

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО»
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО»
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Кваліфікаційна наукова
праця на правах рукопису

Кузнєцова Наталія Володимирівна

УДК 303.732.4, 519.226

ДИСЕРТАЦІЯ
МЕТОДИ І МОДЕЛІ АНАЛІЗУ, ОЦІНЮВАННЯ ТА ПРОГНОЗУВАННЯ
РИЗИКІВ У ФІНАНСОВИХ СИСТЕМАХ

Спеціальність: 01.05.04 – Системний аналіз і теорія оптимальних рішень

Подається на здобуття наукового ступеня доктора технічних наук

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

_____ Н.В.Кузнєцова

Науковий консультант
Бідюк Петро Іванович
доктор технічних наук, професор

Київ – 2018

АНОТАЦІЯ

Кузнєцова Н. В. Методи і моделі аналізу, оцінювання та прогнозування ризиків у фінансових системах. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора технічних наук за спеціальністю 01.05.04 – системний аналіз і теорія оптимальних рішень. – Національний технічний університет України “Київський політехнічний інститут імені Ігоря Сікорського”, Міністерство освіти і науки України, Київ, 2018.

Дисертаційна робота присвячена вирішенню важливої науково-прикладної проблеми створення системної методології дослідження ризиків фінансових систем та розробки єдиної інформаційної технології моніторингу і менеджменту ризиків для забезпечення їх мінімізації, підвищення ефективності аналізу, оцінювання та прогнозування фінансових ризиків (ФР) для складних систем різного типу.

Оцінювання ризиків фінансових систем зазвичай проводиться на основі великих обсягів статистичних даних, з певною невизначеністю як самих даних, так і засобів їх збору та обробки. Побудова будь-яких прогнозів також висуває досить суворі вимоги щодо якості і обсягів даних. Вимоги щодо швидкості, якості, правильності опрацювання ризиків у сучасному конкурентному середовищі весь час зростають. Тому актуальним є завдання створення нових методів, моделей та технологій для підтримки прийняття управлінських рішень щодо опрацювання ризиків у фінансових системах.

У дисертаційній роботі критично проаналізовано вплив невизначеності, неточності та нечіткості на формування ризиків. Розроблено класифікацію різних типів невизначеностей і пов’язаних з ними ризиків, існуючих способів і методів обробки пропущених значень та доведено необхідність подолання невизначеностей при моделюванні та оцінюванні ризиків. Виконано огляд існуючих моделей оцінювання фінансового ризику, проаналізовано особливості та обмеження щодо їх застосування.

Аналіз стану проблеми моделювання та прогнозування ФР показав, що відсутній єдиний підхід до оцінювання та моніторингу ФР. Не виконується глибинний аналіз причин появи невизначеностей, варіантів реалізації ризику, не вистачає засобів швидкого напрацювання антиризикових рекомендацій для більшості напрямів фінансової діяльності через відсутність досвіду, статистичної інформації та аналітичного інструментарію.

Виконано дослідження особливостей існуючого інструментарію для ризик-менеджменту та його недоліків, що потребує розробки нових методів і технологій для опрацювання ризиків фінансових систем. Зокрема, запропоновано узагальнити та розвинути застосування скорингових карт на всі типи ФР та впровадити поведінкові скорингові карти для динамічного оцінювання ФР. Запропонований нейро-нечіткий метод доповнення вибірки відхиленими даними, який полягає у тому, що на навчальній вибірці будується нейронна мережа, яка і прогнозує значення ймовірності для відхилених заявок, а ваги визначаються за нечітким підходом.

Розроблена системна методологія дослідження та оцінювання фінансових ризиків базується на принципах та процедурах системного аналізу і враховує основні принципи менеджменту ризиків. Системна методологія передбачає статичне та динамічне оцінювання. Особливістю статичного оцінювання є розширення застосування скорингової карти до аналізу ризиків різної природи, розроблений нейро-нечіткий метод урахування відхилених раніше даних, застосування розробленого критерію урахування інформаційного ризику як прояву невизначеностей різних типів та зовнішніх впливів і інформаційних загроз. Використання власного комбінованого методу обробки неповних даних при моделюванні фінансових ризиків дозволило здійснити глибинний аналіз повноти, інформативності та якості даних, проаналізувати причини появи пропусків та здійснити їх відновлення шляхом побудови множини регресійних моделей та обрання кращої з них.

Динамічне оцінювання базується на запропонованих принципах динамічного оцінювання та адаптивного менеджменту ризиків. Розроблено

ймовірісно-статистичний метод оцінювання ризику фінансових втрат, який ґрунтується на комбінованому використанні оптимального фільтра, регресійного та байєсівського аналізу, і за рахунок використання динамічної мережі Байєса дозволяє динамічно оцінювати втрати від реалізації фінансового ризику. На основі методу розроблено різні типи комбінованих моделей і апробовано їх практичне застосування для оцінювання ринкового фінансового ризику.

У кваліфікаційній роботі розроблено динамічний метод оцінювання ризиків, який передбачає побудову різних типів моделей виживання (параметричних, непараметричних), з можливістю прогнозування рівня небезпеки та ступеню ризику і ключових моментів часу, які характеризують допустимий, критичний та катастрофічний рівень ризику. Розроблені алгоритми визначення моменту настання ризику за заданим критичним (катастрофічним) ступенем та рівнем ризику.

Науковими результатами досліджень також є розвиток принципу адаптивності у менеджменті ФР і запропонований метод структурно-параметричної адаптації, який дозволяє враховувати наявність невизначеностей, нестационарності, нелінійності фінансових процесів, враховувати сезонні ефекти та який відрізняється використанням множини статистичних критеріїв для оцінювання структури і параметрів моделей, і забезпечує побудову адекватних моделей завдяки використанню комплексних критеріїв якості. Метод структурно-параметричної адаптації містить оригінальний квадратичний критерій якості опрацювання ризику та інтегральну характеристику оцінювання ефективності методів менеджменту ризиків, що є корисними для прийняття управлінських рішень.

У дисертаційній роботі створено розширену інформаційну технологію та інформаційну систему підтримки прийняття рішення (ІСППР) на основі запропонованої системної методології, що об'єднує методи і моделі динамічного, ймовірісно-статистичного, регресійного оцінювання фінансових ризиків, комбіновані методи і моделі відновлення пропущених і втрачених

даних, методи і моделі структурно-параметричної адаптації. Показано, що запропонована інформаційна технологія нового типу може бути реалізована у вигляді клієнт-серверної архітектури, мікросервісів та хмар, є гнучкою і адаптивною з практичної точки зору як до задач, так і до інтеграції в існуючу інформаційну систему функціонуючого підприємства.

Наукова новизна одержаних результатів визначається такими теоретичними і практичними результатами, отриманими автором:

Уперше:

➤ запропоновано методологію системного аналізу фінансових ризиків з урахуванням наявності невизначеностей, характерних для процесів моделювання, прогнозування і оцінювання можливих втрат. Методологія відрізняється тим, що містить такі функціональні складові: методи обробки пропущених та втрачених даних, випадкових збурень стану та похибок вимірів; метод адаптації моделей ризиків до нових даних; спосіб побудови адекватних моделей байєсівського типу та прогнозів можливих втрат на задану кількість кроків і забезпечує розв'язання задач попередньої обробки даних, побудови адаптивних моделей і обчислення прогнозних оцінок;

➤ запропоновано принцип інтегрованого динамічного урахування факторів часу, ступеня та рівня ризику у фінансовому ризик-менеджменті і метод динамічного оцінювання та прогнозування фінансових ризиків, що відрізняються визначенням моменту переходу на вищий ступінь ризику і забезпечують можливість прогнозування критичного моменту настання ризику;

➤ розроблено комбінований метод обробки неповних даних, який відрізняється застосуванням мережі Байєса (для виявлення причин і наслідків неповноти даних) та регресійних моделей (для прогнозування пропущених значень), що дозволяє відновити втрачені дані і усунути невизначеності;

➤ запропоновано ймовірнісно-статистичний метод оцінювання ризику фінансових втрат, який ґрунтується на комбінуванні статистичних та ймовірнісних методів з метою поглибленого аналізу фінансових ризиків,

урахування невизначеностей різної природи, побудови адекватних моделей та отримання високоякісних оцінок прогнозів;

➤ запропоновано принцип адаптивного менеджменту ризиків та новий метод структурно-параметричної адаптації ймовірнісно-статистичних моделей, що відрізняються налаштуванням моделей оцінювання фінансових втрат як функцій часу, та використанням множини статистичних критеріїв для оцінювання якості структури і параметрів моделей, і забезпечують побудову адекватних моделей досліджуваних процесів і врахування очікуваних втрат, відхилень та компенсаційних втрат на подолання ризиків.

Удосконалено та розвинуто:

➤ методологію застосування скорингової карти як інструменту оцінювання фінансового ризику, яка відрізняється наявністю нейро-нечіткого методу доповнення вибірки відхиленими заявками при моделюванні фінансових ризиків, урахуванням інформаційної складової при оцінюванні та менеджменті фінансових ризиків, що забезпечує підвищення якості оцінок можливих втрат;

➤ методологію побудови систем підтримки прийняття рішень для розв'язання задач моделювання, оцінювання і прогнозування фінансових ризиків, яка відрізняється комплексним застосуванням ймовірнісно-статистичного та динамічного методів, множини критеріїв якості даних, моделей та прогнозів, і забезпечує побудову адекватних математичних моделей фінансових ризиків та отримання високоякісних оцінок прогнозів можливих втрат.

Практичне значення одержаних результатів полягає у розробці методології системного аналізу ФР в умовах наявності невизначеностей із застосуванням методів динамічного оцінювання та прогнозування ФР, урахуванням часу та оцінюванням моменту настання переходу на вищий ступінь ризику. Для практичного вирішення проблеми неповноти даних запропоновано методику обробки і заповнення пропущених значень та розроблено авторський комбінований ймовірнісно-регресійний метод заповнення втрачених даних.

Важливим практичним результатом є розробка інформаційної технології та інформаційної системи підтримки прийняття рішень на основі запропонованих моделей і методів, що дозволяє підвищити ефективність обробки фінансових даних різних типів, виконати аналіз фінансових та інформаційних ризиків. Розроблені елементи ІСППР для поведінкових моделей дозволили встановлювати найбільш ризикованих, з точки зору відтоку, клієнтів, ранжувати їх за очікуваними втратами та оцінювати фінансове положення компанії в цілому через загальний обсяг можливих втрат.

Запропоновані методи і моделі доведені до рівня практичної реалізації і забезпечують підвищення якості оцінок можливих втрат завдяки урахуванню інформаційної складової в моделях оцінювання та менеджменту ФР.

Скорингові моделі для динамічного оцінювання і прогнозування ФР на основі пропорційних та непараметричних моделей дали можливість завчасно визначити та спрогнозувати поведінку клієнтів, які будуть користуватися послугами, і спрогнозувати до 97% потенційно схильних до відтоку клієнтів. Запропонований принцип інтегрованого динамічного урахування факторів часу, ступеня та рівня ризику дозволив оцінювати клієнтів, які виїжджають за кордон, прогнозувати їх витрати та кількість днів у роумінгу та, відповідно, витрати компанії, що дозволило зменшити втрати на 12%.

Результати дисертаційної роботи використані та впроваджені в Національному банку України, компаніях ТОВ «САС Інстїт्यूт ЕЛ.ЕЛ.СІ.», ТОВ «ЕЛСІКО», телекомунікаційній компанії ТОВ «Лайфселл», іноземній компанії Artcom Venture GmbH та у навчальному процесі кафедри ММСА ІПСА НТУУ «КПІ ім. Ігоря Сікорського», що підтверджується відповідними актами та довідками про впровадження.

Ключові слова: системна методологія, фінансові системи, динамічне оцінювання фінансових ризиків, скорингові карти, структурно-параметрична адаптація, інформаційні ризики, мережі Байєса, ймовірно-статистичний метод оцінювання ризику фінансових втрат, комбінований метод обробки

неповних даних, квадратичний критерій якості опрацювання ризику, принцип інтегрованого динамічного оцінювання ризиків.

ABSTRACT

Kuznietsova N.V. Methods and models for analysis, estimating and forecasting of the risks for financial systems. – Qualifying scientific work as a manuscript.

Thesis for scientific degree of Doctor of Technical Sciences on the specialty 01.05.04 – system analysis and the theory of optimal decisions. – National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ministry of Education and Science of Ukraine, Kyiv, 2018.

The thesis is devoted to solving an important scientific and applied problem of creating a systemic methodology for studying the risks of financial systems and developing a unified information technology for monitoring and risk management to ensure their minimization, to increase the efficiency for financial risk analysis, estimation and forecasting for different types of complex systems.

Risk assessment of financial systems is usually carried out on the basis of large statistical data volumes, with some uncertainty both of the data by themselves and of their collection and processing means. Making any forecasts also imposes very strict requirements regarding the data quality and volumes. The requirements for speed, quality, and correct risks' processing in the current competitive environment are constantly increasing. Therefore, the task of new methods, models and technologies developing is urgent in order to support the adoption of managers' decisions in dealing with risks in financial systems.

In the thesis the influence of uncertainty, inaccuracy and fuzziness on the formation of risks is critically analyzed. The classification of various types of uncertainties and associated risks, existing means and methods of handling missing values and the need to combat uncertainties in modeling and risk assessment has been developed. It has been done a review of existing financial risk assessment models, and the main features and limitations to their application have been analyzed.

The analysis of the problem state of modeling and forecasting of the financial risks showed that there is no unified approach to the financial risks' evaluation, and monitoring. There is no in-depth analysis of the uncertainties occurrence reasons, risks implementations, and also there is a lack of the rapidly developing means recommendations against risks for most areas of financial activity due to the lack of experience, statistical information and analytical tools.

The features of the existing risk management tools and disadvantages that require the development of new methods and technologies for processing risks of financial systems were investigated. In particular, it is proposed to generalize and develop the use of scorecards for all financial risks types and to introduce behavioral scorecards for financial risks' dynamic evaluation. A neural-fuzzy method for adding rejected data in datasets was proposed. The method consists in the fact that a neural network is constructed on the training sample, which predicts the value of the probability for rejected data and the weights are determined by the fuzzy approach.

The developed system methodology of financial risks investigation and evaluation is based on the basic principles and procedures of system analysis and takes into account the basic principles of risk management. The system methodology proposes static and dynamic evaluation. The peculiarity of the static estimation is the expansion of the use of a scorecard to the risks' analysis of various types, developed neural-fuzzy method of incorporating previously rejected data, the application of the developed criteria of taking into account information risk as a manifestation of various types of uncertainties and external influences and information threats. The use of the own combined method of processing incomplete data for financial risks modeling allowed to carry out a deep analysis of the completeness, informativeness and quality of data. It also gives the possibility to analyze the causes of the gaps occurrence and carry out their recovering by constructing a set of regression models and selecting the best model.

The dynamic evaluation is based on the proposed dynamic assessment principles and adaptive risk management. The probabilistic-statistical method of estimating the risk of financial losses based on the combined use of optimal filter,

regression and Bayesian analysis, and by using the Bayesian dynamic network allows estimating dynamically the losses from the financial risk realization. Various types of combined models on the basis of the method have been developed. Their practical application for the market financial risk estimation has been tested.

In the qualification work, a dynamic method of risk assessment was developed that envisage the construction of different survival models types (parametric, nonparametric), with the ability to predict the risk level and the degree, and that moment which characterize the permissible, critical and catastrophic level of risk. Algorithms for determining the moment of risk occurrence by the given critical (catastrophic) risk degree or level were developed.

The scientific results of the research are also in the fact that it was developed the adaptive principle of the financial risks' management. The structural and parametric adaptation method which allows taking into account the existence of uncertainties, non-stationary features and nonlinearities for financial processes was proposed. It also allows to take into account season effects and to use different set of statistical criteria for evaluating the model structure and parameters. It provides the building of the adequate models through the use of complex quality criteria. The method of structural and parametric adaptation contains the original quadratic criteria of the risk quality processing. Also the integral characteristic for the effectiveness of risk management methods is estimated. Both criteria are useful for making decisions.

In the dissertation the extended information technology and information decision support system (IDSS) was created on the basis of the proposed system methodology, combining methods and models of dynamic, probabilistic and statistical, regression evaluation of financial risks, combined methods and models for missing data recovering, structural-parametric adaptation. It is shown that the proposed new type information technology can be implemented in the form of client-server architecture, microservices and clouds. It is flexible and adaptive both to the practical tasks and to the integration into the existing enterprise information system.

The scientific novelty of the obtained results is determined by the following theoretical and practical results obtained by the author:

For the first time:

- the methodology of the financial risks systemic analysis is proposed, which takes into account the uncertainties typical for the simulation, forecasting and evaluation of possible losses. The methodology contains the following functional components: methods for missing and lost data processing, random state disturbances and measurement errors processing; method of the risk models adaptation to new data; method for constructing adequate Bayesian models and possible losses predictions for a given number of steps. The methodology provides solutions of the tasks for preliminary data processing, the adaptive models construction and the forecasts scores calculation;
- the principle of integrated dynamic accounting of time, degree and level of risk factors in financial risk management and the dynamic method of the financial risks estimation and forecasting are proposed. They take into account the time and estimate the moment of the risk transition to a higher degree and provide the possibility of the critical moment of risk occurrence forecasting;
- the combined method of processing incomplete data, which differs by using the Bayesian network (to identify causes and effects of incompleteness in data) and regression models (to predict missing values), has been developed, which allows to recover the lost data and eliminate uncertainties;
- the probabilistic and statistical method for estimating the risk of financial losses based on a combination of statistical and probabilistic methods for the purpose of financial risks in-depth analysis, the uncertainties of different nature account, adequate models construction and obtaining high quality forecast estimates are proposed;
- the principle of adaptive risk management and a new method of structural and parametric adaptation of probabilistic-statistical models were proposed. They are characterized by the set of models for estimating financial losses as functions of time, and use a set of statistical criteria for assessing the quality of models' structure and parameters and provide the construction of adequate models, estimation of expected losses, deviations and compensatory losses to overcome risks.

Further improved and developed:

- the methodology of using a scorecard as a tool for financial risk assessment, which differs by the presence of a neural-fuzzy method of completing the sample with rejected data during financial risks modeling, taking into account the information component in the assessment and management of the financial risks, which provides an improvement in the quality of possible losses scores;
- the methodology for building decision support systems for solving the problems of modeling, estimation and forecasting of the financial risk. It is characterized by the complex application of probabilistic statistical and dynamic methods, the set of data quality criteria, models and forecasts quality scores. It provides the construction of adequate mathematical models for financial risks and obtaining high-quality assessments of loss forecasts.

The practical significance of the obtained results are: development of the methodology for the financial risks systemic analysis in the presence of uncertainties using the methods of dynamic financial risks estimation and forecasting, taking into account the time and evaluation of the moment of risk transition to a higher degree. As a practical solution of the problem of incompleteness of data it was proposed the methodology of missing values processing and recovering. It was proposed the authors' combined probabilistic regression method for imputation the lost data. An important practical result is development of information technology and information decision support system based on the proposed models and methods, which allows improving the efficiency of processing various types of financial data, to perform an analysis of financial and information risks. The developed IDSS components for behavioral models have allowed identifying the customers with the outflow risk, ranking them for expected losses, and evaluating financial position of a company as a whole due to the total amount of possible losses.

The offered methods and models are brought to the level of practical realization and provide an increase in the quality of estimations of possible losses due to consideration of the information component in the models of evaluation and management of the financial risks.

Scoring models for dynamic estimation and forecasting of financial risks based on proportional and non-parametric models have made it possible to identify and predict behavior of the clients who will use the services and predict up to 97% of potentially prone for outflow clients. The proposed principle of integrated dynamic accounting of time, degree and level of risk factors allowed estimating clients who travel abroad, to forecast their costs and the number of days in roaming and, accordingly, company expenses, which allowed reducing losses by 12%.

The results of the thesis fulfillment were used and implemented at the National Bank of Ukraine, companies SAS Institute LLC, ELSIKO LLC, telecommunication company, LLC “Lifecell”, foreign company Artcom Venture GmbH and in the educational process of the Department of MMSA IASA NTUU “Igor Sikorsky KPI”, which is confirmed by the relevant acts and certificates of implementation.

Keywords: system methodology, financial systems, dynamic assessment of financial risks, scoring cards, structural and parametric adaptation, information risks, Bayesian networks, probabilistic statistical method for estimating the risk of financial losses, combined method of processing incomplete data, quadratic risk quality assessment criterion, integrated dynamic risk assessment principle.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Кузнєцова Н. В., Бідюк П. І. Інформаційна технологія аналізу фінансових даних на основі інтегрованого методу. *Системні дослідження та інформаційні технології*. 2011. № 1. С. 22–33. (Входить до наукометричних баз Index Copernicus, DOAJ, ВІНІТІ, РІНЦ).
2. Бідюк П. І., Кузнєцова Н. В., Терентьєв О. М. Система підтримки прийняття рішень для аналізу фінансових даних. *Наукові вісті НТУУ “КПІ”*. 2011. №1. С. 48–61. (Входить до наукометричних баз Index Copernicus, DOAJ, EBSCO, WorldCat, J-Gate та ін.).
3. Bidyuk P. I., Konovalyuk M. M., Kuznetsova N.V., Pudlo I. V. Adaptive Short-Term Forecasting of Selected Financial Processes. *Research bulletin of NTUU*

“KPI”. 2014. N1. P.35–41. (Входить до наукометричних баз Index Copernicus, DOAJ, EBSCO, WorldCat, J-Gate та ін.).

4. Кузнецова Н. В. Деякі аспекти мінімізації інформаційних ризиків у банківській діяльності. *Системні дослідження та інформаційні технології*. 2014. № 1. С. 7–19. (Входить до наукометричних баз Index Copernicus, DOAJ, ВІНІТІ, РІНЦ).

5. Бидюк П. И., Кузнецова Н. В. Прогнозирование волатильности финансовых процессов с помощью моделей условной дисперсии. *Проблемы управления и информатики*. 2014. № 5. С. 47–54. (Входить до наукометричних баз **Scopus**, ISI Thomson Reuters, Mathematical Review, Applied Mechanics Reviews, ВІНІТІ, РІНЦ).

6. Кузнецова Н. В. Практичні підходи до визначення та урахування невизначеностей, що формують фінансові ризики. *Тр. Одес. политехн. ун-та*. 2014. Вып.2 (44). С. 160–170. (Входить до наукометричних баз Index Copernicus, DOAJ, EBSCO, WorldCat, РІНЦ та ін.).

7. Кузнецова Н.В., Бидюк П. И. Нейронні та мережі Байеса у задачі аналізу кредитних ризиків. *Реєстрація, зберігання і обробка даних*. 2015. Т.17, №2. С. 61–71. (Входить до наукометричних баз Index Copernicus, ICI World of Journals).

8. Kuznietsova N. V., Bidiuk P. I. Business Intelligence Techniques For Missing Data Imputation. *Research bulletin of NTUU “KPI”*. 2015. N5. P. 47–56. (Входить до наукометричних баз Index Copernicus, DOAJ, EBSCO, WorldCat, J-Gate та ін.).

9. Кузнецова Н. В. Виявлення та оброблення невизначеностей у формі неповних даних методами інтелектуального аналізу. *Системні дослідження та інформаційні технології*. 2016. № 2. С. 104–115. (Входить до наукометричних баз Index Copernicus, DOAJ, ВІНІТІ, РІНЦ).

10. Kuznietsova N. V. Scoring technology for risk assessment of fraud in banking *CEUR Workshop Proceeding (ISSN 1613-0073)*. 2016. Vol. 1813. P.54-61

[Online]. Available: <http://ceur-ws.org/Vol-1813/paper8.pdf> (last accessed 25.09.2018). (Входить до наукометричних баз **Scopus**, DBLP).

11. Кузнецова Н.В., Бідюк П. І. Моделювання фінансового ризику в телекомунікаційній сфері. *Наукові вісті НТУУ “КПІ”*. 2017. №5. С. 51–58. (Входить до наукометричних баз Index Copernicus, DOAJ, EBSCO, WorldCat, J-Gate та ін.).

12. Кузнецова Н. В., Бидюк П. И. Моделирование кредитных рисков на основе теории выживания. *Проблемы управления и информатики*. 2017. №6. С. 33 – 46. (Входить до наукометричних баз **Scopus**, ISI Thomson Reuters, Mathematical Review, Applied Mechanics Reviews, ВІНІТІ, РІНЦ та ін.).

13. Kuznietsova N. V. Information Technologies for Clients’ Database Analysis and Behaviour Forecasting. *CEUR Workshop Proceeding (ISSN 1613-0073)*. 2017. Vol. 2067. P.56-62 [Online]. Available: <http://ceur-ws.org/Vol-2067/> (last accessed 25.09.2018). (Входить до наукометричних баз **Scopus**, DBLP).

14. Кузнецова Н. В. Розробка скорингових карт для аналізу ризиків банківської діяльності. *Реєстрація, зберігання і обробка даних*. 2017. Т.19. №4.С.35–44. (Входить до наукометричних баз Index Copernicus, ICI World of Journals).

15. Кузнецова Н. В., Бидюк П. І. Динамічне моделювання фінансових ризиків. *Індуктивне моделювання складних систем*. 2017. Вип. 9. 2017. С. 122–137.

16. Кузнецова Н. В. Інформаційні технології для аналізу фінансових зловживань на платформі ProZorro. *Наукові праці ВНТУ: електронне наукове фахове видання*. Електронні дані. Вінниця: Вінниц. Нац. тех. ун-т: 2018. №1. URL: <https://praci.vntu.edu.ua/index.php/praci/article/view/534/525>. (дата звернення 25.09.2018). (Входить до наукометричної бази РІНЦ).

17. Бідюк П. І., Кузнецова Н. В. Ймовірно-статистичний метод оцінювання ризику фінансових втрат. *KPI Science News*. 2018. №2. С. 7–17. (Входить до наукометричних баз Index Copernicus, DOAJ, EBSCO, WorldCat, J-Gate та ін.).

18. Кузнєцова Н. В. Фінансовий ризик-менеджмент з урахуванням інформаційних ризиків. *Реєстрація, зберігання і обробка даних*. 2018. Т.20. №1. С. 30–39. (Входить до наукометричних баз Index Copernicus, ICI World of Journals).
19. Кузнєцова Н. В. Аналіз та прогнозування ризиків шахрайства з кредитними картками. *Інформатика та математичні методи в моделюванні*. 2018. Т.8, №1. С. 16–25. (Входить до наукометричних баз Index Copernicus, РІНЦ).
20. Кузнєцова Н.В., Бідюк П. І. Система підтримки прийняття рішень для аналізу інвестиційних ризиків фінансових ринків. *Вісник Національного університету «Львівська політехніка». Серія «Інформаційні системи та мережі»*. 2018. Вип. 887. С. 115–121.
21. Кузнєцова Н.В., Бідюк П.І. Структурно-параметрична адаптація ймовірно-статистичних моделей для оцінювання фінансових ризиків. *KPI Science News*. 2018. №3. С.23–34. (Входить до наукометричних баз Index Copernicus, DOAJ, EBSCO, WorldCat, J-Gate та ін.).
22. Кузнєцова Н. В., Бідюк П. І. Системний підхід до менеджменту фінансових ризиків. *Системні дослідження та інформаційні технології*. 2018. № 2. С. 124–140. (Входить до наукометричних баз Index Copernicus, DOAJ, ВІНІТІ, РІНЦ).
23. Кузнєцова Н. В., Бідюк П. И. Выявление рисков мошенничества в ходе тендерных закупок методами теории выживания. *Управляющие системы и машины*. 2018. № 2 (274). С.87–96. (Входить до наукометричної бази РІНЦ).
24. Кузнєцова Н. В., Бідюк П. І. Аналіз фінансових ризиків з використанням SAS-технологій обробки даних. *Електротехнічні і комп'ютерні системи*. 2016. № 22(98). С. 267–271. (Входить до наукометричних баз Index Copernicus, Ulrich's Periodicals Directory, РІНЦ, ВІНІТІ).
25. Кузнєцова Н. В. Практичне застосування методології розробки скорингових карт для аналізу ризиків автокредитування. *Електротехнічні та комп'ютерні системи*. 2017. № 24(100). С. 104 – 111. (Входить до

наукометричних баз Index Copernicus, Ulrich's Periodicals Directory, ПІНЦ, ВІНІТІ).

26. Кузнєцова Н. В. Динамічно-рівномірна дискретизація даних у гібридних мережах Байєса. *Вісник Хмельницького Національного Університету*. 2011. №5. Т.3(180). С. 247–251. (Входить до наукометричних баз Index Copernicus, ПІНЦ, Polish Scholarly Bibliography).

27. Кузнєцова Н. В. Методика оцінювання ризику зниження фінансової стабільності за допомогою мереж Байєса. *Вісник Хмельницького Національного Університету*. 2013. №2. Т.3 (198). С. 187–190. (Входить до наукометричних баз Index Copernicus, ПІНЦ, Polish Scholarly Bibliography).

28. Kuznietsova N. Requirements management. *Summer School IT-SEA 5-20 July'14. Report*. Odessa National Polytechnic University. Odessa, 2014. P. 98–121.

29. Kuznietsova N. Project management. Student's hackathon for mobile application *Summer School IT-SEA 5–20 July'14. Report*. Odessa National Polytechnic University. Odessa. 2014. P. 122–130.

30. Кузнєцова Н. В. Інформаційні технології обробки та аналізу даних у фінансовому ризик-менеджменті. *Інформаційні технології та спеціальна безпека*. 2015. №1. С. 86–98.

31. Кузнєцова Н. В. Скорингові карти для аналізу ризиків банківської діяльності. *Інформаційні технології та спеціальна безпека*. 2018. №1. С. 26–35.

32. Белевець М. С., Бідюк П. І., Кузнєцова Н. В. Розробка системного підходу до менеджменту ризиків. *Системні науки та кібернетика*. 2015. №1. С. 30 – 48. URL: http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc_1_2015.pdf.

33. Гуськова В. Г., Кузнєцова Н. В. Комбінування оцінок прогнозів, обчислених за різними методами для обраних країн світу. *Системні науки та кібернетика*. 2015. №1. С. 49 – 57. URL: http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc_1_2015.pdf.

34. Кузнєцова Н. В., Кінда В. В. Обробка і аналіз даних обрахунку тепловтрат приміщень у реальному часі. *Системні науки та кібернетика*. 2015.

№1. С. 77 – 86. URL:
http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc_1_2015.pdf.

35. Фомін О. В., Кузнєцова Н. В. Скорингові моделі поведінки клієнтів-власників кредитних карток для оцінки їх платоспроможності. *Системні науки та кібернетика*. 2016. № 5. С. 56 – 67. URL:

http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc_5_2016.pdf.

36. Кузнєцова Н. В., Куца К.В., Штогрін С. Р. Застосування методології аналізу виживання для дослідження споживчих ризиків. *Системні науки та кібернетика*. 2017. №6. С. 126–135. URL:

http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc_6_2017.pdf.

37. Kuznyetsova N. Informational technologies for financial data analysis. *Proc. IV Int. Scientific Conf. "Intelligence, Integration, Reliability"*. Kyiv-Warsaw, April 21–22, 2011). K.: NTUU "KPI". 2011. P. 32–33.

38. Кузнєцова Н. В. Інтегровані моделі аналізу фінансових даних. *Системний аналіз та інформаційні технології (SAIT 2011)*: матеріали XIII міжнар. наук.-техн. конф. (Київ 23–28 травня, 2011 р.). К.: ННК "ІПСА" НТУУ "КПІ", 2011. С. 274.

39. Кузнєцова Н. Аналіз даних клієнтів за допомогою мереж Байєса. *Комп'ютерні науки та інженерія*: мат. V міжнар. конф. молодих вчених CSE-2011, (Львів, 24–26 листоп. 2011 р.). Вид. Львівської політехніки, 2011. С.50–53.

40. Кузнєцова Н. В. Аналіз і оцінювання інформаційних ризиків. *Системний аналіз та інформаційні технології (SAIT 2013)*: матеріали XV міжнар. наук.-техн. конф. (Київ, 27–31 травня, 2013 р.). К.: ННК "ІПСА" НТУУ "КПІ", 2013. С. 122–123.

41. Кузнєцова Н. В. Застосування мереж Байєса до оцінювання інформаційних ризиків. *Информационные технологии и безопасность. Оценка состояния*: мат. междунар. науч.-практ. конф. ИТБ-2013. (Киев, 18 июня 2013г.). Вип. 13. К.: НАН України. 2013. С. 95–104.

42. Кузнєцова Н. В., Бідюк П. І. Інтегрований підхід до аналізу фінансових ризиків. *Сучасна інформатика: проблеми, досягнення та*

перспективи розвитку: матеріали міжнар. наук. конф. (12–13 вересня 2013 року, Київ). Київ: Ін-т. кібернетики ім. В.М.Глушкова НАН України, 2013. С. 153–155.

43. Kuznietsova N.V., Bidiuk P.I. Systemic approach to estimation of financial risks. *Науково-технічна конференція «Інформатика, математика, автоматика. ІМА:2015* (Суми, 20–25 квітня 2015 року. Сумський державний університет). 2015. С. 46–47.

44. Кузнецова Н. В. Питання якості, конфіденційності та коректності даних в інформаційних технологіях аналізу фінансових ризиків. *Информационные технологии и безопасность*: мат. междунар. науч.-практ. конф. ИТБ-2015 (Київ, 21 октября 2015 г.). Вып. 15. К.: ИПРИ НАН Украины, 2015. С. 131–136.

45. Bidiuk P. I. Kuznietsova N.V. Decision support system for adaptive processes forecasting in various applied areas with alternative techniques. *Геоинформационные системы и компьютерные технологии эколого-экономического мониторинга*: сб. докл. междунар. научно-техн. конф. (г. Днепропетровск, 13–15 апреля 2016 г.) / под ред. Л. В. Сарычевой: Электрон. данные. Днепропетровск: ГВУЗ «НГУ» МОН Украины, 2016. URL: http://gis.dp.ua/conf2016-publications/sections/iad/1_Bidiuk_Kuznetsova.pdf.

46. Кузнецова Н. В. Скорингові карти як інструмент аналізу ризиків. *Системний аналіз та інформаційні технології (SAIT 2016)*: матеріали 18-ї міжнар. наук.-техн. конф. (Київ, 25 – 29 травня 2016 р.). К.: ННК «ІПСА» НТУУ «КПІ», 2016. С. 106 –107.

47. Кузнецова Н. В. Скорингові технології оцінювання ризиків шахрайства в банківській діяльності. *Информационные технологии и безопасность*: мат. междунар. науч.-практ. конф. ИТБ-2016. (Київ, 1 декабря 2016 г.) К.: ИПРИ НАН Украины, 2016. С. 43 – 47.

48. Кузнецова Н. В., Бідюк П. І. Математичні моделі виживання для прогнозування фінансових ризиків. *Моделирование та прогнозування*

економічних процесів: матеріали XI науково-практичної конференції. (Київ, 6–8 грудня 2017р.). К.: КПІ ім. Ігоря Сікорського, 2017. С. 47–48.

49. Кузнєцова Н. В. Інформаційні технології аналізу клієнтської бази абонентів та прогнозування їх поведінки. *Информационные технологии и безопасность*: мат. междуна. науч.-практ. конф. ИТБ-2017 (Киев, 30 ноября, 2017 г.). К.: ИПРИ НАН Украины, 2017. С. 114 – 120.

50. Кузнєцова Н. В., Фомін О. В. Прогнозування ризику втрати користувачів онлайн-платформи. *Системний аналіз та інформаційні технології (SAIT 2018)*: матеріали 20-ї Міжн. науково-техн. конф., (Київ 21–24 травня, 2018 р.). К.: ННК «ІПСА» НТУУ «КПІ», 2018. №20. С. 139–140.

51. Kuznietsova N., Bidiuk P. Forecasting of Financial Risk Users' Outflow. *IEEE First International Conference on System Analysis & Intelligent Computing (SAIC)*, (Kyiv: 08–12 October). 2018. P.250–255. DOI: <https://ieeexplore.ieee.org/abstract/document/8516782>. (Входить до бази **Scopus**).

52. Kuznietsova N., Seebauer M., Zabelin S. Some Methods for Estimating Financial Risks in Banking. *IEEE First International Conference on System Analysis & Intelligent Computing (SAIC)*, (Kyiv: 08–12 October, 2018) 2018. P.271–274. DOI: <https://ieeexplore.ieee.org/document/8516873>. (Входить до бази **Scopus**).

53. Бідюк П. І., Коршевнік Л. О., Кузнєцова Н. В. Моделі і методи прикладної статистики: **навч. посіб.** з грифом МОН України. Київ: НТУУ «КПІ», 2014. 722 с.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ І СКОРОЧЕНЬ	26
ВСТУП.....	27
РОЗДІЛ 1 Методологічний аналіз проблеми моделювання і мінімізації ризиків фінансових систем	37
1.1. Особливості системного підходу до аналізу ризиків	37
1.1.1. Категорія «ризик»: сутність та основні властивості	38
1.1.2. Ризики фінансових систем: поняття і визначення.....	39
1.1.3. Загальна характеристика процесу аналізу та опрацювання ризику.....	41
1.1.3.1. Рівні ризику	43
1.1.3.2. Аналіз сучасних методів управління ризиками	44
1.1.3.3. Принципи створення ефективних систем управління ризиками.....	46
1.1.4. Ризик-менеджмент на основі існуючих міжнародних та національних стандартів для фінансових систем.....	47
1.1.5. Формалізація рейтингу та рівня ризику	50
1.2. Причини виникнення і класифікація ризиків.....	54
1.2.1. Види фінансових ризиків	54
1.2.2. Аналіз впливу інформаційних ризиків на виникнення та результати оцінювання фінансових ризиків.....	56
1.3. Аналіз застосування економіко-математичних методів і моделей в задачах ідентифікації і оцінювання ризиків фінансових систем	59
1.3.1. Особливості задач оцінювання і прогнозування фінансових ризиків.....	60
1.3.2. Методи інтелектуального аналізу даних для дослідження фінансових процесів і систем.....	62
1.3.2.1. Факторний аналіз	62
1.3.2.2. Нечіткі множини	63
1.3.2.3. Нейронні мережі	64
1.3.2.4. Мережі Байєса	65
1.3.2.5. Дерева рішень	67
1.3.2.6. Методи оптимальної фільтрації	68
1.3.2.7. Регресійні моделі	69
1.3.2.7.1. Лінійна множинна регресія	69
1.3.2.7.2. Логістична регресія.....	71
1.3.3. Сучасні практики оцінювання ризиків у фінансових системах ...	73
1.4. Невизначеності в задачах моделювання і менеджменту фінансових ризиків.....	82
1.4.1. Вплив невизначеностей на формування ризиків різної природи.....	83

1.4.2. Ідентифікація та класифікація невизначеностей	86
1.4.3. Ілюстрація розкриття невизначеностей	88
1.4.3.1. Неповнота даних як реалізація невизначеності.....	89
1.4.3.2. Існуючі методи відновлення даних	93
1.5. Змістовний аналіз проблеми досліджень	100
Висновки до першого розділу і постановка завдання дисертаційної роботи.....	104

РОЗДІЛ 2 Сучасні інструменти інтелектуального аналізу даних фінансових систем.....	105
2.1. Інструментальні засоби аналізу даних і оцінювання ризиків у фінансовій сфері	105
2.1.1. Програмні засоби ІАД компанії Microsoft	108
2.1.2. Аналітична платформа Tableau	109
2.1.3. Qlik – рішення для бізнес-аналітики	109
2.1.4. Інтелектуальні рішення для бізнесу компанії IBM.....	111
2.1.5. SAP – аналітика та консолідація даних	111
2.1.6. Аналітичні можливості рішень TIBCO	112
2.1.7. AgenaRisk – аналітичне рішення з використанням мереж Байеса.....	113
2.1.8. RapidMiner – інструмент машинного навчання.....	114
2.1.9. STATISTICA – засоби для ризик-менеджменту.....	115
2.1.10. Angoss Predictive Analytics – платформа для прогнозної аналітики	115
2.1.11. Analytic Solver – інструмент бізнес-аналітики	116
2.1.12. OLAP-технології	117
2.1.13. Безкоштовні аналітичні засоби	118
2.2. Продукти компанії SAS Institute для обробки та аналізу даних.....	120
2.2.1. SAS Enterprise Miner – базове аналітичне рішення компанії SAS	120
2.2.2. SAS Credit Scoring for Banking	123
2.3. SPSS (PASW) Statistics – інструментарій для аналізу даних.....	125
2.4. Скорингові карти ризику	127
2.4.1. Застосування скорингових моделей для оцінювання фінансових ризиків	127
2.4.2. Скорингова карта поведінки	131
2.4.3. Методологічні аспекти розробки та впровадження скорингових карт.....	132
2.4.4. Розробка нейро-нечіткого методу доповнення відхиленними заявками при моделюванні фінансових ризиків.....	136
2.4.5. Приклад побудови скорингової карти для аплікаційного	

скорингу.....	140
2.4.6. Скорингова карта як узагальнена міра ризику.....	143
Висновки до другого розділу	145

РОЗДІЛ 3 Розробка системної методології менеджменту фінансових ризиків.....

3.1. Основні складові системної методології менеджменту фінансових ризиків	147
3.1.1. Принципи менеджменту ризиків	148
3.1.2. Принцип урахування інформаційного ризику під час оцінювання фінансового ризику	149
3.1.3. Основні критерії у задачах аналізу та оцінювання фінансових ризиків	151
3.1.4. Економічно-фінансові критерії	152
3.1.5. Ймовірно-статистичні критерії оцінювання оптимальності прийнятих рішень	153
3.1.6. Критерії оцінювання якості даних та адекватності моделей	155
3.1.7. Критерії оцінки якості рішень.....	159
3.2. Розробка системної методології аналізу, оцінювання і прогнозування фінансових ризиків	161
3.3. Опрацювання невизначеностей в системній методології аналізу ризиків	166
3.4. Комбінований метод обробки неповних даних при моделюванні фінансових ризиків	169
3.5. Статичне оцінювання фінансових ризиків.....	175
3.6. Динамічне оцінювання ризиків	177
3.7. Ймовірно-статистичний метод оцінювання ризику фінансових втрат.....	180
3.7.1. Приклад оцінювання ринкового ризику на основі ймовірно-статистичного методу	190
Висновки до третього розділу	193

РОЗДІЛ 4 Розробка методу динамічного оцінювання ризиків на основі теорії виживання.....

4.1. Теоретичні і прикладні аспекти теорії виживання	195
4.1.1. Припущення і означення теорії виживання.....	196
4.1.2. Формалізація задач динамічного прогнозування часу безпроблемності обслуговування банківського кредиту.....	198
4.2. Динамічні моделі прогнозування ризиків	201
4.2.1. Модель пропорційних ризиків Кокса	201
4.2.2. Узагальнена лінійна модель	206
4.2.3. Розвиток моделі пропорційних ризиків.....	207
4.2.4. Непараметрична модель оцінювання фінансових ризиків.....	210

4.3. Оцінки для порівняння динамічних моделей	211
4.4. Розробка алгоритмів прогнозування часу настання ризику.....	212
4.5. Метод динамічного оцінювання ризиків	214
4.6. Приклади динамічного моделювання ризиків фінансових систем	217
4.6.1. Динамічне моделювання кредитних ризиків	217
4.6.2. Приклад динамічного моделювання ризиків телекомунікаційної компанії.....	226
Висновки до четвертого розділу	229

РОЗДІЛ 5 Розвиток адаптивного підходу до побудови моделей ризиків фінансових систем	231
5.1. Принцип адаптивного менеджменту ризиків фінансових систем	232
5.2. Адаптивний підхід до менеджменту фінансових ризиків	233
5.3. Структурно-параметрична адаптація ймовірно-статистичних моделей.....	241
5.3.1. Метод структурно-параметричної адаптації менеджменту ризиків	242
5.3.2. Використання структурно-параметричної адаптації під час моделювання кредитних ризиків	253
Висновки до п'ятого розділу	256

РОЗДІЛ 6 Інформаційні технології оцінювання, моніторингу та менеджменту ризиків фінансових систем	257
6.1. Формування вимог до моделюючого інструментарію	258
6.2. Розширена інформаційна технологія менеджменту ризиків фінансових систем	260
6.2.1. Клієнт-серверна архітектура.....	260
6.2.2. Реалізація розробленої інформаційної технології у вигляді мікросервісів	263
6.2.3. Реалізація хмарних сервісів для ІТ фінансового ризик-менеджменту.....	265
6.3. Розробка архітектури ІСППР.....	267
6.3.1. Основні рівні ІСППР.....	269
6.3.2. Структура ІСППР для менеджменту фінансових ризиків.....	271
6.4. Приклад інформаційної технології для оцінювання ризиків і інтеграції в існуючу систему на основі мікросервісів.....	276
6.5. Інформаційна технологія опрацювання інформаційних ризиків в системі фінансового ризик-менеджменту підприємства	279
6.6. Інформаційна система підтримки прийняття рішень для аналізу інвестиційних ризиків на фінансовому ринку.....	281
6.7. Розробка і реалізація ІТ для динамічного оцінювання ризиків відтоку клієнтів	284
6.8. ІТ та ІСППР менеджменту ризиків на базі SAS.....	288

	25
Висновки до шостого розділу.....	290
ВИСНОВКИ.....	292
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	297
ДОДАТОК А Фінансові ризики: класифікація, моделі і методи оцінювання, приклади	333
А.1. Класифікації фінансових ризиків.....	333
А.2. Огляд моделей оцінювання волатильності фінансових процесів і приклади моделювання.....	335
А.3. Динамічні мережі Байєса	350
А.4. Реалізація EM-алгоритму.....	355
А.5. Приклад комбінованого використання мереж Байєса та нейронних мереж.....	356
ДОДАТОК Б Приклади застосування запропонованих у дисертаційній роботі моделей та методів.....	359
Б.1. Приклад оцінювання ринкового ризику.....	359
Б.2. Задача аналізу електронних закупівель на платформі ProZorro	362
Б.3. Задача аналізу ризику втрати клієнтів в існуючій ERP-системі.....	371
ДОДАТОК В Програмна реалізація окремих модулів ІСППР.....	377
В.1. Реалізація динамічного методу для телекомунікаційної компанії в середовищі SAS.....	377
В.2. Динамічні моделі для аналізу ризиків шахрайства на платформі ProZorro.....	380
В.3. Реалізація на мові R модулів інформаційної технології оцінювання фінансових ризиків онлайн-гемблінгу.....	384
В.4. Реалізація на мові Python окремих мікросервісів для інтеграції в існуючу ERP-систему	396
ДОДАТОК Г Список публікацій здобувача за темою дисертації	404
ДОДАТОК Д Акти впровадження результатів дисертаційної роботи	409

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ І СКОРОЧЕНЬ

AUC	Площа під кривою (Area Under Curve)
EL	Очікувані втрати (Expected Losses)
EM	Алгоритм максимізації математичного сподівання (Expectation-maximization)
IRB	Підхід на основі внутрішніх рейтингів (Internal Ratings-Based Approach)
IV	Інформаційне значення (Informational value)
NPV	Чиста приведена вартість (Net Present Value)
ROC	Операційна характеристика отримувача (Receiver Operation Characteristic)
SEMMA	Вибірка, дослідження даних, перетворення даних, побудова моделей, оцінка моделей (Sample, Explore, Model, Modify, Assess)
WOE	Вага свідчення (Weight of Evidence)
АР	Авторегресія
АРКС	Авторегресія з ковзним середнім
БЗД	База знань та даних
БД	База даних
ДМБ	Динамічна мережа Байєса
ІАД	Інтелектуальний аналіз даних
ІСППР	Інформаційна система підтримки прийняття рішень
ІТ	Інформаційна технологія
МБ	Мережа Байєса
НМ	Нейронна мережа
ОПР	Особа, що приймає рішення
СД	Сховище даних
СК	Скорингова карта
СППР	Система підтримки прийняття рішень
ФК	Фільтр Калмана
ФР	Фінансові ризики
ФС	Фінансова система

ВСТУП

Обґрунтування вибору теми дослідження. У сучасному світі розвиток промисловості, соціальних, економічних та фінансових систем суттєво залежить від здатності адекватно оцінювати і прогнозувати ризики та пов'язані з ними втрати. Ризик – це властивість, притаманна будь-яким видам діяльності, яка описує множинну зміну стійкості системи і проявляється у вигляді імовірнісної невизначеності реалізації її цільових функцій. Різноманітність множини факторів ризиків та їх проявів, вартість засобів та недостатня швидкість процедур оцінювання ризиків ускладнюють можливість їх ефективного опрацювання. Значними практичними проблемами сьогодні є відсутність достатньої статистичної інформації, науково обґрунтованих рекомендацій для оцінювання і зменшення ризиків, складність формалізації задачі аналізу, недостатність математичного апарату для прогнозування ризиків, недосконалість існуючих систем підтримки прийняття управлінських рішень у ризиковій ситуації. Успішність управління фінансовими системами суттєво залежить від якості оцінювання і прогнозування фінансових ризиків. Фінансові ризики (ФР) – це ймовірність виникнення фінансових наслідків в ситуації невизначеності умов функціонування фінансових систем, вони тісно пов'язані з інформаційними і предметними ризиками різних сфер діяльності.

Існуючі підходи до оцінювання ризиків не враховують потреби аналітиків, експертів та осіб, які приймають рішення (ОПР), у якісній, достовірній, оперативній і точній інформації; вплив невизначеностей різних типів; неточність та неповноту як вхідних даних, так і критеріїв оцінювання.

Для розв'язання зазначених наукомістких проблем у дисертації розроблена системна методологія, яка ґрунтується на комплексному використанні ідеологічно різних математичних моделей, методів інтелектуального аналізу даних (ІАД) для оцінювання і прогнозування ризиків, і створення на її основі інформаційних технологій для підтримки прийняття рішень. Задача створення нової методології дослідження ФР та розробки інтегральних методів аналізу

даних у рамках єдиної технології моніторингу та менеджменту ризиків є, безумовно, актуальною.

Запропонована у роботі методологія системного підходу до аналізу, оцінювання та прогнозування ФР спирається на дослідження та теоретичні напрацювання вітчизняних та зарубіжних вчених: методологія системного підходу розроблена і викладена у роботах М. З. Згуровського, Н. Д. Панкратової, В. М. Кунцевича, В. Д. Романенка, В. Я. Данилова; особливості фінансового ризик-менеджменту розглянуто у роботах О. В. Чугунова, О. А. Лобанова, В. В. Вітлінського; моделювання ризиків висвітлено у роботах В. О. Яценка, В. О. Капустяна, А. Б. Камінського, Т. А. Говорушко; математичному моделюванню і аналізу фінансових процесів присвячені роботи Г. Марковіца, М. Шоулса, Р. Мертона, С. А. Положаєнка, В. М. Подладчікова; ймовірнісний підхід до аналізу даних закладений у роботах Т. Байєса, П. Лапласа, Дж. Перла, А. Л. Тулуп'єва; нелінійному аналізу даних присвячені роботи В. С. Мельника, П. О. Касьянова, Я. Г. Гоойера; формалізація невизначеності, аналіз нечітких даних запропоновані у роботах Л. Заде, Е. Мамдані, М. Сугено, Ю. П. Зайченка, В. Є. Снитюка, Є. В. Бодянського; моделі теорії виживання описані Д. Коксом, Е. Капланом, П. Мейєром, М. Степановою, С. Гланцем; розробка інформаційних технологій та систем підтримки прийняття рішень для фінансових систем виконується.

Підвищення ефективності аналізу, оцінювання та прогнозування ризиків фінансових систем шляхом створення системної методології дослідження ФР та розробки єдиної інформаційної технології моніторингу і менеджменту ризиків для забезпечення їх мінімізації є важливою і актуальною науково-прикладною проблемою.

Зв'язок роботи з науковими програмами, планами, темами. Тема дисертаційної роботи відповідає науковим напрямам Інституту прикладного системного аналізу (ІПСА) Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» (КПІ ім. Ігоря Сікорського), зокрема дослідженням систем і методів прийняття рішень,

прогнозування, побудови інтелектуальних інформаційних систем. Робота виконувалась у рамках науково-дослідних робіт: «Розробка інформаційної технології моделювання та оцінювання фінансово-економічних ризиків із врахуванням невизначеностей різної природи (на основі байєсівських моделей)» (№ ДР 0113U000650, 2013–2014 рр.); «Розробка методології системного аналізу моделювання та оцінювання фінансових ризиків» (№ ДР 0115U000356, 2015–2016 рр.); «Проектування сучасних систем сервісів на прикладі мобільної медичної системи для мешканців прифронтових селищ в зоні АТО» (№ ДР 0117U002435, 2017–2018 рр.); та міжнародних європейських проектів: «Innovation hybrid strategy of IT-outsourcing partnership with enterprises 530319-TEMPUS-1-2012-1-DE-TEMPUS-JPHES» (2012–2015 рр.); «Development of a network infrastructure for youth innovation entrepreneurship support on fablab platform 561536-EPP-1-2015-1-UK-EPPKA2-CBHE-JP» (2015–2018 рр.).

Мета і завдання дослідження. *Метою дослідження є підвищення якості моделювання, оцінювання та прогнозування ФР на основі розвитку системної методології аналізу ризиків завдяки розробці нових методів, моделей та засобів опрацювання ФР в умовах нелінійності та нестационарності фінансових процесів, а також неповноти та невизначеності даних.*

Для досягнення поставленої мети необхідно виконати такі *завдання*:

- проаналізувати існуючі методи та підходи до моделювання, оцінювання і прогнозування ФР;
- застосувати системний підхід до аналізу та менеджменту ФР і розробити системну методологію дослідження ФР на основі методів статистичного та інтелектуального аналізу даних, яка враховує динамічний аналіз природи, сутності і причини появи ФР;
- дослідити причини виникнення ризиків у фінансових системах, характер фінансових ризиків і запропонувати спосіб врахування інформаційних впливів;
- удосконалити методику аналізу причин появи невизначеностей різного характеру з врахуванням особливостей фінансових процесів та неповноти

фінансових даних шляхом застосування динамічних та гібридних мереж Байєса;

- розробити новий комбінований метод та моделі аналізу даних на основі комплексного підходу до обробки фінансових даних з урахуванням їх неповноти, неструктурованості, нечіткості та неоднозначності;

- запропонувати математичний інструментарій для оцінювання, ранжування і прогнозування фінансових ризиків з урахуванням попереднього досвіду скорингу, вимог аналітиків та неповноти фінансових даних;

- розробити метод прогнозування ризику фінансових втрат з можливістю оцінювання рівня та ступеню ризику;

- розробити метод динамічного оцінювання та прогнозування ФР з можливістю прогнозування критичного часу та рівнів втрат (допустимого, критичного та катастрофічного);

- запропонувати критерії оцінювання якості опрацювання ризиків та інтегрувати їх разом з існуючими в єдину інформаційну технологію (ІТ) менеджменту ФР;

- запропонувати спосіб адаптації моделей оцінювання моменту настання катастрофічних втрат з урахуванням нестаціонарності та нелінійності фінансових процесів з метою їх подальшого використання в інформаційних системах підтримки прийняття управлінських рішень у фінансових системах;

- розробити інформаційну систему підтримки прийняття рішень (ІСППР) нового типу із застосуванням системного підходу до розробки архітектури, функціональної схеми та процедур аналізу даних і можливістю генерування множини взаємодоповнюючих альтернатив та об'єктивного вибору кращої з них за допомогою кількох множин статистичних критеріїв якості;

- розробити програмний інструментарій підтримки прийняття рішень на основі запропонованої системної методології для подальшого впровадження в існуючі фінансові системи;

- застосувати розроблені інформаційні технології та системи для розв'язання прикладних задач аналізу та менеджменту фінансових ризиків.

Об'єктом дослідження є ризики у фінансових системах, зокрема в банківських установах, телекомунікаційній сфері, страхуванні та підприємницькій діяльності, які потребують коректного опрацювання з метою виявлення засобів їх зниження та підвищення ефективності процесу прийняття управлінських рішень.

Предмет дослідження – методи і моделі дослідження ризиків фінансових систем; методи моделювання нелінійних нестационарних фінансових процесів в умовах невизначеності, неповноти та нечіткості фінансових даних; методи оцінювання параметрів і станів динамічних систем.

Методи дослідження. Особливості об'єкту і мета досліджень обумовлюють необхідність застосування методів: для аналізу й обробки інформації – системного аналізу, інтелектуального і статистичного аналізу даних; методів фільтрації та обробки невизначеностей; для моделювання й оцінювання ФР – ймовірнісно-статистичного та регресійного аналізу даних, байєсівських мереж; для побудови й оцінювання прогнозів – методів прогнозування на основі часових рядів, різницевих рівнянь, регресійного аналізу і байєсівських мереж та методів аналізу виживання; для побудови практичних реалізацій – методів, засобів та інформаційних технологій проектування ІСППР і сучасного програмування.

Наукова новизна одержаних результатів визначається такими теоретичними і практичними результатами, отриманими автором:

Уперше:

➤ запропоновано методологію системного аналізу фінансових ризиків з урахуванням наявності невизначеностей, характерних для процесів моделювання, прогнозування і оцінювання можливих втрат. Методологія відрізняється тим, що містить такі функціональні складові: методи обробки пропущених та втрачених даних, випадкових збурень стану та похибок вимірів; метод адаптації моделей ризиків до нових даних; спосіб побудови адекватних моделей байєсівського типу та прогнозів можливих втрат на задану кількість

кроків і забезпечує розв'язання задач попередньої обробки даних, побудови адаптивних моделей і обчислення прогностичних оцінок;

➤ запропоновано принцип інтегрованого динамічного урахування факторів часу, ступеня та рівня ризику у фінансовому ризик-менеджменті і метод динамічного оцінювання та прогнозування фінансових ризиків, що відрізняються визначенням моменту переходу на вищий ступінь ризику і забезпечують можливість прогнозування критичного моменту настання ризику;

➤ розроблено комбінований метод обробки неповних даних, який відрізняється застосуванням мережі Байєса (для виявлення причин і наслідків неповноти даних) та регресійних моделей (для прогнозування пропущених значень), що дозволяє відновити втрачені дані і усунути невизначеності;

➤ запропоновано ймовірнісно-статистичний метод оцінювання ризику фінансових втрат, який ґрунтується на комбінуванні статистичних та ймовірнісних методів з метою поглибленого аналізу фінансових ризиків, урахування невизначеностей різної природи, побудови адекватних моделей та отримання високоякісних оцінок прогнозів;

➤ запропоновано принцип адаптивного менеджменту ризиків та новий метод структурно-параметричної адаптації ймовірнісно-статистичних моделей, що відрізняються налаштуванням моделей оцінювання фінансових втрат як функцій часу, та використанням множини статистичних критеріїв для оцінювання якості структури і параметрів моделей, і забезпечують побудову адекватних моделей досліджуваних процесів і врахування очікуваних втрат, відхилень та компенсаційних втрат на подолання ризиків.

Набула подальшого розвитку методологія побудови систем підтримки прийняття рішень для розв'язання задач моделювання, оцінювання і прогнозування фінансових ризиків, яка відрізняється комплексним застосуванням ймовірнісно-статистичного та динамічного методів, множини критеріїв якості даних, моделей та прогнозів, і забезпечує побудову адекватних математичних моделей фінансових ризиків та отримання високоякісних оцінок прогнозів можливих втрат.

Удосконалено методологію застосування скорингової карти як інструменту оцінювання фінансового ризику, яка відрізняється наявністю нейро-нечіткого методу доповнення вибірки відхиленними заявками при моделюванні фінансових ризиків, урахуванням інформаційної складової при оцінюванні та менеджменті фінансових ризиків, що забезпечує підвищення якості оцінок можливих втрат.

Практичне значення одержаних результатів полягає у розробці:

– структури та базових модулів інформаційної системи підтримки прийняття рішень на основі запропонованих моделей і методів, що дозволяє підвищити ефективність обробки фінансових даних різних типів, виконати аналіз фінансових та інформаційних ризиків;

– методології системного аналізу ФР в умовах наявності невизначеностей із застосуванням методів динамічного оцінювання та прогнозування ФР, урахуванням часу та оцінювання моменту настання переходу на вищий ступінь ризику;

– методики обробки і заповнення пропущених значень та авторського комбінованого методу обробки неповних даних.

Запропоновані методи і моделі доведені до рівня практичної реалізації:

– забезпечують підвищення якості оцінок можливих втрат за рахунок урахування інформаційної складової в моделях оцінювання та менеджменту ФР, наприклад, під час перевірки і моніторингу банків Національним банком України;

– скорингові моделі для динамічного оцінювання і прогнозування ФР на основі пропорційних та непараметричних моделей дали можливість завчасно визначити та спрогнозувати поведінку клієнтів, які будуть користуватися послугами, і спрогнозувати до 97% потенційно схильних до «відтоку» клієнтів;

– розроблені елементи ІСППР з використанням поведінкових моделей дозволили встановлювати найбільш ризикованих, з точки зору відтоку, клієнтів, ранжувати їх за очікуваними втратами та оцінювати фінансове положення компанії в цілому через загальний обсяг можливих втрат;

– принцип інтегрованого динамічного урахування факторів часу, ступеня та рівня ризику дозволив оцінювати клієнтів, які виїжджають за кордон, прогнозувати їх витрати та кількість днів у роумінгу та, відповідно, витрати компанії, що дозволило зменшити втрати на 12%.

Результати дисертаційної роботи використані та впроваджені в Національному банку України, компаніях ТОВ «САС Інстїт्यूт ЕЛ.ЕЛ.СІ.», ТОВ «ЕЛСІКО», телекомунікаційній компанії ТОВ «Лайфселл», іноземній компанії Artcom Venture GmbH та у навчальному процесі кафедри математичних методів системного аналізу ІПСА НТУУ «КПІ ім. Ігоря Сікорського», що підтверджується відповідними актами, відгуками та довідками про впровадження.

Особистий внесок здобувача. Усі теоретичні та практичні результати дисертаційної роботи, які виносяться на захист, отримані автором особисто. У друкованих працях, написаних у співавторстві, здобувачеві належать: схема аналізу фінансових даних та ІТ з використанням інтегрованого методу [1]; структурна схема ІТ для аналізу фінансових показників та архітектура ІСППР комерційного банку [2]; огляд методів прогнозування, адаптивна модель прогнозування фінансових даних на основі фільтра Калмана та авторегресії з ковзним середнім (АРКС) та їх порівняння [3]; моделі стохастичної волатильності та узагальненої авторегресії з умовною гетероскедастичністю для прогнозування фінансових процесів [5]; комбінована модель на основі нейронних мереж для аналізу ФР [7]; методологія для аналізу невизначеностей і вибору методу заповнення пропусків [8]; оцінювання ФР на основі моделей ІАД та поведінкових моделей [11]; динамічна модель аналізу кредитних ризиків [12]; динамічний підхід до аналізу та оцінювання ФР з прогнозуванням можливого часу настання ризику, ймовірності та рівня втрат [15]; ймовірнісно-статистичний метод оцінювання ризику фінансових втрат [17]; ІСППР для аналізу та розрахунку втрат за інвестиційним портфелем [20]; метод структурно-параметричної адаптації на основі ймовірнісно-статистичних моделей [21]; системна методологія менеджменту ФР [22]; поведінкові моделі

для аналізу заявок і закупівель на електронній платформі ProZorro [23]; архітектура системи аналізу ФР [24]; класифікація моделей для дослідження операційних ризиків, аналіз використання МБ в умовах наявності невизначеності [32]; застосування комбінованого методу прогнозування з використанням авторегресійних моделей та нейронних мереж, експериментальне моделювання [33]; програмна реалізація на основі системного підходу комплексу для оцінювання тепловтрат [34]; розробка поведінкових моделей клієнтів-власників кредитних карт [35]; поведінкові моделі на основі біхевіоріального скорингу [36]; інтегрований підхід для прогнозування ФР [42]; основні положення системного підходу для оцінювання ФР [43]; архітектура системи для прогнозування фінансово-економічних процесів [45]; моделі виживання для прогнозування ФР [48]; моделі ієрархічної кластеризації, xgboost та моделі на основі теорії виживання для прогнозування відтоку користувачів онлайн-платформи [50]; поведінкові моделі для аналізу ризиків втрати клієнтів [51]; методи оцінювання банківських ризиків [52]; розділи 6, 12, 14, 16-21, присвячені факторному та байєсівському аналізу, моделям дискретного вибору та прийняття рішень [53].

Апробація результатів дисертації. Основні положення роботи і наукові результати дисертаційної роботи доповідались і обговорювались на:

– міжнародних та національних науково-технічних конференціях «Системний аналіз та інформаційні технології» (Київ, 2011, 2013, 2016, 2018); «Intelligence, Integration, Reliability» (Kyiv–Warsaw, 2011); «Комп'ютерні науки та інженерія: CSE-2011» (Львів, 2011); «Сучасна інформатика: проблеми, досягнення та перспективи розвитку» (Київ, 2013); «Інформаційні технології і безпека» (Київ, 2013, 2015, 2016, 2017); «Геоінформаційні системи і комп'ютерні технології еколого-економічного моніторингу» (Дніпро, 2016); «Електротехнічні та комп'ютерні системи: теорія і практика (ELTECS-2016)» (Одеса, 2016); «Моделювання та прогнозування економічних процесів» (Київ, 2017); «Інформатика, математика, автоматика» (Суми, 2015); IEEE First International Conference on System Analysis & Intelligent Computing (Київ, 2018);

– міжнародних наукових семінарах (Нансі (Франція), 2016; Анкара (Туреччина), 2017; Будапешт (Угорщина), 2018).

Публікації. За результатами досліджень опубліковано **53** наукові праці, з них **25** публікацій у фахових виданнях (21 публікація у виданнях України, що входять до наукометричних баз, та 2 у закордонних виданнях), **11** публікацій в інших виданнях; **16** публікацій у матеріалах доповідей і збірках праць конференцій; **1** навчальний посібник з грифом МОН України.

Структура та обсяг дисертації. Дисертаційна робота складається зі вступу, переліку умовних позначень, шести основних розділів, висновків, списку використаних джерел і п'яти додатків. Робота викладена на 415 сторінках і містить 294 сторінки основної частини, 83 сторінки додатків, 71 рисунок, 34 таблиці, список використаних джерел із 372 найменувань на 36 сторінках.

РОЗДІЛ 1

МЕТОДОЛОГІЧНИЙ АНАЛІЗ ПРОБЛЕМИ МОДЕЛЮВАННЯ І МІНІМІЗАЦІЇ РИЗИКІВ ФІНАНСОВИХ СИСТЕМ

1.1. Особливості системного підходу до аналізу ризиків

Системний підхід як загальний методологічний напрям наукових досліджень [76, 132] дозволяє провести всебічний аналіз досліджуваної складної системи, враховуючи її цілісність, цілеспрямованість, складність організації, взаємодію із зовнішнім середовищем, виконати глибинний аналіз всіх сторін функціонування, знайти шляхи вирішення системних проблем.

У певному сенсі системний підхід до аналізу ризиків є способом дослідження інтегральних залежностей і взаємодії факторів появи чи прояву ризиків, прогнозування та оцінювання наслідків їх реалізації [78], напрацювання рекомендацій щодо опрацювання ризиків у рамках систем управління ризиками з метою їх мінімізації.

Застосування системного підходу до менеджменту ризиків [28] передбачає певну етапність дій у дослідженні і опрацюванні ризиків, а в реальній діяльності потребує виконання постійного моніторингу ризиків [142], удосконалення процесів збору, обробки і оцінювання інформації, що використовується у процедурах моделювання і напрацювання управлінських рішень.

Наразі відсутні єдині універсальні системи, які можна було б рекомендувати для використання у будь-якій «ризиковій» ситуації, для будь-якого процесу опрацювання ризиків [115]. Це пов'язано не тільки з різноманітністю і множиною факторів ризиків чи їх проявів, а й з фізичною можливістю їх опрацювання (вартість засобів, обсяги даних, швидкість процедур опрацювання, тощо) [126].

Разом з тим, напрацьовано певну кількість методів управління ризиками, які можуть бути поєднані у системи опрацювання ризиків [142].

1.1.1. Категорія «ризик»: сутність та основні властивості

Ризик в основному розуміється як потенційна небезпека, можливість збитку чи шкоди, хоча інколи ризик може мати й позитивний вплив. Ризик за означенням є слабо структурованим, оскільки виникає в результаті невизначеності подій та ситуацій, зовнішнього середовища, незнання щодо дій інших систем, неповноти і неточності формалізації задачі, нечіткості як самих вхідних даних, процесів, так і моделей, що їх описують [93]. В результаті діяльності фінансової системи як суб'єкта впливу ризику може змінюватись як рівень самого ризику за рахунок зміни наслідків та можливих очікуваних втрат (як несприятлива реалізація ризику), так і ймовірність реалізації ризику (від малоімовірних до найімовірніших).

Ризик як філософська категорія [126, 156] характеризується певними властивостями, які є спільними для всіх типів ризику (рис. 1.1).

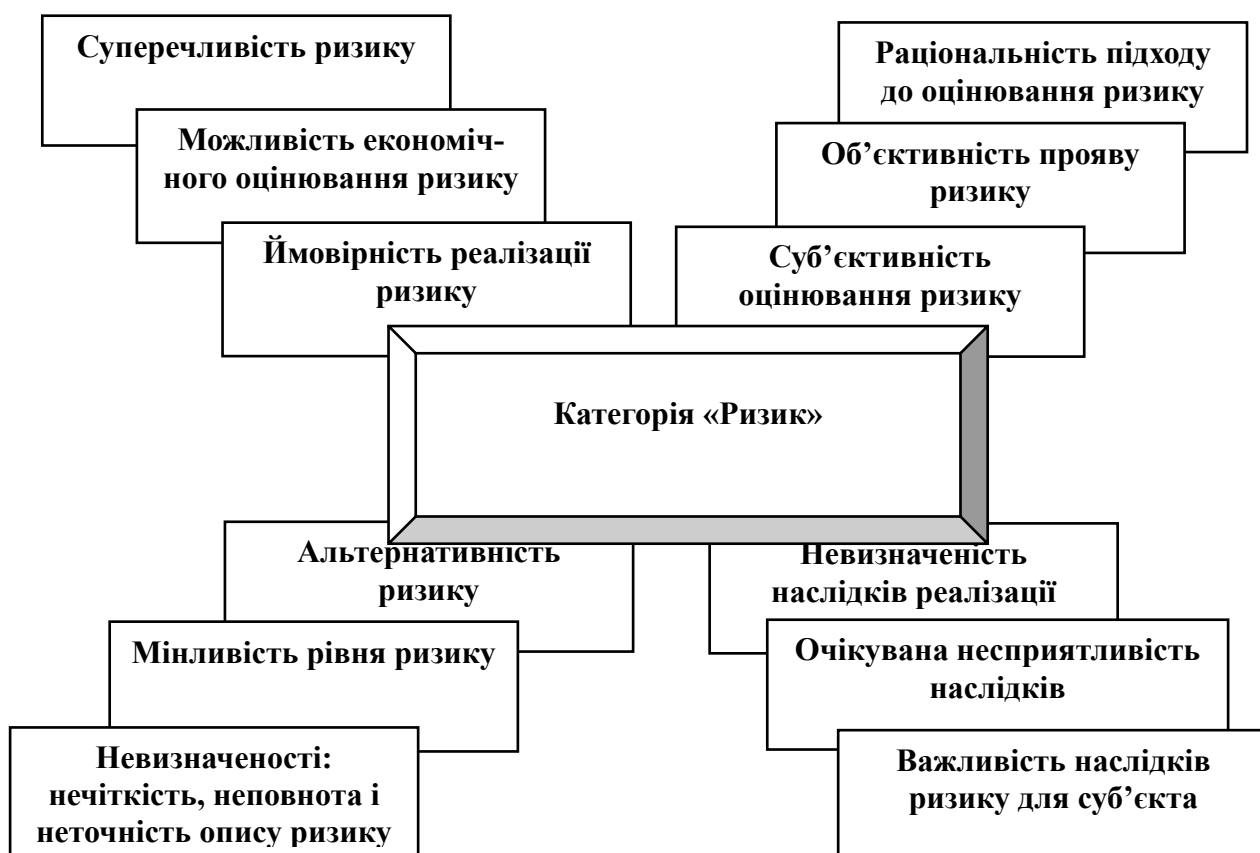


Рис. 1.1 Категорія «Ризик» та її основні характеристики

Вважають, що ризик є причиною несприятливих наслідків (інакше кажучи, розглядають лише його негативну сторону), а тому важливим для

суб'єкта є тип ризику і його безпосередній вплив. Насправді ризик може трактуватись не лише як несприятливі наслідки, а й як неочікуваний дохід [183]. Тут виявляється певна альтернативність, неоднозначність і суперечливість ризику, пов'язана з суб'єктивністю його оцінювання [64]. Під альтернативністю ризику розуміють наявність двох або більше можливостей, варіантів, що створює ризик обрати не найкращу можливість. Якщо можливість вибору відсутня, то не виникає ризикової ситуації, і ризику, відповідно [184].

Суб'єктивність ризику пов'язана з тим, що люди по-різному сприймають одну і ту саму величину економічного ризику, оскільки мають різний світогляд, матеріальне становище та психологічне сприйняття ситуації [55]. Об'єктивна природа ризику пов'язана з тим, що ризик не з'являється нізвідки, а проявляється як наслідок різноманітності природних, соціальних і технологічних взаємозв'язків, в які вступають об'єкти, відображає реальну ситуацію в суспільстві та віддзеркалює природний стан речей. Можна зробити висновок, що для всебічного оцінювання ризику і врахування всіх його характеристик та особливостей слід використовувати раціональний підхід, який базується на оцінюванні ризиків через ймовірність його настання та можливі втрати.

Математичне оцінювання величини ризику R виконується за формулою:

$$R = P \otimes U, \quad (1.1)$$

де P – ймовірність настання несприятливої події; U – її наслідки (втрати або прибуток); \otimes – певним чином визначена композиція, яка найчастіше інтерпретується як добуток [1].

1.1.2. Ризики фінансових систем: поняття і визначення

Визначення. *Фінансова система* (ФС), як сукупність відокремлених, але взаємопов'язаних сфер і ланок фінансових відносин, може бути визначена на глобальному, регіональному або рівні компанії і фактично складається з набору процедур, що супроводжують фінансову діяльність. Інколи фінансову систему визначають як систему, яка охоплює фінансові операції та обмін грошима між

інвесторами, кредиторами та позичальниками [31]. Реальні фінансові системи складаються зі складних відносин і складних моделей, які описують фінансові послуги, інститути та ринки, що пов'язують вкладників з інвесторами.

Фінансова система компанії – це набір реалізованих процедур, які відстежують її фінансову діяльність. Багатокомпонентні складові формують фінансову систему різних рівнів. У межах компанії ФС охоплює всі аспекти фінансів, зокрема облікові заходи, графіки доходів і витрат, заробітну плату та перевірку балансу [141]. Регіональні ФС включають банки та інші фінансові установи, фінансові ринки, фінансові послуги [151]. Світова ФС – це більш широка регіональна система, яка охоплює всі фінансові установи, позичальників і кредиторів у рамках глобальної економіки [31], та включає Міжнародний валютний фонд, Центральні Банки, Світовий Банк і великі банки, які практикують міждержавне кредитування [171, 188, 201].

Фінансові системи суворо регулюються, оскільки вони безпосередньо впливають на фінансові ринки. Стабільність фінансових ринків відіграє вирішальну роль у монетарному захисті споживачів. Ці ФС в основному керуються фінансовими установами, які включають комерційні, центральні, державні та кооперативні банки [67, 136]. ФС існують не лише в банківських фінансових установах, а і установах, що надають ринкові послуги, послуги інвестування та страхування ризиків, тощо [49, 56]. Торгівельні компанії також вважаються небанківськими фінансовими установами, які мають фінансові системи. Крім фінансових установ та ринків, ФС також проявляються у фінансових інструментах, що включають грошові (кредити, депозити, цінні папери, облігації, векселі, ф'ючерси, опціони, варанти, факторинг, франчайзинг, лізинг та інші) та похідні еквіваленти.

Умовно ризики фінансових систем можна розділити на три категорії: бізнес-ризик, фінансові і операційні ризики [9, 11, 28, 29, 39, 64, 83, 149, 193].

Визначення. *Фінансові ризики (ФР)* це ймовірність виникнення фінансових наслідків у ситуації невизначеності умов здійснення фінансової діяльності [119, 228, 229]. Кількісною характеристикою ризику найчастіше є

його ймовірнісний та вартісний показники. Остаточними наслідками реалізації ризику можуть бути як втрати, так і доходи [126, 183, 187].

При оцінюванні і прогнозуванні ФР слід враховувати, що фінансові дані характеризуються надзвичайно великою кількістю величин, які надходять з різних джерел і у різноманітному вигляді, отже виникає потреба у репозитарії для їх збереження і достатньо складних інструментах для їх обробки. Виникає необхідність автоматизації консолідації та аналізу даних і візуалізації результатів. Фінансові дані можна визначити за такими характеристиками, як періодичність, однотипність, вони можуть містити пряме відношення або опис певного фінансового процесу, а також потребують ретельного збору, перевірки та прогнозування [104, 118].

1.1.3. Загальна характеристика процесу аналізу та опрацювання ризику

Фінансові ризики за своєю природою є ризиками несистемними, вони пов'язані з цілою сукупністю факторів, які спричиняють їх появу, їм притаманні неформалізованість і складність оцінювання та урахування, відсутність єдиної методології опрацювання, тощо. Системний підхід [132] має стати базою активного опрацювання ризику, а не пасивною, безпорадною реакцією на ризик, що приносить збитки, оптимізація яких не піддається управлінню.

Ефективність систем опрацювання ризику визначається тим, наскільки оптимально і правильно, для конкретного оточення і процесів аналізу та управління ризиками, визначено рівень прийнятності ризику [117]. Якщо цей показник має точну кількісну характеристику, то можна вважати, що система управління ризиком має досить фундаментальну числову основу, яку необхідно якісно і своєчасно актуалізувати і обґрунтовано розвивати відповідно до цілей компанії та наявних ресурсів [31, 41, 45, 63, 112, 157, 165, 179].

Для коректної розробки методології аналізу та опрацювання ризиків необхідно представити всю процедуру менеджменту ризику через основні етапи: виявлення, аналіз та мінімізацію ризику (рис. 1.2) [112].

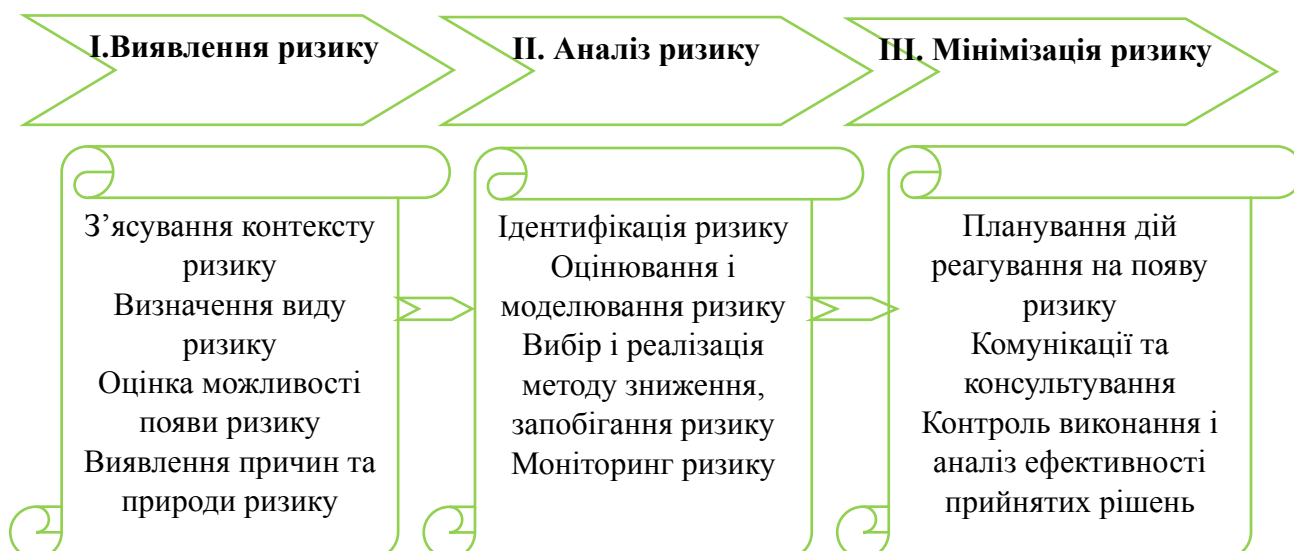


Рис. 1.2 Процес менеджменту ФР

Виявлення ризику полягає у визначенні того, до яких видів ризику найбільш уразливий об'єкт аналізу. Для цього необхідно розглядати проблему ризику в цілому, з урахуванням всіх факторів, що впливають на нього [35, 53].

Оцінювання ризику – це кількісне визначення витрат, пов'язаних з видами ризику, які були виявлені на першому етапі управління ризиком, шляхом побудови економіко-математичних моделей і проведення відповідних розрахунків [60, 79, 80, 91, 96, 98].

Оцінювання ФР зазвичай проводиться на основі великих обсягів статистичних даних, з певною невизначеністю як самих даних, так і засобів їх збору та обробки. Побудова будь-яких прогнозів також висуває досить суворі вимоги щодо якості і обсягів даних [18, 22, 25, 112, 169]. Вимоги щодо швидкості, якості, правильності опрацювання ризиків у сучасному конкурентному середовищі весь час зростають, тому актуальним є створення нових засобів та способів підтримки прийняття управлінських рішень щодо опрацювання ФР, з урахуванням їх складної природи [126]. Прийняте рішення, як мінімум, має бути ефективним для ситуації, в якій воно приймається, а також, за можливості, враховувати потенційні майбутні ризики, які можуть проявитися, з урахуванням тренду розвитку процесу і прогнозу ризикової складової [25, 206].

Управління ризиком – це динамічний процес із зворотним зв'язком [26, 28, 163, 179], де прийняті рішення повинні періодично аналізуватися і переглядатися. Час іде, обставини змінюються і несуть з собою зміни: з'являються нові види ризику, нові відомості про наявні види ризику, дешевшає стратегія управління ризиком.

1.1.3.1. Рівні ризику

Опрацювання ризиків потребує чіткого визначення меж мінімізації чи усунення ризику, що можна зробити, спираючись на концепцію прийняттого ризику [1, 160], метою якої є вироблення оптимального компромісу і яка побудована на методології диференціації ризику (його величини) залежно від стадій його прояву. Розрізняють такі рівні ризику:

- *початковий рівень ризику*. Ступінь ризику, при якому не враховуються результати заходів з аналізу, ідентифікації і оцінки конкретного ризику / виду ризиків. Цей рівень є неідентифікованим, неоціненим і, отже, може бути використаний тільки як концептуальна, верхньорівнева оцінка [160];

- *оцінений рівень ризику*. Це рівень ризику з урахуванням заходів щодо ідентифікації, оцінки та аналізу (різних видів) ризику. Величина оціненого рівня ризику є реальною, в умовах поточних процесів компанії по роботі з ризиками, оцінкою рівня ризику, який є ризиком нижчого рівня, ніж початковий рівень ризику;

- *остаточний рівень ризику* – це рівень ризику з урахуванням розроблених і виконаних заходів щодо зниження початкового рівня ризику;

- *кінцевий (прийнятний) рівень ризику* – рівень ризику, який є прийнятним, для конкретних умов певної ситуації, з точки зору критеріїв ризику. Величина кінцевого рівня ризику може дорівнювати будь-якій величині з наведених вище рівнів ризиків. Важливою умовою є обрана система критеріїв ризику. Концепція прийняттого ризику [160] полягає в тому, що ризик не обов'язково повністю або частково усувати. З урахуванням інформації, необхідної для прийняття управлінського рішення за певним ризиком, можливим результатом процесу прийняття рішення може / повинно бути:

- ❖ зниження ризику до прийняттого рівня;
- ❖ ігнорування;
- ❖ прийняття (толерантність);
- ❖ ухилення.

1.1.3.2. Аналіз сучасних методів управління ризиками

Умовно, методи управління ризиками можна розділити на 4 категорії:

❖ *Методи ухилення від ризиків.* В основі методів цієї групи лежить постулат про повне виключення ризикових ситуацій з критичних процесів / проектів. Практичні рішення [28, 29] передбачають відмову від проектів і процесів, в яких задіяні ненадійні партнери, клієнти, технології, тощо. В техніках ухилення від ризиків використовуються дії, в результаті яких відповідальність за можливі збитки переноситься на третю особу – наприклад, страхування відповідальності [42].

❖ *Методи локалізації ризиків,* які базуються на ідентифікації критичних «місць» в процесах і активностях, де спостерігається найбільш ймовірний прояв ризику [112]. Надалі такі «точки» процесів виділяються в окрему діяльність, яка в подальшому потребує високого ступеня контролю і управління. Методи локалізації застосовуються тоді, коли на основі кількісних методів аналізу ризиків вдається конкретно виявити всі можливі джерела ризиків [112].

❖ *Методи диверсифікації ризиків.* В основі знаходиться розподіл основних цінностей, вироблених процесами організації, між множиною дій, які не повинні бути прямо пов'язані між собою [29]. Таким чином, загальна цінність має бути розподілена між множиною процесів, що дозволить контролювати загальну «вартість» і знизити ризики її втрати [112].

Методи диверсифікації є найбільш поширеними засобами зниження сумарної величини можливого збитку і застосовуються для зменшення наслідків випадкових видів погано ідентифікованих ризиків [126]. Якщо мова йде про системний ризик, то застосування методу диверсифікації є

несистемним і не може бути частиною процесу, а буде тільки тимчасовим рішенням щодо ризикової ситуації, яка виникла.

❖ *Методи компенсації ризиків* спрямовані на попередження небезпеки превентивними способами. До них, наприклад, потрібно віднести стратегічне планування діяльності підприємства в цілому як засіб компенсації ризику. В ході планування / перепланування виявляється багато існуючих в організації невизначеностей, стає можливим передбачити появу багатьох «вузьких місць» в бізнес-процесах [112]. Так само до даних методів можна віднести процедури прогнозування використання інноваційного програмного забезпечення, та інші процедури, спрямовані на модернізацію та інновації в усталених бізнес-процесах компанії [29, 177].

Процеси управління ризиками мають бути спланованими і постійно актуалізуватися, контролюватися і вдосконалюватися на кожній стадії свого виконання, з урахуванням конкретних умов зовнішнього і внутрішнього середовища [112].

Особливістю процесів опрацювання ризиків є ітеративність, інтерактивність, комплексність, потреба в опрацюванні великих обсягів даних, тому надзвичайно актуальною є розробка автоматизованих систем опрацювання ризиків [23,117,124].

Для автоматизації процесів опрацювання ризиків можуть застосовуватись мультифункціональні рішення та конкретні «вузькі» галузеві програмні засоби. Умовою для автоматизованої системи є єдине інформаційне середовище [112]. Воно має забезпечити найбільш ефективний контроль всіх процесів, необхідних для обробки ризику, вибору шляху його обліку, якщо він ідентифікований і оцінений, а також напрацювання і впровадження механізму моніторингу [124].

Постійний моніторинг та удосконалення наявних процесів щодо їх відповідності цілям діяльності компанії є обов'язковим етапом ризик-менеджменту [6, 7, 165]. Ця процедура повинна так само виконуватися в єдиній інформаційній системі, що може прискорити виконання ризикових процесів і знизити ймовірність можливих помилок.

1.1.3.3. Принципи створення ефективних систем управління ризиками

Ефективність системи управління ризиками може досягатися за рахунок використання наступних принципів [29]:

- максимально можливе охоплення системою сукупності ризиків, обґрунтованої цілями діяльності підприємства, та ресурсами, виділеними на її функціонування;
- мінімізація, ухилення або прийняття ризику, обґрунтоване для кожного конкретного випадку [126];
- адекватна реакція на причини і можливі наслідки ризиків;
- використання сучасних і кращих практик при роботі з / над ризиками;
- моніторинг і контроль процесів обробки ризиків [126] і подальшого їх супроводу, засновані на числових індикаторах і метриках, застосування яких дозволяє фахівцям по роботі з ризиками оперативно реагувати на проблеми;
- залучення до процесу управління ризиками співробітників компанії, відповідальних за певні функціональні напрямки, що знаходяться на різних структурних рівнях, що дозволяє охопити комунікації компанії в цілому і контролювати можливі ризикові «вогнища»;
- автоматизація процедур контролю з метою зниження впливу «людського фактору» і трудовитрат на виконання процедур з аналізу та управління ризиками [165];
- мінімізація часу, витраченого на повний цикл по роботі з ризиками (від виявлення ризику до прийняття рішення);
- надання адекватної звітності, що дозволяє оцінити процес ризик-менеджменту і його ефективність;
- удосконалення вироблених раніше процедур за рахунок організації безперервного аналізу проведених робіт і валідації результатів діяльності з управління та аналізу ризиків [165].

Наведений список процедур претендує на достатність і повноту дій, дотримання яких може гарантувати ефективність конкретної системи з управління та аналізу ризиків [142].

1.1.4. Ризик-менеджмент на основі існуючих міжнародних та національних стандартів для фінансових систем

Сьогодні основні підходи до менеджменту ризиків формуються такими організаціями та стандартами [295, 312, 328]:

✓ **FERMA** (Federation of European Risk Management Association). Європейська федерація асоціації ризик-менеджменту – в її документах запропоновано модель ідентифікації подій.

✓ **ERM COSO** (Enterprise Risk Management – Integrated Framework Committee of Sponsoring Organizations of the Treadway Commission) – принципи ризик-менеджменту, розроблені Комітетом спонсорських організацій Комісії Тредвея разом з компанією PricewaterhouseCoopers [126].

✓ **ISO/IEC Guide 73** – стандарт, розроблений Міжнародною організацією по стандартизації (International Organization of Standardization), у якому описано системний підхід до оцінки і управління ризиками.

✓ **PMBOK** (Project Management Body of Knowledge) – стандарт управління проектами, сформований Американським інститутом управління проектами (Project Management Institute, PMI) [294]. Описує всі стадії життєвого циклу проекту, включаючи елементи проектного ризик-менеджменту.

✓ **Basel II** – Угода комітету по банківському нагляду [212], містить нормативи покращення техніки оцінки кредитних ризиків і управління ними.

✓ **Basel III** – Стандарти Базельського комітету по банківському нагляду вимірювання капіталу банків і доопрацьована угода (2010-2011), яка посилює вимоги до капіталу банку і вводить нові нормативні вимоги по ліквідності для підвищення якості управління ризиками [211].

Найчастіше аналізом ФР займаються відділи ризик-менеджменту банків та фінансових установ, які спираються на основні положення Базельської угоди (Basel II та Basel III) [186, 211, 212, 280, 304, 358]. Основною метою є виявлення і прогнозування ризиків, які можуть з'явитися в організації, класифікація ризику, визначення його рівня та рейтингу. Значення критичного рівня ризику

[186] може визначатися згідно встановленої політики банку або Базельської угоди.

Класифікацію фінансових установ щодо рівня обслуговування їх зобов'язань та фінансового стану, зокрема виконують такі міжнародні рейтингові організації як Standard & Poor's, Moody's, Fitch тощо (табл. 1.1).

Таблиця 1.1

Класифікація ризиків юридичних осіб

КОРПОРАЦІЇ	ING new	S & P	Moody's	PD Range	Average PD (інтерполяція)	Average PD (основний клас)	Рекомендований для оцінювання ринкової вартості	
PERFORMING							NO (GLOBAL) INVESTMENT GRADES applicable	
	1	AAA	Aaa	<0,01	0,00	0,00		
	2	AA+	Aa1	0,01 - 0,02	0,01	0,01		
	3	AA	Aa2	0,02 - 0,04	0,02			
	4	AA-	Aa3	0,04 - 0,05	0,04			
	5	A+	A1	0,05 - 0,06	0,05	0,05		
	6	A	A2	0,06 - 0,08	0,07			
	7	A-	A3	0,08 - 0,11	0,09			
	8	BBB+	Baa1	0,11 - 0,17	0,15	0,37	A	0,318%
	9	BBB	Baa2	0,17 - 0,29	0,25			
	10	BBB-	Baa3	0,29 - 0,51	0,42			
	11	BB+	Ba1	0,51 - 0,89	0,71	1,45	BBB	0,504%
	12	BB	Ba2	0,89 - 1,54	1,19		BBB	1,005%
	13	BB-	Ba3	1,54 - 2,67	2,00		BBB	2,004%
	14	B+	B1	2,67 - 4,62	3,35	6,59	CCC+	3,175%
	15	B	B2	4,62 - 8,01	5,63		CCC	3,997%
							CC+	6,332%
CC							7,969%	
16	B-	B3	8,01 - 13,88	9,45	C+	12,625%		
Special Mention (спеціальна оцінка)	17	CCC	Caа1	13,88 - 26,63	15,89	15,89	C	15,890%
NON – Sub-Standard	18	CC	Caа2	26,63 - 44,7	34,14	34,14	D	34,14
PERF Doubtful	19	C	Caа3	44,70 - 100	44,70	44,70		44,70
ORM Liquidation – no Loss	20	D	D	100	100,00	100,00		100,00
ING Liquidation – with Loss	21	D	D	100	100,00	100,00		100,00

Стандартна міжнародна класифікація передбачає існування 22 категорій (одна категорія – завжди визначається як NR – no rating, тобто неможливість застосування рейтингу з об'єктивних чи суб'єктивних причин). За методологією RAROC (risk-adjusted return on capital) [212] необхідно визначити та співставити класифікацію і рейтинги, запропоновані міжнародними компаніями (S&P, Moody's) та вітчизняною класифікацією [139, 140], що складається з 12 категорій (11 та категорією NR – без встановлення рейтингу), наведеною у таблиці 1.2.

Таблиця 1.2

Українська класифікація компаній з точки зору інвестиційних ризиків

Рейтинг	Опис
1	Високоякісні корпорації, ймовірність виплати якими процентів та основної суми є вищою, ніж в уряді країни. Ймовірність впливу кризи у країні на ці компанії є незначною.
2	Найкращі корпорації та компанії з державними гарантіями. Ймовірність виплати процентів та основної суми є високою. Вони іноді є вразливими перед несприятливими змінами в діловому, економічному та зовнішньому середовищі.
3	Успішні компанії та кращі з середніх компаній. На них, ймовірно, вплинуть несприятливі зміни в економіці.
4	Ймовірність виплати процентів та основної суми є середньою. Несприятливі зміни в ділових, економічних та інших зовнішніх умовах неминуче призведуть до нездатності або неготовності компанії сплачувати проценти та погашати основну суму.
5	Існує певна ймовірність виплати процентів та основної суми.
6	Ймовірність виплати процентів та основної суми є пониженою.
7	Особливі позичальники. Виплата процентів та основної суми залежатиме від того, наскільки сприятливими будуть економічні умови. Цих позичальників необхідно контролювати раз на півроку.
8	Особливі позичальники. Виплата процентів та основної суми залежатиме від того, наскільки сприятливими будуть економічні умови. Цих позичальників необхідно контролювати раз на квартал.
9	Неякісні позичальники. Скоріш за все, не сплачуватимуть ні проценти, ні основну суму. Кредити їм, як правило, не надаються. Позичальників, що належать до цієї категорії, доводиться примушувати погашати кредити.
10	Неякісні позичальники. Скоріш за все, не внесуть навіть перший платіж на погашення процентів основної суми. Нові кредити таким позичальникам надаються лише за виняткових обставин.
11	У компанії дефолт. Вона не сплачує проценти та основну суму протягом більш ніж 3 місяці.

Слід зазначити, що розглянуті методології здебільшого розроблені для стрес-тестування під час оцінювання ризиків і зовсім не ставлять перед собою

завдання виявлення першопричин ризику для запобігання навіть їх появі [126]. Ці методології є досить суб'єктивними, і невідомі вхідні параметри та критерії, за якими насправді здійснюється оцінювання, неможливо здійснити перевірку коректності виставлених оцінок – рейтинги цінуються лише завдяки авторитету організацій, які їх формують.

Крім наведеної вітчизняної методології, всі інші методології пропонуються іноземними організаціями, а тому навіть з урахуванням специфіки країни та адаптації розрахункових показників не можуть гарантовано давати об'єктивний результат [126]. Отже дедалі актуальнішою стає не лише адаптація існуючої методології оцінювання ФР до українських реалій, а й розробка власної системної методології.

1.1.5. Формалізація рейтингу та рівня ризику

Для компанії можуть визначатися різні рівні критичного ризику. Наприклад, якщо компанія знаходиться на рівні AAA, то критичним рівнем для такого підприємства буде можливість переходу/ів в іншу/і категорію/і (на клас нижче), тобто опускання рейтингу компанії до рівня BBB+, тоді:

$$R_{кр(AAA)} = P(D_{BBB+}), \quad (1.2)$$

де імовірність дефолту компанії для даної категорії BBB+ визначається в межах $P(D_{BBB+}) \in [0,11;0,17]\%$ згідно описаної вище класифікації в таблиці 1.1. При такій постановці задачі можна визначити (формалізувати) категорії та рейтинг ризику протягом певного періоду часу (кількість місяців) $0 < t < 12$, тобто протягом року. Для первинної категорії (класифікації) компанії або підприємства чи банку ризик вважається *прийнятним*, якщо:

$$R_{пр} < R_{кр}, \quad (1.3)$$

де $R_{пр}$ – прийнятний ризик для даної категорії [126], $R_{кр}$ – критичний ризик, тобто ліва межа ймовірності дефолту, що характеризує наступний (нижчий) рівень категорії або класифікації в таблиці за стандартними оцінками рейтингів (рівнів) компанії.

Перехід компаній з одного класу (категорії або рівня) в інший за один рік може бути представлений у вигляді матриці міграції, що ілюструє (рис. 1.3) [126], який відсоток компаній з кожної категорії протягом одного року змінює свій рейтинг, підвищуючи його.

	AAA	AA	A	BBB	BB	B	CCC	D	NR
AAA	93,37	4,82	0,00	0,00	0,00	0,00	0,00	0,00	1,81
AA	0,74	90,59	3,51	0,92	0,00	0,37	0,00	0,00	3,88
A	0,00	1,88	89,56	3,95	0,19	0,56	0,00	0,00	3,86
BBB	0,00	0,27	3,88	86,50	2,67	0,94	0,13	0,13	5,48
BB	0,00	0,00	0,19	8,97	75,95	4,77	0,00	0,38	9,73
B	0,00	0,00	0,67	0,45	7,62	73,32	2,69	3,36	11,88
CCC	0,00	0,00	0,00	0,00	0,00	14,82	51,85	11,11	22,22

Рис. 1.3 Матриця переходів компанії (міграцій)

Для банків допустимий ризик – це той рівень ризику, на який банк погоджується для досягнення мети своєї діяльності (толерантність до ризику) та виконання його стратегічних завдань в рамках стандартного плану розвитку ситуації і можливих стресових ситуацій [62, 87, 139,140].

Кількісна оцінка толерантності до ризику банку визначається як:

$$\tau = \frac{2 * (\bar{X} - X_{rf}) * \sigma_X^2}{(\bar{X} - X_{rf})^2},$$

де τ – кількісна оцінка толерантності ризику суб'єкта господарювання;
 \bar{X} – очікувана величина цільового показника (математичне сподівання);
 X_{rf} – безризикове абсолютне значення цільового показника X ; σ_X^2 – дисперсія цільового показника X .

Встановлені такі ранжування толерантності до ризику [139]:

- ❖ мінімальний обсяг прийнятого ризику – фактична толерантність в інтервалі 0 – 50% регулятивного капіталу;
- ❖ нормальний обсяг прийнятого ризику – фактична толерантність в інтервалі 51 – 90%;

- ❖ збільшений обсяг прийнятого ризику – фактична толерантність в інтервалі 91 – 94%;
- ❖ великий обсяг прийнятого ризику – фактична толерантність в інтервалі 95 – 97%;
- ❖ кризовий – фактична толерантність більше 98% [126].

Існуючі методики оцінювання рівня ризику пропонують різні трактування формули (1.1) для оцінки величини ризику: в одних на перший план висувається імовірнісна складова ризику, а в інших – значення збитку [1]. Спираючись на концепцію «прийнятого ризику», можна визначити, що:

$$\text{для } \forall U \exists P: P^*(U) = P. \quad (1.4)$$

В загальному вигляді шукана залежність може бути представлена як:

$$P^* = P^*(U),$$

де $P^*(U)$ – монотонно незростаюча функція, яка відображає прийнятну ймовірність виникнення несприятливої події, яка призводить до збитку U ;

$U = \frac{U^*}{U_{кр}}$ – нормоване значення збитку [126]; $U^* \in [U_{нз}, U_{кр}]$, $U_{нз}$ – втрати, які не

є значимими для особи, що приймає рішення (ОПР) ($P^*(U_{нз})$ може бути близькою до 1), $U_{кр}$ – критичний (максимально прийнятний) для ОПР рівень збитків, ймовірність виникнення якого має прямувати до нуля: $P^*(U_{кр})=0$ [126].

$\bar{U} = \frac{U_{нз}}{U_{кр}}$ – нормоване значення незначних збитків (втрат).

Як P^* може бути використана функція виду [1]:

$$P = a \cdot e^{-b(U - \bar{U}_{нз})}, \quad (1.5)$$

де a і b – деякі константи: a – відповідає ймовірності прийняття незначного збитку $\bar{U}_{нз}$; b – характеризує швидкість падіння допустимої ймовірності нанесення збитку. Функція толерантного ризику (тобто того, який

ми приймаємо) проходить вище кривої прийнятного ризику (рис. 1.4), а тому рівняння даної кривої має вигляд [1]:

$$P^T = \beta \cdot P^*(U), \quad (1.6)$$

де β – деяка константа; а прийнятний ризик

$$R^{np} = \frac{I}{I - \bar{U}^{n3}} \cdot \int_{\bar{U}^{n3}}^I P^*(U) dU. \quad (1.7)$$

Підставивши (1.5) у (1.7) і проінтегрувавши, для певного рівня прийнятного ризику отримаємо:

$$R^{np} = \frac{a}{b(I - \bar{U}^{n3})} \cdot [I - e^{-b \cdot (I - \bar{U}^{n3})}]. \quad (1.8)$$

Відповідно, рівень толерантного ризику може бути знайдений за виразом:

$$R^T = \beta \cdot R^{np}. \quad (1.9)$$

Тоді область прийнятного ризику може бути побудована наступним чином: ОПР визначає скінченну множину значень $\hat{R}^* = \{(U_i; P_i^*)\}$, значення U_i нормуються за значенням $U_{кр}$ [1]; задані в опорних точках U_i ймовірності P_i^* апроксимуються неперервною функцією виду (1.5).

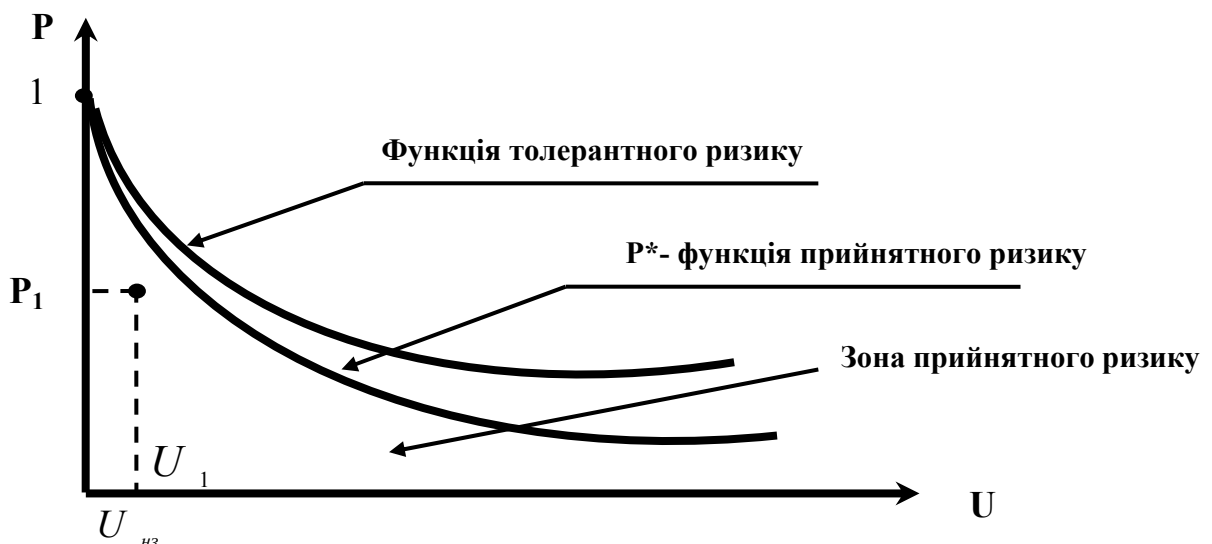


Рис. 1.4 Функції прийнятного і толерантного ризику

Визначив і формалізував поняття толерантного ризику, тобто такого, що може на себе прийняти організація з урахуванням можливих втрат, будемо розуміти, що система ризик-менеджменту має розроблятися в організації для

визначення також критичного фактору ризику, тобто ризику, збитки за яким є гранично допустимими для компанії. Далі виникає необхідність у формуванні єдиної системної методології, яка буде об'єднувати всі описані фактори, категорії та методи оцінювання ризику в єдину систему підтримки прийняття рішень та буде застосовувати відповідні критерії для перевірки як якості вхідних даних, так і вихідних рівнів та рейтингів фінансового ризику, чому й присвячена дисертаційна робота.

1.2. Причини виникнення і класифікація ризиків

Відмінність точок зору і ступеня деталізації веде до визначення як завгодно великої кількості видів ризиків. Класифікувати ризики у фінансових системах важко внаслідок тісного взаємозв'язку і заміщення ризиків. Ризики господарської і інвестиційної діяльності характеризуються великою різноманітністю, і з метою ефективного управління ними доцільно класифікувати їх за різними ознаками. Класифікація ризиків дозволяє чіткіше структурувати проблеми і впливає на аналіз ситуацій і вибір методу ефективного управління діяльністю організації [27, 28, 38, 45, 50, 64].

Складність класифікації ризиків пояснюється не тільки їх розмаїттям, а також появою нових видів ризиків внаслідок економічного, соціального, технологічного розвитку сучасного світу. Основні припущення класифікації ризиків, запропоновані у [157], передбачають застосування системно-цільового підходу, розрізнення ситуацій невизначеності і ризиків, відповідність рівнів деталізації цілям класифікації, урахування характерних ознак, таких як джерело ризику, об'єкт, що несе ризик, суб'єкт, що сприймає ризик.

1.2.1. Види фінансових ризиків

Фінансові ризики класифікують за сукупністю інструментів, що досліджуються (індивідуальний та портфельний), за об'єктом, що характеризується (ризик окремої фінансової операції, різних видів фінансової діяльності), за джерелом виникнення (зовнішній, систематичний або ринковий

ризик та внутрішній, не систематичний або специфічний ризик). Ризики можуть розрізнятися і за фінансовими наслідками (ризик, що спричиняє лише економічні втрати, ризик втраченої вигоди, ризик, що спричиняє як економічні втрати, так і додаткові доходи) [133, 193], і за характером прояву у часі (постійний та тимчасовий ФР), і за рівнем фінансових втрат (допустимий, критичний та катастрофічний) [126], і за можливістю передбачення (прогнозований та непрогнозований), за можливістю страхування [2, 8, 173]. Детальніше побудова класифікацій ризиків за різними критеріями описана у додатку А.1.

Види ризиків залежать від виду діяльності, частина з них може бути об'єднана в один вид, інші – взагалі не розглядатимуться. Наприклад, якщо підприємство веде економічну діяльність лише всередині країни і не працює з валютою, то для нього немає сенсу розглядати валютний ризик; а, якщо воно не проводить інвестиційну та кредитну політику, ці види ризиків також немає сенсу розглядати.

Фінансовий ризик може складатися з різновидів систематичних і несистематичних ризиків [58, 133, 193]. Види ризиків та їх основні характеристики наведені у таблиці 1.3.

При дослідженні фінансових ризиків підприємств було встановлено [107], що інколи доцільно об'єднувати ризик зниження фінансової стабільності та ризик неплатоспроможності (а інколи і кредитний ризик [37]). Це пов'язано з подібністю характеристик для оцінювання цих ризиків. Зокрема, використання фактор-листів фінансової діяльності, звітності підприємств. Ці показники відбивають структуру капіталу підприємства, наявність у нього оборотних активів, показники ліквідності, активів, прибутку, виручки тощо.

Аналіз фінансових ризиків проводиться на основі статистичних даних, в інформаційному середовищі обробки даних, за допомогою певних інформаційних, програмних, алгоритмічних засобів, і його результати та прийняті рішення суттєво залежать від різноманітних чинників впливу на якість інформації [47, 65, 105, 190].

Таблиця 1.3

Види фінансових ризиків

Вид ризику	Характеристика
Інфляційний	обезцінення реальної вартості національної валюти, очікуваних доходів в умовах інфляції;
процентний ризик	непередбачена зміна процентної ставки;
системний фінансовий ризик	невдала диверсифікація, що викликає підвищення загальної ненадійності;
кредитний ризик	ризик неплатежу, несвоєчасного розрахунку, перевищення бюджету;
криміногенний ризик	– фіктивне банкрутство; – рейдерство; – підробка документів; – викрадення активів; – корупція;
структурний ризик	неефективні поточні витрати;
податковий ризик	– введення нових видів податків і зборів; – збільшення діючих податків і зборів; – зміна строків і умов платежів; – відміна податкових пільг;
валютний ризик	недоотримання доходів через зміну курсу валют;
депозитний ризик	неповернення депозитних вкладів;
інші види	– ризики стихійних лих; – «форс-мажор»; – емісійний ризик; – тощо.

1.2.2. Аналіз впливу інформаційних ризиків на виникнення та результати оцінювання фінансових ризиків

Інформаційні ризики – це небезпека витоку, руйнування та втрати необхідної, використання необ'єктивної інформації для прийняття рішень, а також поширення невігідної, негативної, небезпечної для суб'єкта інформації, що може завдати йому збитки, матеріальну або моральну шкоду [94, 150].

Під інформаційними ризиками розуміють загрозу виникнення втрат чи збитків в результаті використання інформаційних технологій. Інформаційні ризики пов'язані зі створенням, передачею, збереженням і використанням інформації. Якщо причини виникнення інформаційного ризику всередині підприємства, то ризики вважають внутрішніми; зовнішніми інформаційними ризиками є ризики, які виникають внаслідок дії зовнішніх факторів.

Інформаційні ризики слід розглядати і враховувати як економічні (майнові, виробничі, фінансові) [65]. Приклад класифікації інформаційних ризиків у банківській діяльності наведено на рис. 1.5.

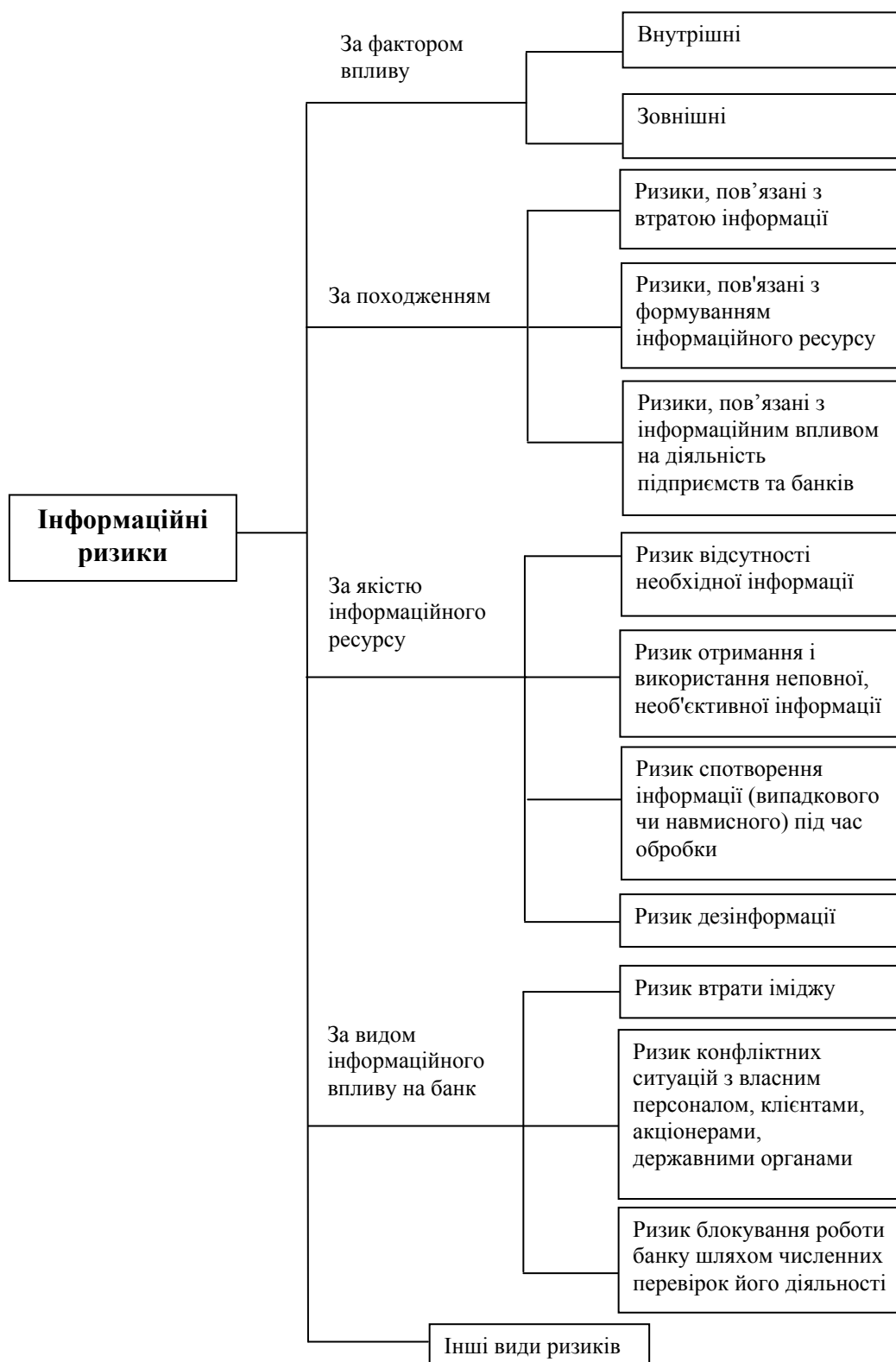


Рис. 1.5 Класифікація інформаційних ризиків у банківській діяльності

Інформаційні ризики за своїм походженням поділяються на три категорії:

- ризики, пов'язані з втратою (витоком, руйнуванням, знищенням) інформації. Особливо небезпечним є ризик втрати інформації, яка становить банківську таємницю, або є інформацією з обмеженим доступом;
- ризики, пов'язані з формуванням інформаційного ресурсу (використання неповної, неправдивої інформації, дезінформація);
- ризики, пов'язані з інформаційним впливом на діяльність фінансових систем (поширення неправдивої інформації, інформаційно-психологічний вплив, інформаційний тероризм) [65].

Інформаційний вплив – це використання спеціальних ІТ для зміни поведінки осіб, об'єктів, діяльності. Ризики інформаційного впливу можуть бути постійними або виникати в результаті цілеспрямованої дії. Найхарактернішими є інформаційні атаки, коли майже одночасно розсилається негативна інформація про банк, що може призвести до втрати клієнтів [94].

Сутність *інформаційного ризику* можна розглядати як випадкову подію [65], яка призводить до порушення функціонування інформаційної системи, зниження якості інформації нижче допустимого рівня, і в результаті під впливом цих наслідків підприємство зазнає значних втрат.

Залежно від властивостей та особливостей даних для фінансових систем, використовуються різні інформаційні технології, яким властиві [96]:

- *Висока динамічність.* Інформаційні технології постійно оновлюються, бізнес-процеси автоматизуються, з'являються нові ризики інформаційної природи.
- *Висока комплексність взаємозв'язків у ІТ-середовищі.* На підприємствах створюються інформаційно-комунікаційні мережі, бази даних і знань, вводиться електронний документообіг, дані з різних систем інтегруються в електронних інформаційних ресурсах. Внаслідок тісних причинно-наслідкових зв'язків між окремими показниками та даними під дією окремого ризикового фактору можливий «ефект доміно» і втрата інформаційних ресурсів в зовсім інших частинах системи.

Класичний алгоритм, що дозволяє оцінити кількісно ризик інформаційних втрат, був розроблений ще у 1974 році, і передбачав обчислення за формулою [47]:

$$AV \times EF \times ARO = ALE,$$

де AV (Asset Value) – вартість ресурсу; EF (Exposure Factor) – міра уразливості ресурсу до загрози; ARO (Annual Rate of Occurrence) – оцінка ймовірності реалізації загрози [96]; ALE (Annual Lost Exposure) – підсумкові очікувані втрати від конкретної загрози за певний період часу.

1.3. Аналіз застосування економіко-математичних методів і моделей в задачах ідентифікації і оцінювання ризиків фінансових систем

Сучасним фінансовим процесам притаманні висока динаміка, нестационарність та нелінійність, велика та змінна у часі волатильність, наявність детермінованих і випадкових компонент у часових рядах даних [217, 341, 361]. Як правило, фінансові процеси функціонують під впливом множини випадкових збурень різної природи (шумові складові), які вносять в аналіз даних суттєві невизначеності. Два найбільш важливих аспекти, які характеризують ризики – по-перше, волатильність, або змінність фінансових індикаторів, імовірність або частота подій, і, по-друге, чутливість (exposure) критеріїв діяльності до їх наслідків.

Якісні методи оцінювання ризиків застосовуються для визначення виду ризику і виділення тих ризиків, які вимагають швидкого реагування і є найбільш суттєвими для фінансових систем [43]. Найчастіше для якісного оцінювання використовуються метод дерев рішень (дозволяє визначити кінцеву кількість варіантів розвитку подій, встановити ймовірність їх реалізації, визначити якісні і кількісні характеристики ризику для кожного варіанту) та метод аналізу сценаріїв [76], що розглядає чутливість критерію чистої вартості (NPV) до змін ключових змінних і діапазон їх імовірнісних значень [123, 233].

Кількісні методи оцінювання ризиків дозволяють визначити ймовірність виникнення ризику і наслідків впливу ризику на діяльність компанії. Серед

основних методів кількісної оцінки ризиків – ймовірнісні методи, теоретико-ігрові, аналіз точки беззбитковості, імітаційна модель Д. Хертца, метод еквіваленту, метод оцінки доходності тощо [43].

Для аналізу і дослідження ФР застосовуються статичні та ймовірнісні моделі, побудовані на базі мереж Байєса (МБ або БМ), регресії, нейронних мереж, тощо. МБ дають можливість відобразити в моделі виявлені причинно-наслідкові зв'язки між різними чинниками ризику і змінами середовища [9] і, на відміну від регресійних моделей, дозволяють враховувати не лише безпосередні залежності рівня ризику від факторів ризику, а також і залежності між факторами ризику. МБ дозволяють опрацювати неповні дані.

У МБ є можливість комбінації експертного знання (наприклад, для побудови мережі шляхом визначення залежностей між змінними), і ймовірнісних методів опрацювання статистичних даних. У подальшому буде показано застосування МБ до моделювання ризиків різних видів.

Задачам моделювання і прогнозування, дослідженню волатильності присвячено велику кількість наукових робіт, зокрема [247, 259, 272, 318], але проблема отримання високоякісних прогнозів цього важливого для прийняття рішень статистичного параметра потребує виконання подальших досліджень, оскільки не всі існуючі типи моделей волатильності дають можливість досягти необхідного ступеня адекватності процесу, а особливості протікання фінансових процесів потребують подальшого удосконалення структур математичних моделей та методів оцінювання їх параметрів.

1.3.1. Особливості задач оцінювання і прогнозування фінансових ризиків

Вирішення задачі оцінювання і прогнозування фінансових ризиків потребує розуміння фінансового процесу, побудови відповідної моделі процесу, визначення її статистичних характеристик і використання для прогнозування ймовірності настання ризику та можливих втрат. Суттєвою складністю є значна волатильність фінансових процесів, їх нестационарність та нелінійність,

наявність трендів та сезонних ефектів. Тому побудова такої моделі є досить складною ітеративною процедурою.

Прогнозування розвитку процесів різної природи – одна із поширених наукових задач, яка має практичне застосування у системах підтримки прийняття рішень, автоматичному керуванні технічними системами і технологічними процесами, менеджменті бізнесу та ризиків, формуванні стратегій управління підприємствами тощо [110, 339, 362]. Часто об'єктом прогнозування є волатильність (ступінь мінливості), яку вимірюють стандартним відхиленням фінансового процесу. Волатильність використовують як один із основних параметрів при оцінюванні можливих втрат у ризик-менеджменті, при виконанні торгових операцій і т. ін. Розглядаються лінійні та нелінійні процеси зі змінною волатильністю, що характеризують поведінку, наприклад, валютного ринку. Найчастіше аналізуються моделі стохастичної волатильності (МСВ) [268, 356], авторегресія з умовною гетероскедастичністю (АРУГ) [259], узагальнена модель авторегресії з умовною гетероскедастичністю (УАРУГ або GARCH – у англійській термінології) [224], експоненційна УАРУГ (Е-УАРУГ) [318], частково інтегрована узагальнена АРУГ (ЧІАРУГ) [209], частково інтегрована експоненціальна узагальнена АРУГ (ЧІЕУАРУГ) [225], частково інтегрована авторегресія з ковзним середнім [266, 275]. Докладно побудови описаних моделей стохастичної волатильності, УАРУГ, Е-УАРУГ, ЧІАРУГ, ЧІЕУАРУГ, приклади їх застосування у прогнозуванні волатильності наведені у додатку А.2.

Зрозуміло, що якість прогнозних оцінок суттєво залежить від якості самих даних, від застосування адекватних методів попередньої обробки даних, спрямованих на поліпшення їх статистичних характеристик, коректного застосування процедур оцінки структури і параметрів, а також методів для генерації самих прогнозів. Одним з основних моментів у процесі розробки моделі та оцінки прогнозів є контроль відповідних обчислювальних процедур [217]. Це завдання може бути вирішене шляхом розробки і практичного застосування адаптивних систем прогнозування на основі ідей системного

аналізу, таких як ієрархічна структура системи, застосування процедури оптимізації, де це можливо, математичний опис і урахування невизначеності, генерація альтернатив для прийняття рішень і т. д. [78].

Крім того, існує можливість комбінування прогнозних оцінок, згенерованих ідеологічно різними методами, що часто приводить до кращих прогнозів, ніж окремі оцінки, за кожним конкретним підходом [52]. У будь-якому випадку такий підхід до прогнозування може привести до значного зменшення дисперсії помилок прогнозування при правильному застосуванні.

1.3.2. Методи інтелектуального аналізу даних для дослідження фінансових процесів і систем

При використанні засобів інтелектуального аналізу даних [230, 269, 271, 319, 326, 363, 366, 367] основними перевагами є наявність великого спектру методів та алгоритмів, які дозволяють обробляти великі масиви даних, виявляти наявні закономірності, моделювати процес різними підходами та визначати кращі моделі. Слід зазначити, що ефективність того чи іншого методу чи алгоритму суттєво залежить від специфіки та постановки задачі, а тому у сучасних програмних засобах задіяні не лише стандартні методи, а й власні розроблені алгоритми як комбінації кількох стратегічно різних підходів, які дозволяють вирішувати нові задачі та отримувати високі оцінки якості моделей та прогнозів.

1.3.2.1. Факторний аналіз

Факторний аналіз, як комплекс моделей і методів, застосовується у задачах ризик-менеджменту для виявлення множини основних параметрів, які характеризують певну фінансову систему з точки зору її функціонування в умовах невизначеності. Він використовується для з'ясування прихованих зв'язків між ними, скорочення кількості параметрів шляхом узагальнення і «стиснення», отримуючи таким чином значно менший обсяг структурованих даних для подальшого аналізу і визначення ступеню впливу тих чи інших факторів ризику на загальну величину ризику.

Основна задача факторного аналізу – виявити ступені впливу на залежну змінну (сумарну величину ризику) інших змінних (факторів – елементарних ризиків різної природи). Опис алгоритму побудови основної моделі факторного аналізу в задачах ризик-менеджменту наведений у [17, 26].

Часто у фінансовій звітності підприємств зустрічаються комбіновані показники (Debt2 – EBITDA/Процентні витрати, Lev1 – Валовий прибуток/Чиста виручка, Qual_str – Борг/Власний капітал тощо), які формуються співвідношенням стандартних характеристик (прибутків, боргів, співвідношення активів і пасивів). Використання таких показників дозволяє оцінити темпи росту підприємства. Найбільш трудомістким процесом є виділення з усієї множини показників фінансової діяльності найбільш статистично важливих змінних та видалення корельованих між собою [26, 107]. Використовують різноманітні засоби, оцінюють статистичну значущість, проводять однофакторний аналіз (перевіряють наявність і силу зв'язку між однією залежною (показник фінансової стабільності) і незалежною змінною (фінансовий показник), що дозволяє визначити, які змінні є найбільш точними характеристиками для моделі.

Однофакторний аналіз проводиться у декілька кроків [107]:

- визначення статистичних характеристик при групуванні на класи за принципом максимально однакової кількості спостережень у кожному класі;
- аналіз отриманих статистичних характеристик, визначення найбільш сильних предикаторів у категорії [107];
- порівняння отриманих закономірностей з фундаментальними знаннями з предметної області, аналіз динаміки показника *WOE* від класів;
- перегрупування на класи за допомогою зміни інтервалів з метою підвищення статистичної значущості.

Використання комбінованих показників потребує додаткової перевірки корельованості змінних, але саме ці показники є найбільш важливими і найчастіше включаються у модель, що прогнозує обраний вид ризику [107].

1.3.2.2. Нечіткі множини

Застосування математичного апарату нечітких множин і нечіткої логіки у задачах аналізу ризиків відповідає як змісту задач (ризик виникає як прояв певної невизначеності) [36, 71], так і рівню початкового знання щодо самої події виникнення ризику. Аналізуються події, відносно яких є лише певні прогнози, не має достатньої інформації щодо їх впливу на процеси, які мають випадковий характер. При дослідженні ризиків важливо використати експертні знання, які найчастіше мають нечіткий, гнучкий характер, щодо визначення значущості певних видів ризиків у комплексі всіх притаманних системі ризиків, провести оцінювання можливих рівнів ризиків, оцінити втрати, які можуть бути представлені не тільки у вигляді кількісних оцінок, а й лінгвістичними змінними.

Характеристикою нечіткої множини виступає функція приналежності $\mu_c(x)$ – ступінь приналежності до нечіткої множини C [220]. Для нечітких множин визначені основні логічні операції. Перетином двох нечітких множин A і B є: $\mu_{A \cap B} = \min\{\mu_A(x), \mu_B(x)\}$. Об'єднанням двох нечітких множин називають: $\mu_{A \cup B} = \max\{\mu_A(x), \mu_B(x)\}$ [267].

Для нечітких множин використовуються поняття нечіткої і лінгвістичної змінної. Нечітка змінна визначається набором (N, X, A) , де N – назва змінної, X – універсальна множина, A – нечітка множина над X [369]. Значеннями лінгвістичної змінної можуть бути нечіткі змінні. Кожна лінгвістична змінна складається з назви, множини значень (термів), універсальної множини X та синтаксичного і семантичного правил.

1.3.2.3. Нейронні мережі

Нейромережа – це статистична модель, яка складається з множини нейронів, згрупованих в шари, що створюють мережу. Кожен нейрон – це елемент із заданою одиничною функцією, який обробляє значення, які надходять до нього. Зв'язки між нейронами створюють мережу, що дозволяє визначити взаємозв'язок між окремими даними [71, 219].

Основна властивість нейронної мережі – здатність її до навчання, тобто за наявності певної кількості прикладів розв’язання задачі, нейромережа може знайти розв’язок, не маючи алгоритму рішення задачі.

Нейронні мережі часто застосовуються для прогнозування, зокрема, динаміки ринків, ситуації на фондових ринках, курсу валют, визначення курсів облігацій і акцій підприємств, а також в управлінні кредитними ризиками, у страховій діяльності (для оцінки ризиків страхування) [71, 301]. Як системи здатні до самонавчання, нейронні мережі імітують діяльність людського мозку. Основні види нейронних мереж такі: мережі зворотного розповсюдження, Хопфілда, стохастичні мережі, тощо [370]. Вони складаються з великої кількості пов’язаних між собою однотипних нейронів, що імітують нейрони головного мозку, а відрізняються їх будовою, топологією зв’язків між ними та алгоритмами навчання [36, 71, 172, 219, 220].

Далі у дисертаційній роботі буде показана практична доцільність використання нейронних мереж для оцінювання ризиків фінансових систем. Зокрема, буде запропоновано розширення їх застосування для практичного використання у нейро-нечіткому методі доповнення відхиленими заявками.

1.3.2.4. Мережі Байєса

Мережа Байєса є різновидом імовірнісних мереж, у яких отримання нових знань про ймовірності у вершинах мережі здійснюється за допомогою формули Байєса та її узагальнень [66, 323]. МБ дозволяє встановити причинно-наслідкові зв’язки між подіями та визначити ймовірності настання тієї чи іншої ситуації при отриманні нової інформації [90, 99, 125, 321]. Складність використання методу полягає у тому, що необхідно коректно поставити задачу, обрати відповідні змінні, що характеризують фінансовий процес, отримати адекватні статистичні дані для навчання мережі. Інтерпретація результату, отриманого за допомогою побудованої мережі, потребує глибокого розуміння процесу, що моделюється [21, 104, 175].

Для аналізу фінансових процесів і подій застосування МБ надає такі можливості: створювати моделі високої розмірності; використовувати

статистичні дані і експертні оцінки в одній моделі; формувати висновок у «прямому» та «зворотному» напрямках; будувати складні комбіновані ймовірнісно-статистичні моделі, які включають у себе МБ, регресійні та інші типи моделей. Сфери застосування МБ такі: розпізнавання ситуацій; прогнозування динаміки розвитку процесів; побудова діагностичних систем у медицині, економіці, техніці; підтримка прийняття управлінських рішень і т. ін. [21, 54, 69, 227, 218, 314].

Визначення. *Мережа Байєса* (МБ) – це трійка $N = \langle V, G, J \rangle$, де V – множина змінних; G – спрямований ациклічний граф, вузли якого відповідають випадковим змінним модельованого процесу; J – спільний розподіл ймовірностей змінних $V = \{X_1, X_2, \dots, X_n\}$. Виконується марковська умова: кожна змінна мережі не залежить від усіх інших змінних, за винятком батьківських попередників цієї змінної [125].

Змінну (вузол) називають *нащадком*, якщо вона залежить від однієї або більше інших змінних. *Батьківська* змінна має одну або більше змінних-нащадків. До множини *нащадків* відносять змінні-нащадки, які є нащадками однієї і тієї самої батьківської змінної, а також змінні-нащадки змінних-нащадків вищого рівня.

Визначення. *Гібридна мережа Байєса* $B = (X, G, P)$ визначається через спрямований ациклічний граф $G = (X, E)$ і його функції $P_i = \{P(x_i | pa_i)\}$, де pa_i – множина батьківських вузлів x_i . X – множина змінних, розділених на дискретні Δ і неперервні Γ змінні, тобто $X = \Gamma \cup \Delta$. Неперервні змінні графу G не можуть мати дискретні змінні як їх вузли-нащадки [20, 24, 95]. Умовний ймовірнісний розподіл (УЙР) неперервних змінних задається лінійною гаусівською моделлю: $P(x_i | I = i, Z = z) = N(\alpha(i) + \beta(i) \times z, \gamma(i))$ $x_i \in \Gamma$, де Z та I – множини відповідно неперервних і дискретних батьків x_i , $N(\mu, \sigma)$ – мультиваріантний нормальний розподіл. Мережа представляє собою спільний розподіл усіх змінних графу, заданий добутком таблиць умовних ймовірностей.

Визначення. Динамічна мережа Байєса (ДМБ) є парою (B_1, B_{\rightarrow}) , де B_1 – МБ, що визначає апіорну ймовірність $P(Z_1)$, а B_{\rightarrow} – двошарова МБ, що визначає $P(Z_t | Z_{t-1})$ за допомогою спрямованого ациклічного графа як:

$$P(Z_t | Z_{t-1}) = \prod_{i=1}^N P(Z_t^i | Pa(Z_t^i)),$$

де Z_t^i – i -й вузол в момент часу t , що може бути компонентою X_t, Y_t або U_t , а $Pa(Z_t^i)$ – батьки Z_t^i на графі [13, 144, 290]. Вузли першого шару двошарової МБ не мають жодних параметрів, що з ними асоціюються, але кожен вузол другого шару має зв'язаний з ним розподіл умовної ймовірності, що визначають $P(Z_t^i | Pa(Z_t^i))$ для всіх $t > 1$ [288].

Основною перевагою застосування МБ для задач аналізу ризиків фінансових систем є можливість одночасного врахування кількісних та якісних ринкових показників, динамічного оновлення інформації, використання явної залежності між існуючими факторами, що впливають на фінансові показники, а також наочність моделювання. У мережах Байєса можливе використання як дискретних, так і неперервних змінних, врахування невизначеностей та великого обсягу характеристик [20, 95, 99, 123]. У порівнянні з іншими методиками оцінювання ймовірності настання ризиків, МБ є гарним інструментом для класифікації, оскільки не мають обмежень на закони розподілу змінних та не вимагають повноти інформації [147, 176].

При побудові МБ існують проблеми знаходження точного імовірнісного висновку на основі навчальної вибірки, вибору змінних, обчислення ймовірності вузлових вершин [20, 75, 77, 123]. Докладніше питання навчання ДМБ, застосування до задач класифікації, а також приклад комбінованого використання нейронних мереж та мереж Байєса, наведені у додатку А.3.

1.3.2.5. Деревя рішень

Деревя рішень використовують для аналізу ризиків подій, що мають невелику кількість варіантів розвитку, а рішення, що приймаються у момент

часу $t = n$, сильно залежать від рішень, прийнятих раніше, і визначають сценарії подальшого розвитку подій [123, 189].

Математичний апарат дерев рішень використовується в автоматизованому аналізі даних, коли правила представляються у вигляді послідовної ієрархічної структури, де кожному вузлу відповідає єдиний, що дає можливість прийняти рішення [189].

Сьогодні відома велика кількість алгоритмів, що реалізують дерева рішень: CART, C4.5, CHAID, тощо [189]. CART (Classification And Regression Tree) – метод побудови бінарного дерева, кожний вузол якого при розбитті може мати лише двох нащадків [234]. Метод найчастіше використовується для вирішення задач класифікації і регресії.

Алгоритм CHAID (CHi-squared Automatic Interaction Detector) використовує тест χ^2 -квадрат для оцінки якості розбиття. Може застосовуватись для числових і категоріальних змінних. Для неперервних змінних потребує їх трансформації у категоріальні змінні [189, 234].

Моделі CHAID чи CART формують дерева рішень за найбільш інформативними атрибутами. Дерева рішень широко використовуються завдяки доволі високій точності прогнозу для більшості видів бізнес-даних та легкості розуміння і інтерпретації результату [123, 234]. До основних переваг методу дерев рішень відносять: швидкий процес навчання, генерування правил, коли знання складно формалізувати, зрозуміла класифікаційна модель, висока точність прогнозу.

1.3.2.6. Методи оптимальної фільтрації

Методи оптимальної фільтрації (зокрема, фільтр Калмана) є одними з найбільш часто застосованих у зв'язку з наявністю різних корисних можливостей функціональності фільтра, включаючи оцінки змінних, які не піддаються вимірюванню, компонентів вектора стану, оцінки статистичних параметрів для стохастичних збурень і похибок вимірювань, оцінки деяких параметрів моделі [17]. На етапі оцінювання структури і параметрів моделі зустрічається ще множина невизначеностей, викликаних помилками оцінки

моделі порядку, відсутністю інформації про нелінійності процесів, час затримки і тип розподілу даних. Структура моделі, як правило, покращується завдяки реалізації відповідної схеми адаптації [74], застосування тестів, які допомагають виявити структуру статистичних даних у кращому вигляді. Невизначеність даних і структури моделі, безумовно, впливають на якість оцінок параметрів, вказують, що вони можуть призвести до більш високої дисперсії та/або зміщення оцінок. Для зменшення впливу невизначеностей, вводять адаптивні схеми оцінювання як для структури, так і для параметрів моделі. Крім того, ключовим моментом для оцінки параметрів математичної моделі є коректне визначення типу розподілу даних, що обробляються. Якщо він далекий від нормального, то застосовується відповідний алгоритм Монте-Карло для марківського ланцюга, адаптований до специфічного розподілу статистичних даних [263, 314].

При аналізі фінансових процесів виникає необхідність зменшення похибок, які виникають внаслідок дії зовнішніх факторів, помилок вимірів, тощо. Оптимальний фільтр використовується для фільтрації даних (згладжування шумів та боротьби зі збуреннями), оцінювання невимірюваних параметрів і змінних, короткострокового прогнозування, комплексування, тобто об'єднання різних потоків інформації для уточнення оцінок векторів станів.

У дисертаційній роботі методи оптимальної фільтрації використано для боротьби зі збуреннями та відновлення втрачених даних і надали покращення результатів прогнозування. Детально реалізацію алгоритму побудови оптимального фільтра для задачі оцінювання ризиків показано у третьому розділі.

1.3.2.7. Регресійні моделі

1.3.2.7.1. Лінійна множинна регресія

Методи регресійного аналізу [57, 123, 189, 213] широко використовуються для вирішення задач оцінювання попиту, доходності акцій, у макроекономічних розрахунках, тощо.

Загальна множинна регресійна модель має такий вигляд:

$$y = f(x_1, x_2, \dots, x_p) + \varepsilon$$

де y – залежна змінна; x_1, x_2, \dots, x_p – фактори (незалежні змінні); ε – похибка.

Множинна регресійна модель є лінійною, якщо вона може бути представлена як [17]:

$$y = b_0 + b_1 x_1 + \dots + b_p x_p + \varepsilon.$$

Позначимо i -е спостереження змінної y через y_i , а факторів – x_1, x_2, \dots, x_p .

Тоді модель можна подати у вигляді [26]:

$$y_i = b_0 + b_1 x_{i1} + \dots + b_p x_{ip} + \varepsilon_i, i = \overline{1, n},$$

або у матричній формі:

$$y = Xb + \varepsilon,$$

де $y = [y_1, y_2, \dots, y_n]^T$ – вектор (матриця-стовпець) значень залежної змінної;

$b = [b_1, b_2, \dots, b_p]^T$ – вектор (матриця-стовпець) коефіцієнтів регресійної моделі;

$\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n]^T$ – вектор (матриця-стовпець) похибок [17];

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \text{ – матриця значень факторів.}$$

Відзначимо основні припущення регресійного аналізу:

– В моделі похибка ε_i (або залежна змінна y_i) є випадковою величиною, а фактори x_{ip} невинпадкові величини ($i = \overline{1, n}$) [17].

– Математичне сподівання похибки ε_i дорівнює нулю, дисперсія похибки ε_i (або залежної змінної y_i) постійна, тобто виконується умова гомоскедастичності.

– Похибки ε_i та ε_j некорельовані [26]:

$$M[\varepsilon_i, \varepsilon_j] = 0, i \neq j.$$

- Похибка ε_i є нормально розподіленою випадковою величиною.
- Матриця значень факторів невироджена, тобто її ранг дорівнює $p + 1$:

$$\text{rang}X = p + 1 < n.$$

Модель, для якої виконуються всі припущення, називається класичною нормальною лінійною моделлю множинної регресії (CNLMR-model). Оцінкою цієї моделі за вибіркою є рівняння регресії [26]:

$$y = b_0 + b_1x_1 + \dots + b_px_p,$$

де y – оцінка математичного сподівання залежної змінної $M_x[y]$; $b_i (i = \overline{0, p})$ – оцінка коефіцієнтів регресійної моделі (або коефіцієнти регресії).

Для оцінки коефіцієнтів CNLMR-model використовують метод найменших квадратів (МНК). Значення коефіцієнтів рівняння регресії, які в матричній формі мають вигляд:

$$b = (X^T X)^{-1} X^T y,$$

де $b = [b_1, b_2, \dots, b_p]^T$ – вектор (матриця-стовпець) коефіцієнтів рівняння регресії. Оцінки b_j є незміщеними, обґрунтованими та ефективними [26]. Оцінка дисперсії похибок є незміщеною.

$$S^2 = \frac{\sum_i^n (y_i - \hat{y}_i)^2}{n - p - 1}.$$

1.3.2.7.2. Логістична регресія

Суть логістичної регресії [287] полягає у використанні спеціальної функції перетворення (сигмоїдальної функції), що відповідає логістичному розподілу з нульовим середнім та середньоквадратичним відхиленням, що дорівнює числу 3, поділеному на корінь з π .

Основними перевагами логістичної регресії є отримання кількісної оцінки ймовірності (наприклад, дефолту позичальника кредиту) у явному вигляді в допустимому інтервалі $[0, 1]$, а також те, що вона характеризується високою стійкістю і менше характеризується властивістю «підлаштовуватись» під навчальну вибірку. Аналіз літератури [24, 123, 145, 213, 287] показав,

логістична регресія найчастіше використовується для оцінювання кредитних ризиків позичальників і показує хороші результати застосування. Недоліками логістичної регресії є певні неточності при описі перелічувальних змінних та її нечутливість до деяких значень числових змінних [137, 172].

Для застосування такої регресійної моделі на практиці вхідні дані мають відповідати таким вимогам:

– обмеження для незалежних (екзогенних) змінних: дані, необхідні для функціонування моделі, повинні бути представлені у вигляді двовимірної таблиці. Вся інформація про об'єкти з предметної області повинна описуватися у вигляді скінченної множини параметрів [287]. Кожен параметр повинен мати категорійне або числове значення. Сама множина незалежних змінних не має змінюватися від прикладу до прикладу, а кількість параметрів має бути фіксованою для всіх прикладів;

– вимоги до залежної змінної (змінної, що прогнозується): кожен приклад повинен бути асоційований однозначно з конкретною характеристикою, яку необхідно прогнозувати [60].

Нелінійна модель у формі логістичної регресії дозволяє встановити обмеження, що стосуються інтервалу для ймовірності $p \in (0;1)$:

$$\ln\left(\frac{p}{1-p}\right) = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n,$$

де x_i – параметри моделі, w_i – ваги параметрів.

Необхідно визначити ймовірність появи події залежно від значень параметрів $x = \{x_1, x_2, \dots, x_n\}$. Виходом є значення логіт-функції [137], яка приймає значення від 0 до 1, тобто:

$$g(z) = \frac{1}{1 + e^{-z}}.$$

Тоді модель матиме вигляд:

$$\text{Pr} = g(p_0 + p_1x_1 + \dots + p_mx_m).$$

Припустимо тепер, що є вибірка: $(x_1^i, x_2^i, \dots, x_m^i, y^i) (i = 1, \dots, n)$, де y^i – значення залежної змінної; n – кількість спостережень.

Коефіцієнти моделі оцінюються за принципом максимальної правдоподібності тими значеннями, для яких досягається максимум функції правдоподібності [26]:

$$L(\bar{p}) = \prod_{i=1}^n g(\overline{px^i})^{y^i} [1 - g(\overline{px^i})]^{1-y^i}.$$

Для зручності позначимо:

$$\bar{p} = (p_0, p_1, p_2, \dots, p_m),$$

$$\overline{x^i} = (x_1^i, x_2^i, \dots, x_m^i),$$

$$\overline{px^i} = p_0 + p_1 x_1^i + p_2 x_2^i + \dots + p_m x_m^i.$$

Зазвичай замість функції правдоподібності використовується її логарифм, що не змінює суті задачі, але дозволяє позбавитись від добутку [26]:

$$l(\bar{p}) = \sum_{i=1}^n y^i \ln(g(\overline{px^i})) + (1 - y^i) \ln(1 - g(\overline{px^i})).$$

Для застосування логістичної регресії необхідні набагато більш складні розрахунки для отримання вагових коефіцієнтів. Перевага методу логістичної регресії полягає у можливості розподілу випадків як на дві групи (0 – поганий, 1 – хороший), так і на кілька груп (1, 2, 3, 4 групи ризику). Це і обумовило успішність їх використання у існуючих скорингових системах.

1.3.3. Сучасні практики оцінювання ризиків у фінансових системах

Оцінювання ризиків фінансових систем вимагає використання серйозного математичного апарату [89, 191, 192], напрацювання методик застосування згідно визнаних світових нормативів і впровадження відповідних програмно-технічних засобів для досягнення необхідної швидкості та актуальності динамічного оцінювання показників [88, 137, 181].

Серед усієї множини методів слід виокремити *метод коригування норми дисконту* з урахуванням ризику, який найчастіше використовується на практиці

і передбачає корегування деякої базової норми дисконту [49] (завдяки введення премії за ризик), яка вважається без ризиковою або мінімально прийнятною.

За допомогою *методу достовірних еквівалентів* здійснюється корегування очікуваних значень суми платежів через формування знижуючих коефіцієнтів з метою приведення очікуваних надходжень до величини платежів, отримання яких практично не викликає сумнівів (гарантовано будуть отримані) і значення яких може бути достовірно визначено.

Метод сценаріїв дозволяє поєднати дослідження чутливості показника результату з аналізом імовірнісних оцінок його відхилень, і отримати досить наочну картину для різних варіантів подій [82]. Метод є певним розвитком аналізу чутливості, оскільки передбачає одночасну зміну декількох чинників.

Методи експертних оцінок є комплексом методів і процедур обробки результатів опитування групи експертів (використання їх знань та досвіду), які є єдиним джерелом інформації [51]. Суттєвою перевагою методу є те, що він може використовуватися в умовах браку інформації, лише необхідно забезпечити виключення взаємного впливу експертів і узгодження їх оцінок.

Метод оцінки фінансової стійкості (аналізу доцільності витрат) передбачає визначення потенційних зон ризику. Виділяють такі різновиди методу: 1) оцінка фінансової стійкості підприємства; 2) співставлення фінансової стійкості підприємства до і після впровадження аналізованого проекту (оцінка доцільності витрат). Віднесення фактичного або прогнозованого стану підприємства до однієї з областей фінансової стійкості, і відповідно, областей ризику, виконується на основі аналізу достатності забезпечення підприємства обіговими коштами (власними або позиковими ресурсами) для формування запасів і покриття витрат, пов'язаних з виконанням розглянутих видів діяльності [49, 182].

Рейтинговий метод оцінювання передбачає можливість підбору коефіцієнтів, виходячи з конкретної мети аналізу. Система рейтингового оцінювання [76] складається з таких елементів: системи оціночних

коефіцієнтів; шкали ваги коефіцієнтів (за необхідності); шкали оцінки значень одержаних показників; формули розрахунку остаточного рейтингу [304].

Переваги: передбачає аналіз великих масивів даних; одразу відбувається ранжування отриманого результату за певною шкалою [78]; обсяг необхідних математичних знань – в межах елементарних фінансових розрахунків [30, 191].

Недоліки: проблема вибору еталону для порівняння [82], оскільки вимагає уточнення для кожного виду ризику [49]. У дисертаційній роботі використовувався рейтинговий підхід світових організацій (S&P, Moody's, ING) та НБУ для оцінювання інвестиційних та кредитних ризиків.

Нормативний метод базується на використанні системи фінансових коефіцієнтів: ліквідності, заборгованості, автономії, маневреності, покриття, тощо. Перевагою методу є легкість і швидкість розрахунків. Система нормативів є своєрідним удосконаленням рейтингового методу, що передбачає формування заздалегідь певної шкали оцінки із мінімумом значень ранжування [82]. Метод дозволяє встановити ступінь ризику з відносною точністю: порівняння з нормативом відбувається за шкалою «низький», «нормальний», «високий», а тому не дає можливості врахування всіх нюансів конкретної ситуації. Недоліками методу є його невисока точність оцінки, неможливість врахувати особливості конкретної ситуації [82].

Фундаментальний метод. Загальний фінансовий ризик розраховують за допомогою фундаментальних показників: мінливості прибутковості активу, малих розмірів компанії (P/BV), незбалансованого зростання (рентабельність власного капіталу (Return of equity – ROE) вище, ніж коефіцієнт збалансованого зростання) та інші. Значимими є внутрішні і зовнішні чинники – структура витрат на одиницю виручки [39]; періодичність операційних процесів і торгової політики в стосунках з дебіторами і кредиторами; додаткова вартість капітальних вкладень; структура фінансування, також ці чинники завдяки аналітичній обробці зазнають мінливості за кожним чинником як міра розбиття значень ключових орієнтирів (деталізація по ROE). Опис характеристик ризику

можна проілюструвати за допомогою деталізації показника рентабельності власного капіталу (ROE) [38 – 40].

В основу оцінювання фінансового ризику покладено структурну формалізацію показників за факторним аналізом, а саме – обчислення міри ризику завдяки визначенню вагомих чинників, які кількісно можна виміряти або ідентифікувати. Інтерпретацією фундаментального методу є метод оцінювання ризику за Шарпом [82], що базується на величині очікуваного прибутку, яка враховує статистичні дані про його рівень протягом певного часового тренду і розподіл ризику на систематичний і не систематичний [276]. Величина очікуваного прибутку визначається, виходячи з середньогалузевої норми дохідності і тенденції розвитку ринку в цілому [82].

Метод аналогів використовується тоді, коли вживання інших методів за будь-яких причин є неприйнятним. Застосовується база даних аналогічних об'єктів для виявлення загальних залежностей і перенесення їх на досліджуваний об'єкт. Результати аналізуються на основі попереднього досвіду для виявлення потенційних факторів ризику, що є перевагою за відсутності чіткої бази для порівняння [49]. Недоліки методу: ігнорується фактор постійного розвитку, не враховується динамічність системи і зміни зовнішнього середовища.

Технологія “Risk Metrics” розроблена компанією “J.P. Morgan” [283] для оцінювання ризиків ринку цінних паперів [32]. Методика має на меті визначення міри впливу ризиків на подію через обчислення “міри ризику”, тобто максимально можливої потенційної зміни ціни портфелю [82], із заданою вірогідністю за заданий проміжок часу.

Stress Testing method – це метод кількісного оцінювання ризику, що полягає у визначенні величини неузгодженої позиції, що наражає банк на ризик, і у визначенні шокової величини зміни зовнішнього фактора – валютного курсу, процентної ставки тощо [27]. Поєднання цих величин визначає, яку суму збитків або доходів отримає банк, якщо події розгортатимуться за закладеним сценарієм. Органи фінансового регулювання

використовують два підходи до стрес-тестування: «зверху вниз» (top down approach) і «знизу вгору» (bottom up approach). У першому варіанті розрахунки проводяться самим регулятором. Підхід «згори вниз» вимагає використання не тільки макроекономічної статистики, але й мікроданих, а також передбачає економіко-математичне моделювання. Перевагою підходу «зверху вниз» є те, що після розробки методики стрес-тестування [212], його регулярне проведення не вимагає великих витрат часу і праці. При використанні підходу «знизу вгору» органи фінансового регулювання визначають сценарії і величину шоку, а фінансові організації на основі отриманої інформації проводять самостійні розрахунок втрат, використовуючи внутрішні моделі. Термін проведення стрес-тестування істотно збільшується у порівнянні з підходом «зверху вниз» [148].

Основні переваги методу: відносна простота методу; швидке проведення розрахунків; більш гнучке формулювання можливих подій.

Основні недоліки методу: ігнорує багато значних деталей, необхідних для реального представлення фінансових ринків; не враховує, який внесок в ризик вносить фінансовий ринок.

Моделі, що дозволяють отримати в явному виді функцію розподілу втрат, можуть використовуватися як для оцінювання середніх втрат, так і для оцінювання максимальних втрат на заданому рівні значущості [11]. В той же час спрощені моделі або моделі, засновані на експертному оцінюванні, не дозволяють отримати кількісні оцінки цих параметрів і застосовні для вирішення вузького класу завдань (табл. 1.4).

За класичним ІМА-підходом (Internal Measurement Approach) [212] внутрішня оцінка очікуваних витрат спочатку визначається для кожної пари «бізнес-лінія – категорія втрат», вводиться множник, який дозволяє перерахувати очікувані втрати у неочікувані. Для інформаційних ризиків, наприклад, такий перерахунок здійснити нереально, оскільки самі по собі інформаційні ризики комбінують багато категорій, для яких можуть відрізнятися як бізнес-процеси, так і активи. ВІА-метод (Business Impact Analysis) є засобом оцінювання впливу на бізнес, що дозволяє дослідити як

ключові види відмов/ порушень можуть вплинути на основні види діяльності і процеси організації. Він також дозволяє ідентифікувати і кількісно визначити ключові критичні бізнес-процеси, функції, взаємозв'язки, які існують в організації, та необхідні можливості управління організацією в цих умовах.

Таблиця 1.4

Класифікація моделей оцінювання ризиків

Клас моделей	Застосування для розподілу ризиків по групах критичності	Застосування для визначення середніх втрат	Застосування для визначення максимально можливих втрат	Застосування для виявлення подій, що ведуть до реалізації ризику
<i>Моделі, основані на аналізі наслідків (Top Down)</i>				
BIA			+	
IMA			+	
LDA	+	+	+	
<i>Моделі, основані на аналізі факторів ризику (Bottom Up).</i>				
Sb-AMA	+	+	+	
Метод функціональних кореляцій	+	+	+	
Регресійні моделі	+	+	+	+
Байєсівські мережі	+	+	+	+
Методи нечіткої логіки	+			+

Метод LDA (Loss Distribution Approach) був запропонований Базельським комітетом для оцінювання операційних ризиків [116]. Він ґрунтується на використанні VaR-моделі оцінювання втрат. Для оцінки функції розподілу втрат застосовуються або лише внутрішні, або лише зовнішні дані про втрати. Підхід SCA (Scorecard Approach) передбачає використання експертних знань для оцінки функцій розподілу. За AMA-підходом (Advanced Measurement Approach) великі компанії вже мали оцінювати свій операційний ризик за більш досконалими моделями, які водночас були більш чутливими до якості ризик-менеджера. AMA-підхід складається з трьох складових, а саме: внутрішньої оцінки (IMA-підхід), розповсюдження даних про втрати (LDA-підхід) і

застосування скорингової панелі (SCA-підхід), який буде детальніше розглянутий та допрацьований у другому розділі.

Методика оцінки фінансових ризиків на основі розрахунків ймовірностей з точки зору фінансового менеджера є незручною для застосування, оскільки визначається імовірнісний розподіл збитків і не здійснюється конкретна вартісна оцінка фінансового ризику [35].

Value-at-risk (VaR-метод) оцінки ФР, що базується на аналізі статистичної природи ринку. Це універсальна методика оцінки різноманітних видів ризиків (цінового, валютного, кредитного та ризику ліквідності). VaR-метод став загально визнаним методом оцінки ризику серед учасників західної фінансової системи і регулюючих органів. Фактично, методика VaR наразі просувається як стандарт оцінки ризику [10, 148, 216]. Нефінансові корпорації можуть використовувати техніку VaR для оцінки ризиків грошових потоків і прийняття рішень про хеджування (захист капіталу від несприятливого руху цін). Одним з тлумачень VaR є кількість незастрахованого ризику, яку приймає на себе корпорація [35, 216]. Інвестиційні аналітики використовують VaR для оцінювання різних проектів. Інституційні інвестори, такі як пенсійні фонди, використовують VaR для розрахунку ринкових ризиків.

Основні переваги методу: висока точність розрахунків; дозволяє агрегувати ризики окремих позицій в єдину величину для всього портфеля. Недоліки методу: відносно висока складність моделей; високі вимоги до обчислювальних потужностей і, як результат, значні витрати часу на проведення розрахунків.

Суть методу полягає в тому, щоб визначити вартість фінансового ризику як найменшу можливу величину капіталу, необхідного для забезпечення заданого рівня ймовірності ризику.

В інвестиційних компаніях і банках методологія VaR може застосовуватися для виконання таких завдань [10]:

1. Внутрішнього моніторингу ринкових ризиків. Інституційні інвестори можуть обчислювати і виробляти моніторинг значень VaR за декількома

рівнями: агрегованого портфелю, по класу активу, по емітенту, по контрагенту, по трейдеру, портфельному менеджеру і т.д. З точки зору моніторингу точність оцінювання величини VaR відходить на другий план, оскільки важлива величина відносного, а не абсолютного значення VaR [216], тобто VaR керуючого або VaR портфелю в порівнянні з VaR еталонного портфелю, індексу, іншого менеджера або того ж менеджера в попередні моменти часу.

2. Зовнішнього моніторингу. VaR дозволяє створити уявлення про ринковий ризик портфелю без розкриття інформації про склад портфелю і оцінити чи прийнятий ризик є допустимим (прийнятним) [10].

3. Моніторингу ефективності операцій зниження ризику. Менеджер може оцінити ефективність конкретної стратегії управління ризиком шляхом порівняння величин VaR інвестиційних портфелів з прийнятою стратегією хеджування і зовсім без неї [216]. Якщо різниця між ними невелика, то виникає питання про доцільність хеджування або чи правильно хеджування застосовується.

4. Автоматичного аналізу можливих управлінських рішень. Методологія VaR дозволяє дати більше свободи і автономії персоналу, відкидає необхідність затвердження тих чи інших угод. Здійснюється моніторинг транзакцій (угод) з використанням VaR, може бути лише встановлене правило для брокерів-дилерів: «Ніяка операція не повинна призводити до збільшення значення VaR більш ніж на $X\%$ початкового капіталу».

За допомогою методології VaR стає можливим обчислити оцінки ризику різних сегментів ринку і ототожнити найбільш ризикові позиції. Оцінки VaR можуть використовуватися для диверсифікації капіталу, установки лімітів, а також оцінки діяльності компанії [35]. У деяких банках оцінка операцій трейдерів, а також їх винагорода обчислюються як прибутковість на одиницю VaR.

Сама методологія VaR не є засобом управління фінансовим ризиком, оскільки вона не звільняє від фінансових втрат. VaR-метод не може визначити оптимальну величину ризику, який необхідно прийняти на себе компанії (це

завдання фінансового ризик-менеджера), проте дозволяє оцінити величину вже прийнятого ризику. VaR-метод є частиною комплексного аналізу ФР [10] і повинен використовуватися не замість, а разом з іншими методами оцінки ризику.

SAR-метод (Shortfall-at-Risk) оцінки ФР. Досить часто для оцінки ризику інвестора цікавить не ймовірність отримання збитків, а саме очікувана величина збитку, бо ймовірність отримання збитку може бути дуже мала, але розмір збитку настільки великим, що наслідки несприятливого результату можна вважати катастрофічними [10]. Тому необхідна оцінка ризику, що враховує і величини можливих збитків. Таким методом оцінки фінансового ризику є SAR -метод.

Оскільки ризик обумовлений невизначеністю результату, то, чим менше дисперсія можливих значень випадкової величини, тим більше її передбачуване значення, а, отже, менше ризик. Такого роду міркування привели до поширення точки зору, що мірою ризику інвестиційного проекту вважається середньоквадратичне відхилення його прибутковості. Проте існує достатня кількість прикладів, коли збільшення дисперсії знижує ймовірність. В умовах, свідомо програшних для інвестора, йому слід обирати стратегії, що призводять до збільшення дисперсії (ризiku).

Метод еквівалентного фінансового інструменту. Найбільш зрозумілою для інвестора моделлю оцінки фінансового ризику є метод еквівалентного фінансового інструменту. Так, якщо деяка фінансова стратегія (фінансовий інструмент) повністю страхує від ризику, то зведена вартість поточних витрат на обслуговування стратегії і є ціною ризику, яку необхідно обчислити [124]. Більше того, якщо сам інструмент торгується на ринку, то його ринкова ціна визначає міру того фінансового ризику, який страхується даними фінансовим інструментом [179].

Розглянуті методи відбивають економічний погляд на аналіз ринків, є зрозумілими для фінансових спеціалістів, але майже не використовують всі

можливості сучасної математики та інформаційних технологій для більш ефективного управління ризиками.

1.4. Невизначеності в задачах моделювання і менеджменту фінансових ризиків

Появу ризику необхідно досліджувати разом з проблемою невизначеності (неточності, недостовірності, ненадійності) знань про умови та процеси, які відбуваються в об'єкті та зовнішньому середовищі, з імовірнісним характером виникнення небажаних подій [107]. Тому вивчення ризиків пов'язане з аналізом невизначеності, а ефективні шляхи для попередження небажаних подій, зменшення рівня ризиків при прийнятті рішень пов'язані із цілеспрямованою мінімізацією невизначеності. Але ототожнювати ці два поняття теж не можна, оскільки ризик – категорія суб'єктивна, а невизначеність – об'єктивна [156].

Під *невизначеністю* розуміють стан неоднозначності розвитку подій в майбутньому, стан нашого незнання і неможливості точного передбачення основних величин і показників розвитку деякого процесу. Невизначеність передбачає наявність факторів, при яких результати дій не є детермінованими, а ступінь можливого впливу цих факторів на результати невідома [76]. У практичних завданнях можна зустрітися із складністю визначення всіх варіантів розвитку подій в майбутньому та встановлення ймовірностей настання тих чи інших подій.

При застосуванні імовірнісного аналізу зустрічаються обмеження, пов'язані з визначенням вхідних даних, розподілом їх ймовірностей і стохастичних залежностей, часто навіть неможливо зробити відповідні припущення і встановити закони розподілу невизначених чинників, а з імовірнісного аналізу не випливає висновок про вплив окремих вхідних величин на результат [151].

Для задач аналізу ризиків фінансових систем важливим є дослідження невизначеностей, класифікація методів обробки пропущених та втрачених даних, пов'язаних з проявом впливу невизначеностей навколишнього світу і

об'єкту моделювання. Актуальною є розробка рекомендацій для створення коректного підходу при обробці неповних даних, які дадуть можливість підвищити прогнозуючу якість моделей, побудованих на відновлених за цією методикою пропущених даних [107]. Зокрема, ця задача є важливою при обробці вибірок невеликих розмірів, коли некоректне оцінювання пропущених змінних є вкрай суттєвим і може викликати похибки подальшого прогнозування поведінки системи і побудови прогнозних моделей. Неповнота або відсутність інформації за окремими категоріями є найбільш важливим фактором, що ускладнює роботу при оцінюванні ФР і може суттєво впливати на отримані результати оцінювання.

Для прийняття рішення в умовах невизначеності використовується низка критеріїв, найпоширеніші серед них такі: Лапласа, Вальда, Севіджа та Гурвіца [132, 151].

1.4.1. Вплив невизначеностей на формування ризиків різної природи

Фактично саме реальність і неідеальність зовнішніх умов, неможливість завчасно їх визначити чи спрогнозувати, і спричиняють появу ФР з конкретними фінансовими величинами – втратами та ймовірністю їх появи. Можна сподіватись, що вчасне визначення всіх зовнішніх умов та факторів хоча б на рівні певної величини x_i з проміжку $[a_i, b_i]$ дало б можливість аналізувати і реально формувати модель для прогнозування стабільного розвитку підприємства. Ризик прояву зовнішніх факторів в такому випадку буде визначатися [107] як ймовірність попадання даної характеристики (фактору) в очікуваний інтервал $P(R) = P(x_i \in [a_i, b_i])$, де x_i – i -й фактор, а a_i і b_i – відповідна ліва та права межа інтервалу, в який оцінюється попадання фактору. Оскільки ймовірність оцінюється в межах від 0 до 1, то тоді ймовірність попадання фактору x_i в інтервал $1 \leq j \leq N$ оцінюється як $\frac{1}{a_{j+1} - a_j}$,

де $[a_j; a_{j+1}]$ – межі інтервалу для фактору x_i з розбиттям на N однакової ширини інтервалів (наприклад, ризик 0,1;0,2;0,3 і т.д.).

Інтегральна оцінка ризику за сукупністю факторів визначається за формулою: $R = \sum n_i \cdot f_i$, де R – інтегральна оцінка ризику, f_i – експертна оцінка по i -му фактору, n_i – нормоване значення (вага) i -го фактора [107].

Нормоване значення i -го фактора визначається за формулою: $n_i = \frac{v_i}{\sum v_i}$, де v_i – вагове значення i -го фактора за 100-бальною шкалою.

Розмір збитків EL від прояву різних видів ФР для підприємства розраховується як:

$$EL = \sum_{i=1}^N P(R_i) \cdot CE_i \cdot LGD_i, \quad (1.10)$$

де $P(R_i)$ – ймовірність (очікувана частота) прояву i -го виду ризику (наприклад, ризику зниження фінансової стабільності), що набуває значення на відрізку $[0,1]$; CE – загроза внаслідок реалізації ризику – сума втрат (заборгованості внаслідок реалізації даного ризику) [107]; LGD – покриття ризику страховкою (в разі її наявності), заставою або ефективність запобіжних засобів, що приймає значення від 0 (ризик повністю покритий заставою) до 1 (ризик не покритий заставою); N – кількість ризиків. Тут відображено логіку розрахунку втрат, але не всі характеристики в цій формулі можна визначити експериментально. В таких випадках слід на початковому етапі скористатися експертним методом для оцінки можливих втрат, покриття ризику та частоти його прояву. Після проведення аналізу та збору результатів діяльності підприємства дані, отримані експертним шляхом, коригуються та уточнюються.

Недостатня інформація про той чи інший об'єкт, неможливість спрогнозувати його подальші дії, неточність або неправдивість вхідної інформації викликає появу різних видів ризиків, інколи навіть декількох [176]. У таблиці 1.5 наведена класифікація невизначеностей та, відповідно, можливих ризиків і загроз, які вони спричиняють [107].

Таблиця 1.5

Класифікація невизначеностей та ризиків

Види невизначеності (зокрема, для фінансових систем)	Стисла характеристика невизначеності за запропонованою ознакою	Приклад реалізації ризику
1. Перспективна невизначеність (коливання попиту та пропозиції)	Виникає внаслідок появи непередбачуваних факторів, що впливають на хід розвитку і ефективність функціонування об'єктів (процесів); об'єкт, що досліджується, недостатньо визначений	Втрата продукції внаслідок стихійного лиха
2. Ретроспективна невизначеність	Пов'язана з відсутністю інформації про поведінку об'єкту, що досліджується, в минулому. Можливий або перехід до ситуацій невизначеності чи ризику, чи такий перехід принципово неможливий.	Відсутність попиту на продукцію через вичерпаність ринку
3. Технічна невизначеність (порушення фінансової стабільності)	Є наслідком неможливості передбачення точних результатів рішень, що приймаються.	Збої в технологічному процесі виготовлення та/або поставки сировини, товарів та послуг
4. Стохастична невизначеність	Виступає результатом ймовірнісного (стохастичного) характеру процесів та явищ, що досліджуються. Можливі наступні випадки: <ul style="list-style-type: none"> • наявна надійна статистична інформація; • відомо, що ситуація стохастична, але необхідної статистичної інформації для оцінювання її ймовірнісних характеристик немає; • висувається лише гіпотеза про стохастичний характер процесів та явищ, що досліджуються, яка потребує перевірки. 	Можливість повернення товару на склад через брак, пошкодження або зрив продажів дистриб'юторами
5. Невизначеність стану природи (невизначеність зовнішніх факторів)	Пов'язана з повним або частковим незнанням навколишніх умов, при яких прийдеться приймати рішення.	Поставка продукції, на яку не буде попиту (обігрівачів до Африки)
6. Невизначеність цілеспрямованої протидії (невизначеність дій конкурентів)	Зустрічається в ситуації конфлікту двох або більше сторін, коли кожна сторона не має знань або має в наявності неповну, неточну інформацію про мотиви і характер поведінки протидіючих сторін.	Невиконання плану продажів через агресивні дії конкурентів (демпінгова політика, судові рішення тощо)
7. Невизначеність	Пов'язана з неоднозначністю, а інколи	Неправильне

цілей	неможливістю вибору однієї цілі при прийнятті рішення або побудові оптимізаційної моделі	встановлення планів з продажів продукції – плани виставлені в іноземній валюті, а не національній або кількісних показниках
8. Невизначеність умов (курсова залежність валют)	Виникає при недостатній або відсутній інформації в умовах, в яких приймаються рішення.	Вихід на ринок з новою продукцією, аналогів якої ще не має (start-up)
9. Лінгвістична (смілова) невизначеність	При аналізі економічних процесів, явищ, об'єктів використовуються вербальний (описовий) підхід і відповідні моделі. Характерною ознакою такого підходу є широке застосування недостатньо точно описаних з математичної точки зору термінів, уявлень.	Використання економічних показників, які описуються лінгвістичними змінними
10. Невизначеність дій	Відсутня однозначність при виборі рішень. Можливі наступні випадки: <ul style="list-style-type: none"> • ціль єдина (цільова функція єдина) – необхідно визначити серед усіх допустимих рішень найкраще; • цілей декілька – проблема не зводиться до розв'язку однієї екстремальної задачі; • цілі (одна або декілька) враховані в обмеженнях – виникає проблема знаходження будь-якого елемента з фіксованої множини, вибору кращого в деякому розумінні елементу з цієї множини. 	Неоднозначна стратегія ведення бізнесу підприємством; неконкретне виставлення мети та цілей, складність оцінки їх досяжності. Наприклад, досягнення рівня «вийти на середній щабель».

1.4.2. Ідентифікація та класифікація невизначеностей

Вважаємо, що виявлені фактори невизначеностей є характерними для більшості підприємств, є тенденційними, а тому можуть бути узагальнені для підприємств різних типів діяльності [107].

➤ *Коливання попиту на продукцію.* Це характеристика, яка незалежно від виду діяльності підприємства спричиняє надзвичайно високий *ризик неплатоспроможності* підприємства. Для банків або фінансових установ це спричиняє появу більш конкретного виду ризику – *втрати ліквідності*.

➤ *Невизначеність, пов'язана з порушенням фінансової стабільності постачальників продукції або сировини* є фактором, що може спричинити зрив виготовлення необхідних обсягів продукції, неможливість вчасно її реалізувати (суттєво для сезонних товарів) та є найбільш вірогідним для появи *ризиків зниження фінансової стабільності досліджуваного підприємства* [107]. Урахування такої невизначеності та зниження її впливу можливо шляхом моніторингу діяльності постачальників та диверсифікації поставок сировини.

➤ *Невизначеність, пов'язана з невідомою курсовою залежністю валют,* впливає на підприємства, які закупають продукцію, сировину або енергоносії за кордоном, сплачуючи в іноземній валюті. Коливання курсу в межах 10% є суттєвими і надзвичайно важливими для урахування у вартості продукції *валютного ризику*.

➤ *Невизначеність дій конкурентів.* Знання головних конкурентів є базою для виявлення основних факторів невизначеності, а моніторинг їх дій є способом подолання таких невизначеностей [107].

➤ *Невизначеність зовнішніх факторів* – широка категорія невизначеностей, яка включає такі фактори, як політична та економічна ситуація в країні, ситуація на зовнішніх та міжнародних ринках.

Можливі засоби запобігання появі невизначеності та зменшення її впливу

1. Для зменшення впливу невизначеності факторів, пов'язаних з *коливаннями попиту і пропозиції*, необхідно проводити ряд стратегічних дій та вчасно приймати управлінські рішення, здійснювати періодичний моніторинг купівельної спроможності покупців та порівняння цін, моніторинг цін на власну продукцію та продукцію основних конкурентів, приймаючи управлінські рішення стосовно забезпечення конкурентоспроможності (зменшення ціни, проведення додаткових маркетингових акцій, перепозиціонування тощо) [59].

2. Для подолання невизначеностей, пов'язаних з коливанням *курсу валют*, можна передбачити декілька стратегій. Вся фінансова звітність, щодо продукції підприємств має «зрізатися» по декільком характеристикам: гривні,

валюти (долару/ євро тощо) і умовних одиницях збуту – коробках, ящиках тощо. Це дозволить привести дані до адекватного вигляду, зменшити можливість маніпулювання цифрами попиту/пропозиції у зв'язку з коливанням курсу, відображуючи реальну ситуацію на ринку [107]. Страхуванням від валютного ризику є знаходження постачальника аналогічної чи взаємозамінної продукції у країні підприємства, а також реалізація продукції за кордоном.

3. Моніторинг цін та стратегій конкурентів – основний засіб подолання невизначеності дії основних конкурентів, але невизначеність дій конкурентів не може бути знята цілком, оскільки можлива поява нових конкурентів, нових видів продукції.

4. Невизначеність *зовнішніх факторів* [107] обумовлена проявом одного або декількох з вище описаних факторів; вона є найбільш складною для оцінювання та прогнозування, оскільки частина факторів залежить від природних умов (різкої зміни клімату, катастроф тощо), які є взагалі вкрай складними для оцінювання. Частина факторів пов'язана з економічною ситуацією всередині країни та політичними відносинами з сусідніми країнами та світовою спільнотою, що теж важко спрогнозувати.

Стратегії зменшення впливів невизначеностей та сформованих ними ризиків розглянемо на прикладі реальних підприємств.

1.4.3. Ілюстрація розкриття невизначеностей

Вирішення задачі розкриття концептуальної невизначеності з точки зору системного аналізу потребує розкриття множини різномірних невизначеностей на основі єдиних принципів, прийомів і критеріїв [146]. На практиці розв'язуються задачі дослідження невизначеності цілей розробки і перспектив конкурентоспроможності виробу, аналізуються невизначеність динаміки ринків попиту та пропозиції і невизначеність активної протидії конкурентів, невизначеність у динаміці розробки, виробництва, збуту та експлуатації певного виробу тощо [93].

Одним з проявів інформаційної невизначеності є невизначеність, пов'язана із пропусками даних. Об'єктивні характеристики певних процесів можуть бути змінені або навіть спотворені внаслідок втрати частини даних, процесів при їх отриманні, передачі чи збереженні. Постає питання відновлення таких пропущених даних і дуже суттєвим є підбір відновлюючих алгоритмів.

1.4.3.1. Неповнота даних як реалізація невизначеності

Причинами неповноти даних можуть бути: пропуски, неухважність при введенні інформації; відсутність інформації з об'єктивних причин; незнання; некомпетентність [93, 128]. Залежно від причини, пропуски можуть суттєво впливати на результати та спричиняти суттєві збитки організації, яка вчасно не отримала необхідну інформацію.

Наприклад, інформаційна невизначеність часто виникає у задачах обробки статистичних даних і пов'язана з недоотриманням, запізненням або втратою частини інформації з будь-яких причин [128]. Аналіз таких причин може дати додаткове розуміння суті пропусків і допомогти при виборі моделі їх заповнення.

Поглиблене вивчення процесів у різних областях за допомогою математичних моделей дозволяє дослідити кількісні зв'язки між вхідними та вихідними змінними, визначити, яким чином змінюються вихідні змінні при варіації вхідних змінних в широкому діапазоні та розглянути поведінку процесів на будь-яких часових інтервалах у прийнятному масштабі часу [93]. Побудована математична модель може бути дуже складною і трудомісткою, оскільки вона має враховувати тонкощі взаємодії кількісних і якісних змінних із можливим врахуванням реального часу, тобто з використанням імітаційного моделювання. Прогнозування значень змінних виконується, як правило, на основі простіших моделей, ніж поглиблене вивчення процесів. Таке спрощення моделі також може внести додаткову інформаційну невизначеність [93].

Поняття структури моделі охоплює такі параметри: порядок, розмірність моделі, наявність нелінійностей і їх характер, час запізнення (для часових рядів [4, 61]), тип збурень тощо. Знаходження структури моделі, що адекватна

процесу, є непростю задачею, що вирішується, як правило, в інтерактивному режимі [26]. Спочатку структуру моделі оцінюють наближено на підставі дослідження закономірностей протікання процесу, аналізу кореляційних функцій, візуального аналізу даних. Вибирають декілька найбільш ймовірних структур (кандидаток), далі обчислюють оцінки параметрів моделей-кандидаток і обирають кращу з них, використовуючи відповідні статистичні характеристики якості моделей [17, 134]. Якщо жодна з моделей-кандидаток не є адекватною для конкретного застосування, то необхідно перевірити на інформативність експериментальні дані, може знадобитися повторний чи додатковий збір експериментальних даних.

Неповнота даних досить широко поширений факт в процесах прийняття рішень [265, 343]. Деяка частина відсутніх даних може бути скоригована з інших джерел даних. Не всі методи можуть моделювати пропущені дані. Найбільш реальним підходом є використання середнього або медіани для заповнення відсутніх даних у вибірці. Тим не менше, використання таких методів суттєво впливає на вибірку і згладжує її, що може бути неприйнятним для визначення сплесків і критичних точок. Такий підхід передбачає, що розподіл даних нагадує нормальний розподіл для популяції [351]. Заміна відсутніх значень середнім, медіаною, або іншою центральною мірою є простим підходом, проте не найкращим. Результати заповнення відсутніх даних для різних сфер показують, що необхідна розробка нових методів заповнення, які дозволяють досягти точніших результатів наближення фактичних даних. Вибір «найкращої» технології заміни пропущеного значення за своєю суттю потребує певних припущень про справжні (відсутні) дані [308].

Пропущені дані можуть бути розділені на такі [351] групи:

➤ ознака не може бути застосована для деяких суб'єктів. З точки зору бізнесу або практики це означає, що для певного предмета (предметів) аналізу, ніякого значення для певної ознаки не може бути встановлено. Відповідний факт не застосовується для цього запису;

➤ ознака може бути застосована, але вона не була знайдена. Дані доступні, але вони не були представлені в процесі збору даних або бізнес-процесі, оскільки:

- ❖ інформація не була надана;
- ❖ деякі дані були надані, однак вони не були оброблені далі, наприклад, в процесі збору або введення даних;
- ❖ дані були введені в систему, але значення або заголовки були вилучені під час передачі даних або управління даними;

➤ з різних причин, деякі значення даних були введені неправильно і можуть бути негайно визнані безглуздими або явно неправильними [297]. Це проблема не тільки коректності, але і повноти, оскільки ці значення часто мають бути встановлені як пропуски.

Той факт, що змінна має відсутні значення, не означає, що немає доступної інформації, адже:

- представлення даних у вигляді відповідей на питання з декількома варіантами в опитуванні часто тільки вводиться як (1) значення, для перевірених категорій. Для неперевірених елементів залишають відсутнє значення в даних. Однак, це означає в дійсності «не перевірено» або «не застосовується». У цьому випадку має сенс лише заміна нульовим значенням (0), і немає необхідності в складних методах, щоб виправити помилку [351];
- у контексті даних транзакцій або часових рядів часто підрахунок для інтервалів або категорій без будь-яких спостережень представляється як відсутнє значення, яке має бути інтерпретоване як нульове значення;
- змінна може бути відсутньою для деяких спостережень, але її значення можуть бути обчислені з інших змінних одного і того ж суб'єкта в дослідженні.

Якщо спостереження містять пропущені значення, то за замовчуванням вони не використовуються для моделювання за допомогою таких методів, як нейронні мережі, байєсівські мережі або регресія [340]. Однак, відкидаючи всі неповні спостереження, можна проігнорувати корисну або важливу інформацію.

Пропущені значення можна розділити на дві категорії: систематичні і випадкові. Різниця між цими двома групами має важливе значення для того, щоб з'ясувати вплив цього значення на результати аналізу і обрати варіанти опрацювання відсутніх значень [351].

Випадкове відсутнє значення. Випадкові пропущені значення визначаються тим, що кожне спостереження в даних має таку ж ймовірність наявності відсутнього значення для певних змінних. Якщо всі дані спостережень були доступні спочатку, випадкові відсутні значення можуть бути отримані шляхом прийняття рішення для кожного спостереження, чи слід задавати ці значення як відсутні [297]. Припущення про випадковість у виникненні пропущених значень зазвичай не виконується через дуже суворі обмеження. Будь-яке пропущене значення, можна вважати, має причину.

Випадкові пропущені значення зменшують кількість інформації в базі даних для аналізу. Правильні спостереження доступні для меншого числа спостережень. Відсутні значення, які дійсно є випадковими, не пошкоджують дані, розподіл, взаємозв'язок з іншими змінними, тому що такі пропущені значення не вносять зміщення (картина неповна, але це не є неправильним) [297]. Проте, дійсно випадкові пропущені значення впливають на точність аналізу, але у вибірках великих розмірів ефект дуже малий.

Пропущені значення розглядають або як окрему категорію, з припущенням, що статистичні дані, представлені непропущеними значеннями, є репрезентативними, або із застосуванням логіки заповнення пропущених даних найбільш репрезентативним значенням заміни [231, 322, 333].

Можна вважати, що не тільки існуючі дані представляють справжню картину для аналізу, але також, що методи заповнення можуть бути використані для заповнення спостережень з пропущеними значеннями. Скажімо, платформа SAS компанії SAS Institute Inc. (США) пропонує широкий вибір способів розрахунку відсутніх значень [2, 337, 340, 351], в тому числі методів часових рядів і методів одного ряду – на суб'єкт у вітрині даних. Деякі з

цих методів приписують статичне значення для всіх спостережень; інші методи приписують індивідуальне найбільш ймовірне значення.

Систематичні пропущені значення. Систематичні пропущені значення складніше відновити, оскільки не пропущені значення не можуть розглядатися як репрезентативна вибірка. Мають бути застосовані складніші методи і стратегії заміни [350]. Вирішальним для цих методів є знання щодо походження даних, бізнес-процес, процес збору даних, а також чому вони були пропущені.

Для деяких задач, в яких відсутні дані мають систематичний характер, необхідно знайти період або закон появи відсутніх даних. Несуттєва інформація про межі або статистики відсутнього розподілу даних дає можливість коректного заповнення. Обробка пропущених значень як випадкових і заміна їх середнім може спричинити зміщення даних і результатів аналізу [351].

Для опрацювання відсутніх інтервальних змінних, наприклад, в SAS [351] представлені такі методи: хвилі Ендрю, константа за замовчуванням, метод Хубера, середнє, медіана, середнє з мінімальною відстанню, півсума крайніх значень, None, дерево, сурогатне дерево, критерій Тюка [231, 322]. Різноманітність методів пов'язана з нескінченним числом можливих станів в інтервалі змінного за означенням. Наприклад, для категоріальної змінної пропонуються такі методи заповнення відсутніх даних як підрахунок суми, константи за замовчуванням, розподіл, None, дерево, сурогатне дерево. Методи за замовчуванням для обох типів змінних (інтервальних або категоріальних змінних) дозволяють налаштувати статистику заповнення за замовчуванням, вказавши власні значення заміни відсутніх і не відсутніх даних [297]. Відсутні значення для підготовки, перевірки, контролю і оцінки набору даних замінюються статистикою заповнення, яка розраховується з навчального набору даних. Використання дерева рішень дозволяє визначити яке рішення з альтернативних або сурогатних правил обрати для того, щоб згрупувати відсутні значення в певну категорію. Можна також використовувати вузол кластера для заміни відсутніх значень.

1.4.3.2. Існуючі методи відновлення даних

Існує багато засобів заповнення пропусків уже після етапу збору даних: заповнення середнім значенням, пропорційне розташування спостережень з пропущеними даними за вже існуючими градаціями шкали, розрахунок можливого значення за допомогою регресійної моделі тощо.

Використання будь-яких засобів заповнення пропусків може змістити структуру вибірки, яка буде отримана на основі існуючих неповних даних, в бік структури неповних даних [93], що може спотворити реальний розподіл спостережень у вибірці і знищити фактичну значущість отриманих результатів.

Сьогодні існують алгоритми, які дають можливість обробляти пропуски, такі як метод Hot Deck, метод Барлета, алгоритми Resampling, Zet, Zetbraid, EM-оцінювання, регресійне моделювання та прогнозування значень [68, 73, 81, 168, 250, 333]. Особливістю цих алгоритмів є заповнення пропусків значеннями, які підбираються самим алгоритмом.

Метод Hot Deck використовує підстановку замість пропущеного значення найбільш близького інформаційного об'єкта. Підбір може здійснюватись як з усієї сукупності повних спостережень, так і з деякої підгрупи – кластеру, до якого належить цільовий об'єкт. Для заповнення пропуску по даній характеристиці у цільового об'єкта використовується значення даної характеристики у об'єкта, найближчого до цільового [168]. Тип функції відстані для визначення спостереження, найближчого до цільового (з пропуском), вибирається виходячи з типу досліджуваних даних, представлень щодо характеру зв'язку між змінними і завдань конкретного дослідження.

Метод Барлета складається з двох етапів: підстановки замість пропуску початкових згенерованих значень на першому етапі; проведення на другому етапі коваріаційного аналізу цільової змінної і побудови дихотомічного індикатору повноти спостережень за цільовою змінною [81]. Індикатор повноти спостережень завжди дорівнює 0, за винятком одного єдиного випадку: i -те значення – це цільова змінна і воно є пропущеним, тоді індикатор приймає значення 1 [93].

Алгоритм ZET. Суть цього алгоритму полягає у підборі кожного значення для заповнення пропуску не за усією сукупністю спостережень, а за деякою її частиною, яка називається компонентною матрицею, що складається з компонентних рядків і стовпців [70, 168]. Компонентність деякого рядка являє собою величину, обернено пропорційну декартовій відстані за цільовим рядком (неповного спостереження з пропуском) у просторі, осями якого задані змінні-характеристики об'єктів.

За даними компонентної матриці надалі будується функціональна залежність прогнозного значення від відповідного значення у компонентній матриці, на основі якої потім прогнозується значення пропуску [168].

Алгоритм ZetBraid. Основна відмінність цього методу від попереднього полягає у тому, що в даному алгоритмі закладено механізм об'єктивного відбору розмірності компетентної матриці. При роботі алгоритму відбувається послідовний почерговий відбір компетентних рядків та стовпчиків і кожний раз формується нова компетентна матриця. Потім за заданим критерієм визначається її ефективність при прогнозуванні пропусків [168, 169].

Resampling. Це ітеративний метод, який передбачає, що рядки з пропущеними даними замінюють випадково вибраними рядками з матриці повних спостережень, а далі будується регресійне рівняння для прогнозування пропущеного значення. Процедура регресійного моделювання повторюється декілька разів, після чого значення отриманих регресійних коефіцієнтів усереднюють і отримують кінцеве значення, яке дає максимальну точність прогнозу пропущеного значення [81].

Множинна вставка передбачає підстановку одразу кількох значень замість кожного пропущеного. Значна розбіжність цих значень означає невизначеність моделі і не дозволяє зробити висновки про їх типи і причини появи [333]. Дані, що містять набір заповнених пропусків, зберігаються в окремі масиви, кожен з яких потім аналізується як такий, що містить повні спостереження без пропусків. Наразі цей метод вважається доволі перспективним і реалізований у більшості комерційних програмних додатків.

EM-оцінювання. Метод максимізації математичного сподівання (EM – expectation maximization) або EM-оцінювання дає можливість не лише відтворювати пропущені значення з використанням двоетапного ітеративного алгоритму [254], але й оцінювати середнє значення, коваріаційні та кореляційні матриці для кількісних змінних. EM-алгоритм у загальному випадку представляє собою ітераційну процедуру, призначену для розв'язання задач оптимізації деякого функціоналу через аналітичний пошук екстремуму функції.

Алгоритм працює у два етапи. На E-кроці (крок очікування) очікуване значення функції правдоподібності обчислюється з використанням поточного наближення не вимірних змінних. M-крок використовується для обчислення оцінок параметрів за моделлю [297]. EM-алгоритм є корисним для оцінки ризиків портфелю при аналізі фінансових даних і детальніше розглянутий у додатку А.4.

Регресійне моделювання. Вставка пропущених значень за допомогою регресійних моделей також здійснюється за два етапи [17].

1. На першому етапі за сукупністю повних спостережень будується регресійна модель і оцінюються коефіцієнти рівняння, де як залежна змінна виступає цільова змінна – пропущене значення, яке необхідно відновити.

2. За отриманим на попередньому етапі рівнянням, у яке підставляються відомі значення незалежних змінних (предикторів), для кожного цільового об'єкта розраховується пропущене значення за залежною цільовою змінною. У випадку інтервальних та абсолютних змінних розраховується конкретне значення, а для порядкових і номінальних значень з деякою ймовірністю передбачається категорія, до якої має бути віднесений об'єкт.

Вибір типу регресійної моделі для розрахунку пропущених значень змінної визначається кількістю вимірів цільової залежної змінної (значення якої необхідно відновити) [93] і незалежних змінних, за якими будуть передбачатися пропущені значення.

У роботі [340] розглядаються можливості оцінювання пропусків даних за допомогою *Байєсівського компонентного аналізу та локального методу найменших квадратів* і порівнюються можливості їх сукупного використання. Показано, що спільне використання обох методів надає можливість отримати вищу якість прогнозів пропущених значень, але при цьому суттєво збільшуються обчислювальні витрати [93].

Зрозуміло, що коректність і ефективність роботи цих алгоритмів визначається підбором найбільш подібного значення до пропуску, а для цього необхідно враховувати причину пропуску даних. Сучасні комп'ютерні аналітичні системи, такі як SPSS, GeNIe, SAS Enterprise Miner, ґрунтуються на використанні логічних дерев для умовного обчислення значень та їх заміни на середні величини або медіани. Зокрема, для розробки скорингових карт рекомендуються методи підстановки [343], що враховують інші характеристики даних. Однак, присвоєння найбільш часто вживаних або середніх значень викличе так звані «сплески», що спотворить реальну ситуацію з розподілом груп у вибірках і призведе до втрати надзвичайно важливої інформації. Тому пропонується виносити пропущені дані в окрему групу, замінюючи пропущені значення певним спеціальним значенням.

Узагальнена класифікація методів заповнення пропусків, що використовуються в різних інструментальних засобах інтелектуального аналізу даних, наведена на рис. 1.6. Для докладного аналізу обрано середовище SAS Enterprise Miner, в якому реалізовано різні методи заміни пропущених значень, а також передбачена відсутність обов'язкової заміни. У SAS Enterprise Miner окремо передбачена можливість вставки для вхідних вузлів та цільових змінних, а також є можливість здійснювати таке заповнення не на всій вибірці даних, а спочатку на навчальній вибірці, і у випадку отримання задовільних результатів – розповсюдити таку заміну і на перевіірочну вибірку [93]. Розглянемо детальніше різні методи заповнення пропусків залежно від типу змінних (категоріальні або неперервні).

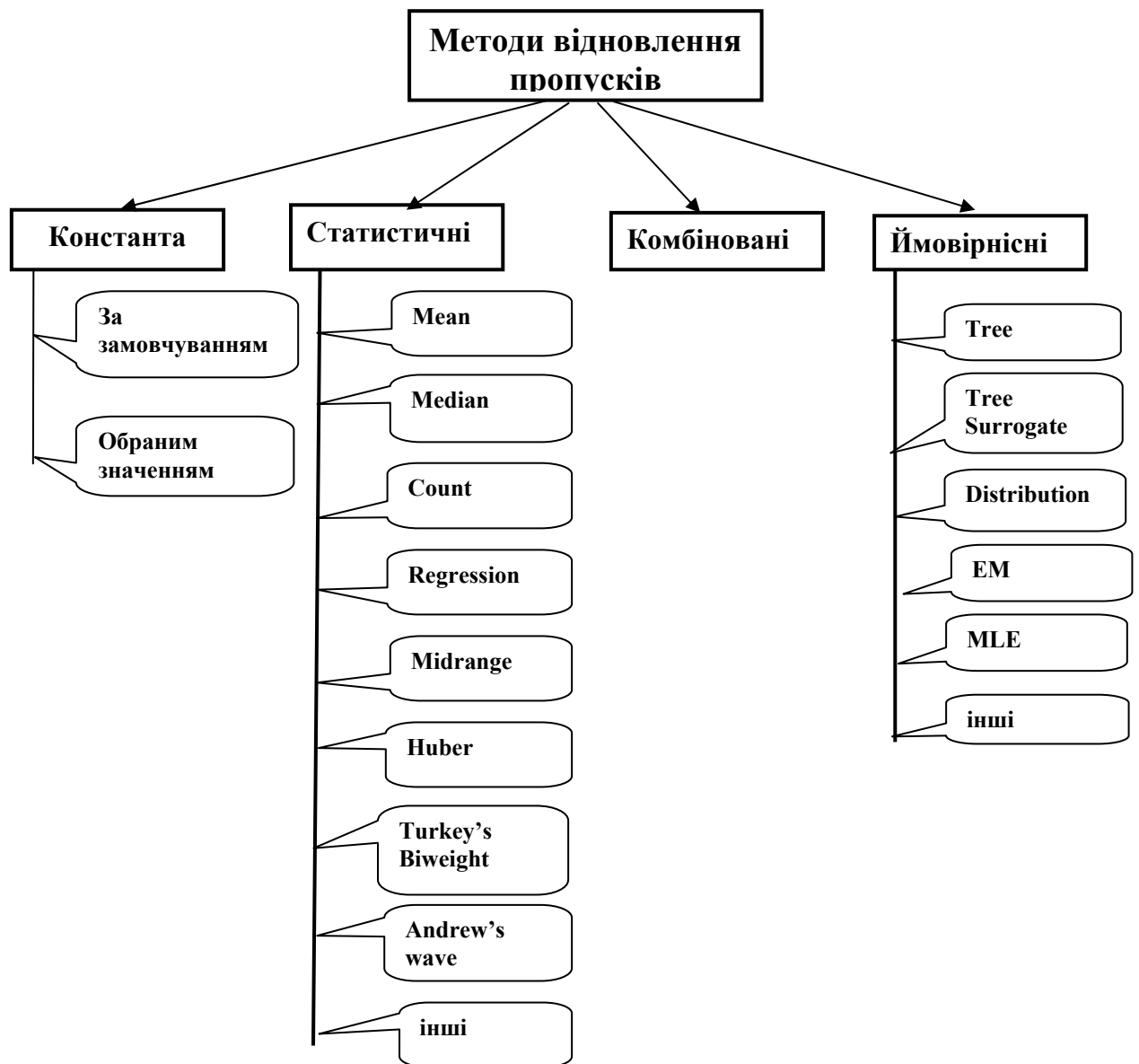


Рис. 1.6 Класифікація методів заповнення пропущених даних

Неперервні змінні:

Mean – заміна пропущених інтервальних значень середнім арифметичним. Це незміщена оцінка середнього популяції. Mean використовується, якщо значення змінної мають приблизно симетричний розподіл (наприклад, дзвоноподібний нормальний розподіл). Цей метод використовують за замовчуванням.

Median – використовується значення, яке є середнім значенням або середнім арифметичним двох середніх значень для множини чисел, розташованих у порядку зростання. Середнє і медіана рівні для симетричного розподілу. Медіана менш чутлива до екстремальних значень, ніж середнє [93], а

тому підходить краще для змінних, які мають спотворені розподіли. Вона використовується також і для порядкових даних.

Midrange – використовується параметр півсуми крайніх значень (середній діапазон) для заміни відсутніх неперервних значень змінної значенням суми максимального та мінімального значення для змінної, розділеної на два [93]. *Midrange* є скоріше відображенням тенденції; його легко розрахувати.

Distribution, Tree, Tree Surrogate – аналогічно як до категоріальних.

Mid-minimum Spacing – використовується середній мінімальний інтервал, застосовується числова константа для визначення пропорції даних, що включаються в інтервал [93].

Huber – метод, у якому для заміни пропущеного значення використовується наступна оцінка [322]. Якщо лінійна регресійна функція втрат швидко зростає зі збільшенням значень залишків визначена як $l(r) = \sum_i r_i^2$ то, альтернативним є використання абсолютного значення функції втрат замість квадрату залишків, тобто $l(r) = \sum_i |r_i|$.

Компромісом між цими двома функціями втрат стала запропонована Пітером Хубером у 1964 році функція виду [93, 322]:

$$l(r) = \sum_i \rho(r_i),$$

$$\text{де } \rho(r_i) = \begin{cases} r_i^2 & \text{if } |r_i| \leq c \\ c(2|r_i| - c) & \text{if } |r_i| > c \end{cases}.$$

Хубер вважав, що правильним вибором є значення $c = 1,345$, і показав, що асимптотично це 95%-й інтервал. Цей метод є ефективним, якщо реальний розподіл близький до нормального (і часто ефективніший у інших випадках).

Tukey's Biweight – метод, у якому оцінка для функції втрат визначається за критерієм *Tukey's Biweight* (відомим також як *Tukey's bisquare*)[322]:

$$\rho'(r_i) = \begin{cases} r_i \left(1 - \left(\frac{r_i}{c}\right)^2\right)^2 & \text{if } |r_i| \leq c, \\ 0 & \text{if } |r_i| > c. \end{cases}$$

Для цієї функції втрат зазвичай використовується значення $c = 4,685$; воно забезпечує асимптотичну ефективність на рівні 95%, так само, як і лінійна регресія для нормального розподілу [231].

Andrew's wave – метод, у якому оцінка визначається так:

$$w(r_i) = \begin{cases} \frac{c}{\pi r_i} \sin\left(\frac{\pi r_i}{c}\right) & \text{if } |r_i| \leq c \\ 0 & \text{if } |r_i| > c \end{cases}.$$

За замовчуванням $c = 1,34\pi$.

Default Constant – пропуск замінюється визначеним символом.

Порівняльний аналіз методів заповнення пропущених даних показав, що поняття «найкращого» методу є некоректним. Вибір методу суттєво залежить не лише від конкретної предметної області, де зустрічаються пропущені значення, а й від припущень щодо типу розподілу реальних даних. Найчастіше застосовується метод середнього для заміни пропущених значень, тобто припускається, що дані належать до нормального розподілу (а це скоріше виключення з правил). Заміна пропущених значень середнім, медіаною або іншою оцінкою є простішим способом, однак це може суттєво спотворити істинний розподіл вибірки, а тому можливі лише у випадку мінімального впливу на характер вибірки. Тому проблема подолання невизначеностей в задачах менеджменту ризиків фінансових систем залишається актуальною.

1.5. Змістовний аналіз проблеми досліджень

Виконаний у попередніх підрозділах аналіз існуючих методів та моделей оцінювання ризиків показав, що вони мають певні обмеження та недоліки, а отже потребують доопрацювання та покращення. У підрозділі 1.1 проаналізовано сутність ризиків, особливості системного підходу до їх аналізу, моделювання та оцінювання, описані принципи та методи управління ризиками.

Проаналізовано міжнародні та національні стандарти ризик-менеджменту. Визначено причини виникнення та надано класифікації ФР.

Показано, що джерелом ризиків у фінансових системах можуть бути неповнота, неточність, нечіткість та недостатня якість інформації. Показано вплив інформаційних ризиків на появу та якість оцінювання ФР. Запропоновано розробити методологію урахування інформаційних ризиків у складі ризиків фінансових систем. У цілому аналіз особливостей оцінювання ФР свідчить про наявність таких невирішених задач:

1. Відсутність системної методології опрацювання невизначеностей різної природи.
2. Необхідність урахування інформаційних впливів у процесі аналізу ФР.
3. Недостатня ефективність існуючих математичних методів та моделей для глибинного аналізу різнотипних фінансових ризиків.
4. Необхідність узгодження, інтеграції і спільного використання ідеологічно різних методів і технологій під час аналізу і оцінювання ФР.
5. Потреба оцінювання і менеджменту ризиків в динаміці з урахуванням надходження нових свідчень, даних та оцінок волатильності фінансових процесів системи.
6. Необхідність створення моделей для оцінювання та прогнозування фінансових ризиків, здатних до адаптації їх структури і параметрів у відповідь на зміну зовнішніх впливів, вимог, цілей, даних та критеріїв оцінювання ФР.
7. Відсутність інформаційної системи підтримки прийняття рішень на базі єдиної методології оцінювання ФР і прийняття рішень з урахуванням різнотипних невизначеностей і ризиків, з адаптацією до змін зовнішнього середовища.

Отже, для фінансової системи на основі статистичних даних $\{y(k)\} \in Dist(\bar{y}(k), \sigma_y^2(k)); \bar{y}(k) \neq const; \sigma_y^2(k) \neq const; k = 1, \dots, N$, з деяким розподілом $Dist(\cdot)$ необхідно побудувати нелінійні моделі досліджуваного фінансового процесу та динаміки його дисперсії загального виду:

$$y(k) = F_1[y(k-i), x(k), \theta_1, \varepsilon(k)], \quad i = 1, \dots, p < N;$$

$$h(k) = F_2[\sigma_y^2(k-j), x(k), \theta_2, v(k)], \quad j = 1, \dots, q < N,$$

де $x(k)$ – множина регресорів моделі; θ_1, θ_2 – вектори параметрів моделей; $\varepsilon(k)$ – випадковий процес, зумовлений зовнішніми випадковими збуреннями і похибками (шумом) вимірів; $h(k)$ – умовна дисперсія процесу; $v(k)$ – залишки моделі дисперсії.

Модель має задовольняти умовам адекватності за статистичними критеріями R^2 , $\sum e^2$, F-статистикою Фішера тощо.

Тоді ризик втрат для фінансової системи визначається як:

$R = f(y(k), h(k), \theta, \sigma(k), \alpha)$, де θ – вектор параметрів моделі, α – квантиль довірчого інтервалу; σ – волатильність (норма мінливості).

Розв'язок оптимізаційної задачі оцінювання ризику втрат для фінансової системи знаходиться з системи рівнянь:

$$y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) + b_1 k + \dots + b_m k^m + \varepsilon(k),$$

$$h(k) = a_0 + \sum_{i=1}^p a_i \varepsilon^2(k-i) + \sum_{j=1}^q \beta_j h(k-j) + v(k)$$

при обмеженнях $\bar{y}(k) < c_1$, $\text{var}[y(k)] < c_2$, де $c_1, c_2 = \text{const} < \infty$.

Для широкого класу фінансових систем, які можуть бути описані фінансовими нелінійними нестационарними процесами виду $\Theta(X, t) = F(x_i, t, \varepsilon)$, $\{\Theta(X, t)\} \in \mathfrak{R}^N$, X – вектор параметрів моделі, де $E[\Theta(X, t)] \neq \text{const}$, $\text{var}[\Theta(t)] \neq \text{const}$, необхідно розробити прогнозуючі моделі для оцінювання фінансових ризиків. Математичні моделі для оцінювання ризиків можуть бути представлені у таких варіаціях:

$$1) R(x) = P(x, t),$$

$$U(t) \leq U_{\text{крит}};$$

$$2) R(x) = U(x, t),$$

$$P(R(t)) \leq P_{\text{крит}};$$

$$3) R(x) = U\left(\sum_{i=1}^n x_i(t) p_i(x_i)\right) \cdot P(R(t)),$$

$$0 \leq U(t) \leq U_{\text{крит}}, \quad 0 \leq P(R(t)) \leq P_{\text{крит}},$$

де $R = \langle P(t), U(x(t)) \rangle$ – модель оцінювання ризиків; $0 \leq t \leq T$ – дискретний час; $x_i(t) \geq 0$ – множина пояснюючих змінних, що характеризують фінансовий процес; $0 \leq p_i(x_i) \leq 1$ – ймовірність реалізації x_i змінних-характеристик факторів, що описують процес; $U(t) = \sum_{i=1}^n x_i p_i$ – можливі втрати від реалізації ризику в часі; $i = \overline{1, \dots, n}$, $U(t) \leq U_{\text{max}}$ – обмеження капіталу на покриття ризиків фінансових систем; $\varepsilon(k)$ – випадковий збурюючий процес з довільним розподілом.

Серед сучасних методів оптимізації слід виокремити ті, які доцільно застосовувати для задачі мінімізації ризику. Оскільки можуть використовуватись математичні моделі довільної структури та різні види обмежень, то задача мінімізації ризику може бути задачею лінійного програмування, нелінійного програмування та безумовної оптимізації. Для розв'язання задачі лінійного програмування використовуються симплекс-метод, двоїстий симплекс-метод тощо. Багатопараметрична задача з обмеженнями типу лінійних нерівностей (а в загальному випадку і нелінійних нерівностей для задачі нелінійного програмування) може бути зведена до задачі безумовної оптимізації. Для задач, де обмеженнями виступає час, задача зводиться до дискретної оптимізації.

Способи розв'язання задачі безумовної оптимізації: лінійний пошук, без використання похідних (пошук обсягу витрат, що є допустимими для покриття ризику визначеної ймовірності), лінійний пошук з використанням похідної (коли задано явно вид функції ризику), градієнтні методи, методи байєсівської оптимізації. Метод байєсівської оптимізації дає можливість розв'язувати оптимізаційні задачі у випадку неявно заданих цільових функцій.

Висновки до першого розділу і постановка завдання дисертаційної роботи

Проведені дослідження засвідчили системний характер виникнення і реалізації ризиків у різних галузях фінансово-господарської діяльності. Аналіз розмаїття ФР представляє складну проблему, яка сьогодні вирішується на основі використання різноманітних галузевих стандартів, методів опрацювання специфічних ризиків, напрацьованих стратегій диверсифікації, уникнення, передачі ризиків тощо. Але дослідження також показали, що глибинного аналізу причин і варіантів реалізації ризиків майже не виконується, засобів швидкого опрацювання і оцінювання ризиків у більшості галузей не вистачає. Винятком є, у деякій мірі, банківська діяльність, але й тут не всі види ризиків однаково успішно опрацьовуються, про що свідчать, зокрема, й світові кризи. Отже, актуальною проблемою є розробка такої методології і інструментарію, які дозволяли б виконувати швидке опрацювання ризиків фінансової діяльності і отримувати прогностні оцінки щодо прийняття рішень. Оцінювання ФР зазвичай проводиться на основі великих обсягів статистичних даних, з певною невизначеністю як самих даних, так і засобів їх збору та обробки. Побудова будь-яких прогнозів також висуває досить строгі вимоги щодо якості і обсягів даних. Вимоги щодо швидкості, якості, правильності опрацювання ризиків у сучасному конкурентному середовищі весь час зростають. Тому актуальним є завдання створення нових засобів та способів підтримки прийняття управлінських рішень щодо опрацювання ризиків у фінансових системах.

Задача дисертаційної роботи полягає у створенні нової методології опрацювання ризиків у фінансових системах на основі розвитку системного підходу і методів інтелектуального аналізу даних, яка відрізняється глибинним аналізом природи, сутності і причин фінансових ризиків як міри невизначеності, комплексним підходом до обробки фінансових даних з урахуванням їх неповноти, побудовою прогнозних оцінок на основі інтегрованих динамічних методів аналізу даних у рамках єдиної технології моніторингу та менеджменту фінансових ризиків.

РОЗДІЛ 2

СУЧАСНІ ІНСТРУМЕНТИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ФІНАНСОВИХ СИСТЕМ

Необхідність розв'язання задач ризик-менеджменту у системах різного типу викликало стрімкий розвиток технологій та інструментарію для оцінювання та прогнозування ризиків з використанням сучасних методів інтелектуального аналізу даних. Дослідження останніх років компанії Gartner підтвердили актуальність і підвищений запит на такі засоби. З'являються нові компанії-розробники програмного забезпечення, укладаються масштабні контракти на придбання окремих продуктів та додатків світовими лідерами у розробці програмного забезпечення, такими як Microsoft та IBM, головними тенденціями стає можливість розвивати ІТ-продукт згідно запиту ринку та уявляти і забезпечувати високий рівень передбачення з появою нових викликів та задач. На передній план виходить необхідність забезпечення візуальної аналітики, надання коротких і зрозумілих звітів, зрозумілий та інтуїтивний інтерфейс, представлення надзвичайно великих обсягів даних та знань у різних вимірах, зрізах, зведених таблицях та простому і зручному для користувача вигляді. Підтримка різних стандартів, платформ, хмарний доступ, можливість встановлення на підприємствах та доступ з будь-якого куточка планети, завантаження і підтримка даних з різних баз, сховищ, ресурсів, обробка онлайн запитів в реальному часі – ось головні вимоги до програмного забезпечення для виконання бізнес-аналітики, оцінювання і прогнозування ризиків фінансових систем.

2.1. Інструментальні засоби аналізу даних і оцінювання ризиків у фінансовій сфері

Сьогодні фінансові установи використовують різноманітні програмні продукти для аналізу даних. Це найбільш відомі зарубіжні системи SAS, SAP, SPSS, а також власні розробки аналітиків і програмістів. Більшість з таких

продуктів застосовує розглянуті вище методи логістичної регресії, нейронних мереж, дерев рішень, нечіткі множини, мережі Байєса, тощо. Дослідницька аналітична компанія Gartner [262] щороку публікує звіти щодо ситуації на ринку інформаційних технологій, використовуючи дві лінійно прогресивні експертні шкали (рис. 2.1): повнота бачення (completeness of vision) та здатність до реалізації (ability to execute) [102].



Рис. 2.1 Магічний квадрант для бізнес-інтелектуальних та аналітичних платформ на лютий 2018 року

Цей магічний квадрант оновлюється кожного року і показує зміну лідерів та ключових гравців порівняно з минулим. Доволі довго утримували лідерські позиції такі інтелектуальні бізнес-платформи, як компанії SAS, SAP, IBM (продуктом якої є SPSS). Проте в 2018 році ситуація дещо змінилась.

Саме ці провідні аналітичні платформи розглянемо детальніше та порівняємо за їх функціональністю та можливостями для обробки та аналізу фінансових даних з метою аналізу і прогнозування фінансових ризиків.

Компанія Gartner розробила у 2017 році додатковий рейтинг «Розширені аналітичні та платформи машинного навчання» (“Advanced Analytic & Machine Learning Platforms”), як окремий від рейтингу “Business Intelligence and Analytic Platforms”. Існує значний збіг у основних гравцях рейтингів, які з'являються в обох, включаючи SAS, IBM, Microsoft, SAP, Alteryx і ряд інших [262, 278, 279, 335, 337] (рис. 2.2) . Крім того є нові компанії, такі як Tableau і Qlik [330, 352].



Рис. 2.2 Рейтинг компаній «Розширені аналітичні та платформи машинного навчання» за 2018 рік

Боротьба основних лідерів бізнес-платформ здійснюється неперервно. Так, минулого року IBM випередила SAS як основного лідера всіх постачальників у цьому сегменті. Цього року відбулося досить помітне зниження рівня здатності до виконання, а також повноти зору у компанії. Тим

не менш, з її величезною побудованою клієнтською базою, компанія IBM продовжує утримувати 9,5% ринку цього сегмента.

2.1.1. Програмні засоби ІАД компанії Microsoft

Корпорація Microsoft є одним з найбільших гравців у сфері розробки інтелектуальних бізнес-рішень та аналітики і протягом декількох років утримує позиції у «верхньому куті» лідерських аналітичних рішень [46]. Провідна платформа компанії, Power BI, реалізується на основі хмарних технологій і на Cloud Azure. Також існують можливості для рішень на підприємстві для окремих користувачів, або для створення потужними користувачами складних веб-додатків, використовуючи внутрішні джерела даних. Power BI є унікальним рішенням, оскільки дозволяє користувачам виконувати підготовку, виявлення даних та інформаційні панелі всередині одного дизайнерського інструменту. Платформа інтегрується з Excel і Office 365, і має дуже активну спільноту користувачів, яка розширює можливості інструменту.

Power BI пропонує вбудовану аналітику в двох різних ітераціях, як для незалежних постачальників програмного забезпечення і розробників, так і для підприємств. API на базі платформи Azure дозволяє користувачам додавати організації, навіть для вже розробленої програми. Незалежно від того, де містяться дані, Power BI дозволяє користувачам вставляти візуальні ефекти в додатки, портали, SharePoint і команди Microsoft [46].

Power BI пропонується в версіях “desktop” та “premium”, і може приєднуватися до електронних таблиць Excel, джерел даних на підприємстві, великих даних, потокових даних та хмарних сервісів [46]. Microsoft надає організації легкість використання при роботі з комплексним аналізом даних. Однак, обмежений хмарний підхід і неможливість обробляти дуже великі обсяги даних є певними обмеженнями у майбутньому для використання потенційними користувачами цього продукту при вирішенні складних бізнес-задач на великих компаніях та підприємствах.

2.1.2. Аналітична платформа Tableau

Tableau пропонує розширену візуальну та аналітичну платформу для бізнесу, і наразі займає лідерські позиції на ринку. Аналітичні рішення компанії доступні через три основні канали: Tableau Desktop, Tableau Server і Tableau Online [351]. Tableau може приєднуватись до великої кількості джерел даних і доступний в режимі «для підприємства» або «хмарному рішенні». Також розробник пропонує вбудовані можливості аналітики, і користувачі можуть візуалізувати та обмінюватися даними з Tableau Public.

Рішення для робочого столу забезпечує виявлення і обробку візуальних даних і підключається до великих даних, SQL-бази даних або хмарних додатків, таких як Google Analytics і Salesforce. Користувачі можуть отримати доступ і поєднувати розрізнені дані без написання коду, об'єднувати і керувати метаданими для оптимізації джерел даних. Tableau Server пропонує керований підхід до аналітики самообслуговування [351], що дозволяє виконувати централізоване управління метаданими та правилами безпеки. Інтегрується з існуючими протоколами безпеки; користувачам забезпечена безпека баз даних завдяки запитам дозволу на передачу даних і фільтрації.

Tableau Online повністю розміщений у хмарі, а тому організації можуть уникнути необхідності налаштування серверів, керування програмними оновленнями та масштабування можливостей апаратного забезпечення. Користувачі можуть підключитися до хмарних баз даних, таких як Amazon Redshift і Google BigQuery, і реальних баз даних на підприємстві або планувати виведення оновлень.

Згідно з дослідженнями компанії Gartner, Tableau досягнув значної зрілості на ринку бізнес-рішень, проте значна кількість обіцяних для підприємств можливостей ще знаходиться в розробці.

2.1.3. Qlik – рішення для бізнес-аналітики

Qlik пропонує широкий спектр інструментів BI та аналітики, які підпорядковані рішенню Qlik Sense. Рішення дозволяє організаціям об'єднувати

всі свої джерела даних у єдину форму. Двигун в пам'яті та асоціативна аналітика показує всі можливі варіанти ставлення до даних, і платформа доступна на підприємстві (Qlik Sense Enterprise) або в хмарі (Qlik Sense Cloud). Продуктовий портфель компанії Qlik [329] наповнений платформою Qlik Analytics (вбудовані та користувацькі інструменти), і QlikView, для перегляду даних.

Платформа QlikView призначена для усунення розриву між традиційними системами бізнес-аналізу та офісними додатками – кінцеві користувачі можуть вирішити аналітичні завдання (оцінка продажів та ефективності маркетингових програм, оптимізація сервісного обслуговування, тощо) майже без допомоги ІТ-спеціалістів, а розгортання рішення здійснюється за кілька днів або тижнів. Відкритість платформи QlikView Business Discovery дозволяє організувати обмін даними з іншими програмними продуктами. Консолідуючи інформацію з систем ERP або CRM, платформа QlikView надає наочні засоби візуалізації, що дозволяють відразу оцінити стан аналізованого об'єкта, наприклад, дізнатися поточну заповнюваність складу або визначити ризики фінансових операцій у конкретній галузі, регіоні, для конкретного клієнта [329]. Зазвичай такі дані «рознесені» по багатьох базах, поступають з декількох джерел, що ускладнює їх комплексний аналіз.

QlikView Business Discovery обробляє всі дані в оперативній пам'яті сервера, що дозволяє швидко управляти мільярдами записів, але, на відміну від традиційних бізнес-аналітичних систем, QlikView встановлює асоціативні зв'язки між даними різної природи, отриманими з різних джерел. Однією з найпопулярніших можливостей QlikView Business Discovery є «соціальний бізнес-аналіз» [329] – об'єднання засобів для колективного обговорення проблем і спільного пошуку рішень.

Платформа Analytics Qlik дозволяє користувачам створювати, розширювати та вставляти візуальну аналітику в існуючі додатки і портали. Вбудовані функції виконуються в межах загальної системи управління та

безпеки. Користувачі можуть створювати і вставляти Qlik як прості колажі або інтегрувати в додатки, інформаційні сервіси або платформи IoT.

2.1.4. Інтелектуальні рішення для бізнесу компанії IBM

Компанія IBM пропонує широкий діапазон інтелектуальних рішень для бізнесу (BI) та аналітичних можливостей за різними напрямками [278, 279].

Платформа Cognos Analytics – це інтегроване рішення для самообслуговування, яке дозволяє користувачам отримувати доступ для створення інформаційних панелей і звітів. IBM Watson Analytics пропонує функцію автоматичного навчання, що включає в себе автоматичне виявлення шаблонів, підтримку запитів природною мовою, а також вбудовані розширені можливості аналізу [278]. Cognos – це передовий аналітичний інструмент компанії IBM, а Watson Analytics забезпечує користувача доступом до нових засобів, що використовують штучний інтелект.

Інтерфейс IBM дозволяє користувачеві створювати контент «на льоту». The Visual Data Explorer автоматично рекомендує найкращі візуалізації для конкретних шаблонів даних. Моделювання даних здійснюється за допомогою повного веб-доступу, який дозволяє отримувати звітність напряму з джерел даних та створювати автоматичні моделі даних на основі ключових слів. Дані захищаються шарами дозволів, аутентифікації та історії. Додаткові можливості включають планування та сповіщення.

Інструменти IBM для інтелектуального бізнесу як програмне забезпечення можуть бути розгорнуті як на підприємстві (on-prem), так і у вигляді хмарного рішення через IBM Cloud. Повний аналітичний портфель також включає такі програмні засоби як SPSS Predictive Analysis, IBM Data Science Experience і IBM Planning Analytics tools [34, 278, 279].

2.1.5. SAP – аналітика та консолідація даних

Компанія SAP пропонує широкий діапазон інструментів BI та аналітики як для підприємств, так і бізнес-користувачів. SAP постачається або через

рішення на підприємстві (BusinessObjects Enterprise), або через хмарне рішення (BusinessObjects Cloud) над областю SAP HANA Cloud [335]. SAP також пропонує набір традиційних можливостей бізнес-інтелектуальних рішень для інформаційних панелей та формування звітів. Інструменти виявлення даних розміщуються в рішенні BusinessObjects, тоді як додаткові функції, включаючи візуалізацію самообслуговування, доступні через набір інструментів SAP Lumira. SAP є популярним як загальнодоступний стандарт BI, особливо якщо організація планує використовувати у подальшому пропозиції компанії.

Інформаційні панелі BusinessObjects дозволяють користувачам персоналізувати засоби оперування елементами інтерфейсу, інтерактивні датчики, діаграми та віджети.

SAP Lumira дозволяє організаціям створювати інтерактивні карти, графіки та інфографіки. Користувачі можуть імпортувати дані з Excel та багатьох інших джерел, виконувати візуальний аналіз BI з використанням інтуїтивно зрозумілих інформаційних панелей і надійного обміну думками та даними всередині підприємства [335]. Є можливості для підприємства автоматизувати підготовку даних, прогнозування, моделювання та розгортання завдяки інтегрованому прогнозуванню на основі скорингу для широкого спектру цільових систем. Користувачі можуть використати можливості розширеної візуалізації і інтегруватися з програмним середовищем R.

2.1.6. Аналітичні можливості рішень TIBCO

Можливості продукту TIBCO розширюються і варіюються від інтеграції даних та API-управління до візуальної аналітики, формування звітів та інтелектуального аналізу даних. Портфель компанії складається з TIBCO Spotfire і TIBCO Jaspersoft [359]. TIBCO Spotfire – це найсучасніша платформа компанії. Вона має інтерактивну візуалізацію, підготовку даних, управління корпоративним класом та розширені аналітичні можливості. TIBCO підтримує Jaspersoft вбудовану функціональність бізнес-рішень для формування традиційної звітності [346].

Spotfire Smart Data Catalog дозволяє користувачам візуально взаємодіяти з даними, автоматично встановлюючи взаємозв'язки, надавати рекомендації через вбудований штучний інтелект. Будь-який користувач може створювати найкращі візуалізації за допомогою Recommendations wizard, використовуючи попередній досвід.

Користувачі можуть підключатися та об'єднувати дані з файлів, реляційних і NoSQL баз даних, OLAP, Hadoop та веб-сервісів. Хмарні рішення, такі як Google Analytics і Salesforce, також підтримуються. Поточна передача даних, таких як IoT, соціальні мережі [334] та обмін повідомленнями можуть бути інтегровані з усіма іншими даними в контекстному аналізі у реальному часі.

TIBCO's Insight Platform [359] підтримує поточну передачу даних, а Spotfire може ідентифікувати історичні закономірності, які безпосередньо є у Hadoop і Spark. Можуть використовуватись існуючі шаблони для прогнозного аналізу. StreamBase і BusinessEvents підтримують багато можливостей, зокрема і інтеграцію коду на R, що використовується багатьма середовищами моделювання.

2.1.7. AgenaRisk – аналітичне рішення з використанням мереж Байєса

AgenaRisk – це середовище розробки та виконання аналізу на основі мереж Байєса. Індивідуальні байєсівські мережі можуть бути побудовані для вивчення параметрів з даних (наприклад, мультиномінальних, нормальних, бета-біноміальних, ієрархічних моделей [161] тощо) [195]. Для побудови таблиць ймовірності вузлів та параметрів використовується EM-алгоритм.

Таке рішення дозволяє [195] виконувати розробку байєсівських мереж з різними типами вузлів (логічні, неперервні, маркіровані, ранжовані, дискретні), побудову графіків ризиків та статистики та їх редагування. Є можливості обробки відсутніх даних, навчання мережі, зокрема на основі як EM-алгоритму, так і алгоритму з учителем. Для обробки ризиків можливе визначення об'єктів ризику, оголошених вузлами введення і виведення, імпорт моделі для

повторного використання об'єктів ризику, а також передача параметрів (статистичних даних) між об'єктами ризику.

2.1.8. RapidMiner – інструмент машинного навчання

RapidMiner Studio оснащений неперевершеним набором можливостей моделювання та алгоритмів машинного навчання як з учителем [249], так і без. Вони гнучкі, надійні і дозволяють концентруватися на створенні найкращих моделей для будь-якого варіанта використання [303].

Дане програмне забезпечення містить:

- байєсівське моделювання (наївні байєсівські моделі можуть оновлюватись, а, отже, добре підходять для великих наборів даних або для інтелектуального аналізу потокових даних).

- регресійний аналіз (лінійна, логістична, узагальнена лінійна модель, лінійний дискримінантний аналіз (LDA), квадратичний дискримінантний аналіз (QDA), регуляризований дискримінантний аналіз (RDA), покроковий вибір вперед і назад, ітеративний t-тест, векторна лінійна регресія, поліноміальна регресія, локальна поліноміальна регресія, гаусівські процеси) [194, 303, 354];

- кластеризацію (визначена користувачем кластеризація або автоматичний вибір кращих кластерів, k-Means, k-Medoids, X-Means, Cobweb, Clore, DBScan, EM, SOM, агломеративна кластеризація, кластеризація зверху вниз);

- нейронні мережі (гнучкі мережеві архітектури з різними функціями активації, глибинне навчання, автоматична оптимізація як швидкості навчання, так і коригування розмірів нейронних мереж під час навчання);

- дерева прийняття рішень (дерева класифікації та регресії (CART), CHAID, решітки, ID3, C4.5, випадковий ліс, багінг [232] та бустинг, багаторівневі дерева, дерева градієнтного бустингу (GBT);

- методи опорних векторів (більше 10 різних методів для підтримки векторної класифікації, регресії та кластеризації, функції ядра включають

радіальну базову функцію, поліноміальну, нейронну, Anova, гаусівську та мультікватратичну комбінацію);

- комбінування моделей.

2.1.9. STATISTICA – засоби для ризик-менеджменту

STATISTICA Risk Management містить передові аналітичні інструменти, які дозволяють компаніям, що надають фінансові послуги, отримувати більш вигідних клієнтів і зменшити ризики [346].

Застосування включає:

- ✓ повний спектр рішень – підготовка даних, побудова атрибутів, вага кодування свідчень, побудова оціночної карти, вибір та оцінка моделі, вибір точки відсікання – включені в один пакет програмного забезпечення;

- ✓ оптимізований процес: рішення Scorecard об'єднують різні інструменти, необхідні для забезпечення комплексного моделювання ризиків;

- ✓ доступні найбільш потужні алгоритми: не лише логічна регресія та пропорційні ризики Кокса, але також інші алгоритми інтелектуального аналізу даних, такі як k-середні, дерева рішення та нейронні мережі, які включені до моделі кредитного ризику;

- ✓ рефлексивні моделі для реального часу: Live Score® миттєво обробляє нових клієнтів і оновлює моделі кредитного ризику в короткі терміни, що стало можливим завдяки інтегрованим рішенням STATISTICA.

2.1.10. Angoss Predictive Analytics – платформа для прогнозної аналітики

Компанія Angoss є світовим лідером у наданні прогнозної аналітики для компаній, що прагнуть підвищити ефективність аналізу стосовно ризиків, маркетингу та продажів. Пакет Angoss включає в себе набір настільних, клієнт-серверних і великих програмних продуктів для аналізу даних та хмарних рішень [204]. Комплект програмного забезпечення для прогнозної аналітики включає в себе: KnowledgeSEEKER, KnowledgeSTUDIO і KnowledgeREADER

все з найкращими в своєму класі дерев прийняття рішень і дерев стратегій. Управлінські рішення KnowledgeCLOUD надають провідну аналітику для організацій, які потребують власної спеціалізації. Рішення також включають KnowledgeSCORE для оцінки лідерства та можливостей, FundGUARD для управління активами та ClaimGUARD для виявлення страхових випадків. Світові компанії, що надають фінансові послуги, страхові компанії, телекомунікаційні, роздрібні, високотехнічні, медичні організації використовують Angoss [204] для збільшення доходів, підвищення продуктивності продажів і підвищення ефективності маркетингу, а також зниження ризику та вартості.

Додаток дозволяє:

- виявляти цікаві закономірності, сегменти та узагальнену статистику;
- досліджувати дані та прогнозувати за допомогою механізму інтелектуального аналізу даних та інтелектуальної аналітики;
- візуалізувати отримані результати за допомогою інтерактивних діаграм;
- будувати дерева прийняття рішень;
- будувати дерева стратегій, які є розширенням дерев прийняття рішень, що дозволяють користувачам аналізувати ряд ключових показників ефективності для прогнозування стратегій [204];
- порівнювати та аналізувати тенденції для візуального відслідковування взаємозв'язків в часі та розподілу за наборами даних;
- створювати моделі логістичної регресії та методи обробки відхиленних заявок для скорингових карт кредитних заявок;
- будувати карту процесів для візуального контролю рядів аналітичного процесу на основі проекту.

2.1.11. Analytic Solver – інструмент бізнес-аналітики

Analytic Solver Comprehensive від Frontline – це інтегрований інструмент, який дозволяє бізнес-аналітику «робити все в Excel» з швидкою побудовою моделей замість програмування на R або Python. Він містить всі можливості

аналітичного інструментарію інтелектуального аналізу даних для візуалізації даних, машинного навчання, аналізу тексту та прогнозування часових рядів; можливості аналітичного рішення для моделювання за методом Монте-Карло та аналізу ризиків, дерева рішень, оптимізацію, стохастичне програмування, тощо [202].

Він може відображати дані або імпортувати текст із джерел, від багатовимірної бази даних PowerPivot від Microsoft, до баз даних Oracle [320], DB2 і SQL Server і кластерів Apache Spark Big Data. Містить компоненти для візуалізації даних, функції вибору та інструменти перетворення, кластеризацію та основні компоненти, асоціативні правила, експоненційне згладжування та моделі прогнозування ARIMA, дуже потужну лінійну і логістичну регресію, дерева класифікації та регресії, нейронні мережі [202]. Навчені моделі можуть використовуватися для «оцінювання» нових даних в базах даних SQL або для класифікації чи прогнозування в електронних таблицях моделей.

XLminer дозволяє користувачам виконувати у Excel: статистичний аналіз даних; імпорт даних з електронних таблиць; візуалізацію даних, з використанням широкої бібліотеки графіків; модифікацію даних; статистичні розрахунки; управління великими обсягами даних; обробку тексту; проведення інтелектуального аналізу; прогнозування часових рядів; регресію на великих наборах даних; пошук важливих змінних; збір даних з різних джерел, їх перетворення та аналіз.

2.1.12. OLAP-технології

Online Analytical Processing (OLAP) – складається з певного набору аналітичних технологій для розв’язання задачі аналітичної обробки даних (в основному фінансових) у реальному часі [321]. Бази даних OLAP спираються на багатовимірні моделі даних, що дозволяє вирішувати складні аналітичні питання за короткий проміжок часу.

Сьогодні немає єдиного підходу до розв’язання задачі опрацювання невизначеностей у сховищах даних (СД), що зумовлене розрізненістю і

несистематичністю досліджень; недостатністю методик проектування схем СД з урахуванням невизначеності; неповно досліджені питання аналізу невизначених даних; існуючі реалізації інформаційних систем аналізують і коректно опрацьовують лише певні типи невизначеної інформації [190].

Проведені дослідження [190] показали, що більшість кінцевих користувачів працюють в основному з агрегованими показниками. Проблема побудови сховищ даних і створення гіперкубу досліджена у роботах [170, 190]. Поточні дані регулярно використовуються і тому зберігаються на накопичувачах з швидким доступом, застарілі детальні дані можуть зберігатися на місткіших накопичувачах з повільнішим доступом. Наповнення гіперкуба даними за недостатньої кількості початкових даних призводить до утворення порожніх комірок. Обробка таких даних може надавати неадекватні результати.

Причини появи невизначеностей у сховищах даних і проблеми, які виникають у зв'язку з цим, детально досліджено Н. Шаховською у роботі [190] і наведені на рис. 2.3.

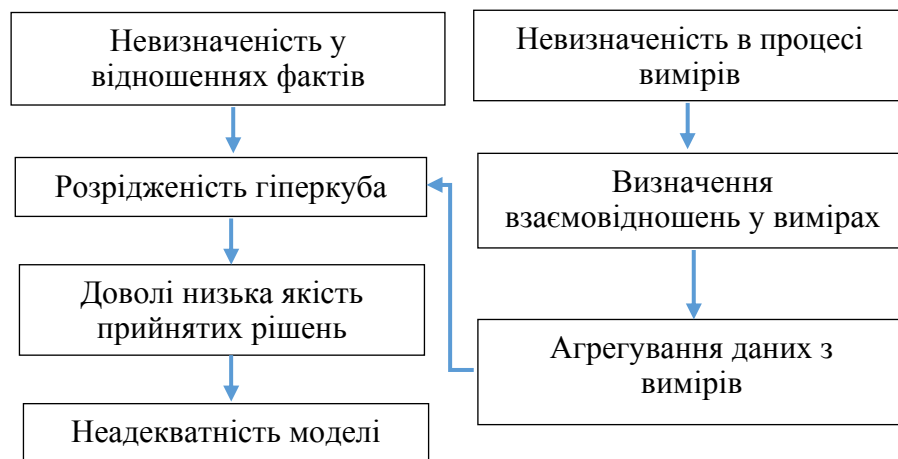


Рис. 2.3 Причини появи невизначеності у сховищі даних

2.1.13. Безкоштовні аналітичні засоби

Слід зазначити, що компанії-лідери розробники програмного забезпечення для інтелектуального аналізу даних надають можливості для безкоштовного ознайомлення зі своїми продуктами (найчастіше протягом 14 днів). Інколи також існують академічні ліцензії для провідних світових вищих

навчальних закладів в галузі статистичного та візуального аналізу даних. Це дозволяє компаніям розширювати перелік своїх клієнтів, які ознайомились з продуктами і переконались у його необхідності для розв'язання власних бізнес-задач, а також залучати нових користувачів, або ж майбутніх співробітників у штат своїх компаній серед студентів.

Крім цього існують відкриті рішення, які є безкоштовними і доволі популярними завдяки окремим своїм можливостям для різних задач. Перелік таких рішень доволі великий, постійно оновлюється, оскільки змінюються задачі для вирішення, або не всі витримують конкуренцію. Ці рішення є популярними через відкритість коду і можливості для безкоштовного користування, а подальша перспектива їх розвитку суттєво залежить від кількості залучених користувачів та їх зацікавленості у цих продуктах. На спеціальних форумах користувачі описують проблеми, з якими зустрічаються при розробці власних програм, обговорюють вирішення задач та алгоритмічні можливості.

Ось неповний перелік популярних безкоштовних програмно-технічних рішень для задач інтелектуального аналізу даних: Orange, Weka, Rattle GUI, Apache Mahout, SCAViS, RapidMiner, R, ML-Flex, Natural Language Toolkit, SenticNet API, ELKI, UIMA KNIME, Chemicalize.org, Vowpal Wabbit, GNU Octave, CMSR Data Miner, Mlpy, MALLET, Shogun, Scikit-learn, LIBSVM, LIBLINEAR, Lattice Miner, Dlib, Jubatus, KEEL, Gnome-datamine-tools, Alteryx Project Edition, OpenNN, AdaM, ROSETTA, ADaMSoft, Anaconda, Yooreeka, AstroML, streamDM, jHepWork, TraMineR, ARMiner, Arules, CLUTO and TANAGRA і т.ін. [205, 210, 269, 273, 284, 303].

Від того, наскільки активно компанії відчують ринок і реагують на пропозиції і звернення користувачів, насправді суттєво залежить і їх подальший розвиток. У великих компаній-лідерів з дорогим програмним забезпеченням можливості в цьому плані досить обмежені, оскільки не завжди їх код є відкритим або інтегрованим з іншими мовами програмування, а по-друге

кількість користувачів значно обмежена придбаними ліцензіями, академічними і тріальними ліцензіями.

2.2. Продукти компанії SAS Institute для обробки та аналізу даних

Компанія SAS Institute (Statistical Analytical System) – американська приватна компанія, яка займається розробкою програмного забезпечення та засобів бізнес-аналітики [337].

SAS надає рішення, які дозволяють [337]:

- оптимізувати аналіз кредитного ризику, впроваджувати своєчасну та точність звітності з урахуванням ризику (RWA) та оптимізувати нормативний капітал;

- відповідати вимогам Базеля III / IV [211] для аналізу кредитного ризику. Вимірювати вплив вимог до капіталу в майбутніх горизонтах часу кожного виду ризику та у всіх портфелях ризиків [306];

- створювати консолідовану базу даних, моделювання та звітності. Виконувати інтеграцію існуючих моделей ризиків та ієрархію даних у спрощену уніфіковану інфраструктуру даних для вимірювання та звітування за кредитами та ризиками контрагентів.

До основних модулів розроблених SAS рішень належать: робоче місце користувача (SAS/ASSIST, Enterprise Guide, Research Analyst), аналітичні можливості (SAS/STAT, Enterprise Miner, SAS/Insight, SAS/OR, SAS/IML, SAS/QC, SAS/ETS, SAS/Spectraview), готові рішення та продукти (FMS (CFO Vision), HR Vision, IT Service Vision, IT Charge Manager, Risk Dimensions, PH Clinical, PH Kinetics, тощо) [102].

2.2.1. SAS Enterprise Miner – базове аналітичне рішення компанії SAS

SAS Enterprise Miner (SAS EM) – це інтегрований компонент системи SAS, створений для виявлення у масивах даних інформації, необхідної для прийняття рішень. Розроблений спеціально для пошуку та аналізу прихованих закономірностей в даних (data mining) [207, 215, 230], Enterprise Miner включає

ефективні методи статистичного аналізу, відповідну методологію дослідження даних (SEMMA) та зручний графічний інтерфейс користувача [337].

Концепція SEMMA, представлена на рис. 2.4, містить такі основні етапи:

- ❖ **Sample** (Вибірка даних). Створення наборів даних для аналізу з джерел «сирих» даних.
- ❖ **Explore** (Дослідження даних). Розвідувальний аналіз даних, включає ряд алгоритмів «навчання без учителя» і різноманітні засоби візуалізації.
- ❖ **Modify** (Перетворення даних). Алгоритми перетворення даних, включаючи алгоритми зменшення розмірності, вибору значущих ознак.
- ❖ **Model** (Побудова моделей). Розробка різних типів моделей.
- ❖ **Assess** (Оцінка моделей). Вибір і порівняння моделей [102].

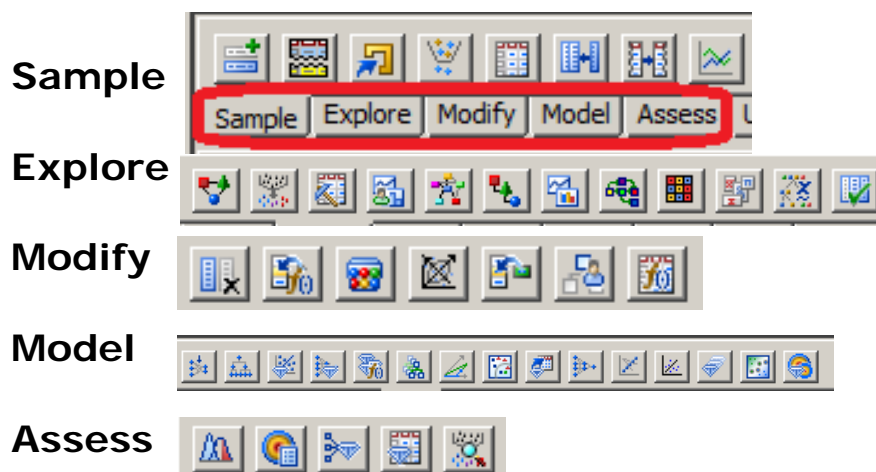


Рис. 2.4 Концепція SEMMA, вбудована в інтерфейс SAS Enterprise Miner

Інструмент SAS EM містить широкий набір методів ІАД – регресійний аналіз, дерева рішень, нейронні мережі, машина опорних векторів, кластерний аналіз та багато інших спеціалізованих та надсучасних методів. У дисертації описані методи та інструментарій, які використовуються для розв’язання практичних задач мінімізації ризиків, прогнозування відтоку клієнтів, розробки скорингових карт, задач аналізу шахрайства з кредитними картами, аналізу ризиків зниження фінансової стабільності, тощо. У роботах [174, 337, 343] описано можливість застосування SAS EM для прогнозування потреб у

ресурсах, зменшення часу простою активів, збільшення ефективності маркетингових кампаній, збільшення продажів товарів, тощо.

Клієнт SAS EM – це лише частина більш широкого набору програм. За своєю суттю клієнт SAS EM – це просто вікно з Java інтерфейсом. Фактично робота з аналізу даних виконується програмним продуктом, відомим як SAS Foundation, який підтримується ще одним програмним продуктом – SAS Metadata Server. *Metadata Server* відслідковує інформацію про доступ до даних і системної архітектури [102].

У конфігурації *Personal workstation* (персональна робоча станція SAS EM) компоненти: клієнт SAS EM, SAS Foundation, Metadata Server – знаходяться на одному головному комп'ютері [337]. Компоненти взаємодіють між собою за допомогою запатентованої компанією SAS технології інтегрованої об'єктної моделі (IOM).

В конфігурації *Enterprise Client* до загального набору доданий додатковий компонент *Analytics Platform* (Аналітична платформа). *Analytics Platform* маршрутизує зв'язок між клієнтом SAS EM, Foundation та серверами Metadata Server. Це дозволяє багатьом клієнтам з'єднуватись з більшістю серверів SAS Foundation [337]. Це найкраща конфігурація для розподілених клієнт / серверних додатків.

Системний адміністратор має встановити і сконфігурувати ці компоненти, зазвичай на декількох незалежних головних комп'ютерах. Слід зауважити, що всі дані зчитуються на SAS Foundation Server, а не на клієнтський персональний комп'ютер [174]. Аналітик має безпосередній доступ лише до інтерфейсу клієнта SAS EM.

Особливості SAS Enterprise Miner [102]:

- розвинуті бізнес-орієнтовані засоби порівняння моделей, побудова звітів, управління моделями, вбудовані можливості підтримки прийняття рішень;
- автоматизований процес застосування моделей «всередині» продукту і «ззовні» («генерація» коду, що реалізує ІАД процес [207]);

- «відкрита» розширена архітектура (можливе монтування власного коду);
- масштабовані обчислення (поки для частини методів);
- широкий набір вбудованих прикладних рішень (не входить у стандартний пакет).

2.2.2. SAS Credit Scoring for Banking

При аналізі фінансових ризиків особлива увага приділяється інформаційним технологіям обробки даних та сховищам, де зберігаються дані. Можливості інформаційної технології аналізу фінансових ризиків програмного продукту Credit Scoring for Banking [113] в основному зосереджені на розв'язанні задач кредитного скорингу [237, 240] і дозволяють: аналізувати клієнтів за різними розрізами характеристик, проводити розрахунки для кредитних продуктів, перераховувати скорингові карти без додаткових інвестицій, формувати алгоритми виявлення шахрайства, здійснювати збір і перетворення даних, формувати звітність.

В результаті моделювання на основі побудованих скорингових моделей формується скорингова карта, яка має бути інтегрована та протестована в рамках пілотного підрозділу банку. Після впровадження отриманої скорингової карти можливий перегляд списку параметрів кредитної анкети, видалення зайвих та введення нових параметрів, які також мають бути відображені у фронт-офісному додатку [92, 342]. Наступним кроком є автоматизація всього процесу розробки та застосування скорингових карт, включаючи збір даних, їх попередню обробку, підготовку аналітичних вітрин, оперативне оновлення і валідацію скорингових карт та впровадження їх на підприємстві (рис. 2.5).

Можливості аналітичного модулю для боротьби з шахрайством

Технічні можливості аналітичного програмного модулю SAS можуть використовуватись і в інших аспектах банківської діяльності, зокрема для запобігання умисним неповернення кредитів – боротьбі з шахрайством [113].

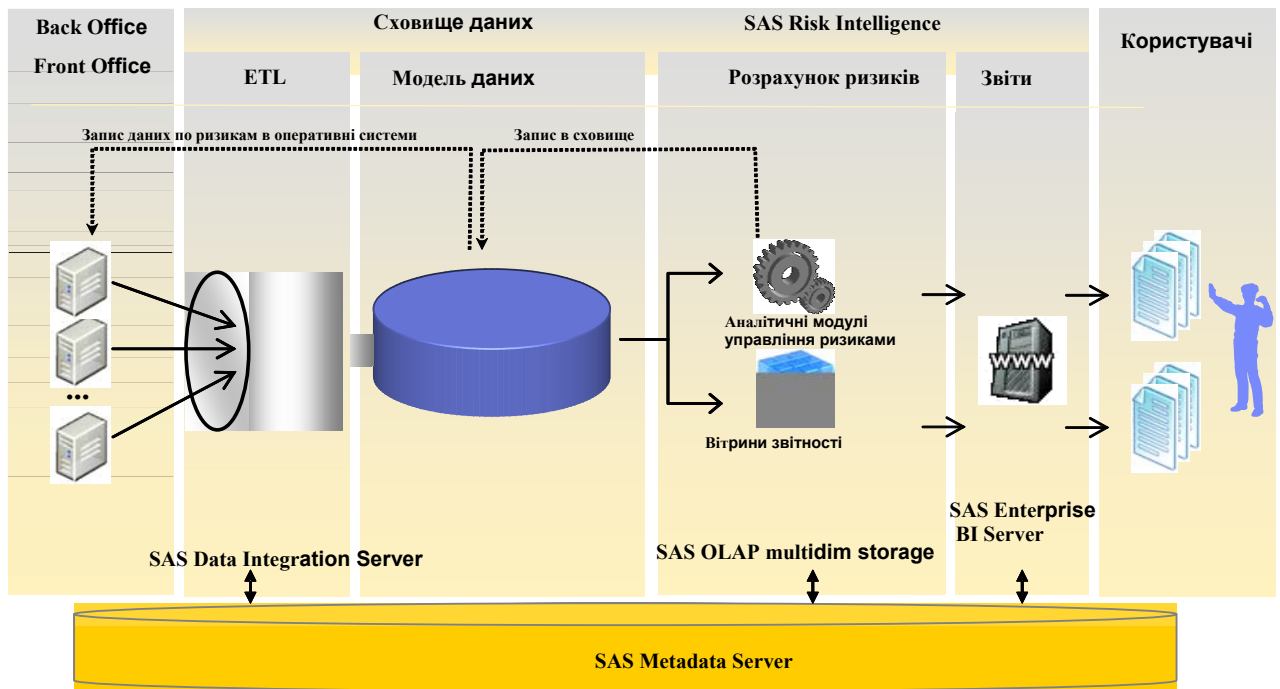


Рис. 2.5 Автоматизація процесу розробки скорингових карт у SAS

Способом зниження кредитних ризиків шахрайства, особливо для беззаставних продуктів, є розробка спеціальних моделей виявлення шахрайства (Application Fraud Scoring). Використання аналітичного модуля (інформаційної технології) SAS [337] для формування таких моделей спрямовано на допомогу банку у створенні послідовної і логічно підтвердженої бази для прийняття рішень, надаючи співробітникам кредитного відділу більш чітку, інтуїтивно зрозумілу міру кредитного ризику.

Проведений аналіз застосування інформаційних технологій компанії SAS для оцінювання фінансових ризиків [113, 174, 337, 338] показав, що можна забезпечити поширення описаної інформаційної технології обробки та аналізу даних на інші види фінансових ризиків. Зокрема, для фінансових ризиків кредитування можливе використання запропонованої технології для попереднього скорингу інших кредитних продуктів для різних регіонів та країн; реалізації скорингу поведінки позичальників та оцінювання дефолту (collection скорингу); для виконання сегментації клієнтської бази; побудови моделей для підвищення рівня перехресних та додаткових продажів продуктів, а також мінімізації відтоку клієнтів.

2.3. SPSS (PASW) Statistics – інструментарій для аналізу даних

SPSS Statistics (англ. «Statistical Package for the Social Sciences» – «статистичний пакет для соціальних наук») – це потужний програмний інструмент статистичного аналізу даних, один з лідерів в області комерційних статистичних продуктів, розроблений для проведення прикладних досліджень в соціальних науках [34], у 2009 році назва SPSS була змінена на PASW (Predictive Analytics SoftWare) Statistics і придбана компанією IBM.

SPSS Modeler зараз є аналітичною платформою від IBM, яка забезпечує інтелектуальний аналіз для повсякденних бізнес-задач [279]. Рішення надає широкий спектр розширеної аналітики, включаючи текстову аналітику [327], аналіз сутностей, аналіз соціальних мереж, автоматичне моделювання, репарацію даних, прийняття рішень та оптимізацію.

Багато організацій використовують прогностичний аналіз для маркетингових кампаній [215, 253]. SPSS Event Builder допомагає ефективно виконувати декілька задач, комбінуючи прогностичні моделі, створені аналітиками, з бізнес-логікою та методами оптимізації, щоб визначити кращий канал, пропозицію та повідомлення для кожного клієнта. Основні можливості системи такі [34]:

- розробка ефективних стратегій маркетингових кампаній: вибір критеріїв включення або виключення клієнтів у кожній кампанії на основі правил чи прогностичних моделей; визначення розподілів каналів, пропозицій, повідомлень або дій і призначень для кожної кампанії; визначення пріоритетів і шляхів оптимізації на основі цілей та обмежень;
- оптимізація результатів [34]: ручне налаштування обмежень і розподілів, перегляд результатів з використанням та без обмежень; автоматизація алгоритмів оптимізації розподілу;
- розгортання результатів: створення звітів щодо планування; запис списків в базу даних або файлову систему [34, 279].

Слід зазначити, що можливості продукту і наявність додаткових модулів можуть варіюватись в залежності від версії продукту. Існує базова версія (модуль), що містить широкий набір засобів для обробки та аналізу даних.

Базовий модуль SPSS (SPSS Base) включає процедури введення, відбору та корегування даних, а також більшість представлених у SPSS статистичних методів [279]. Крім безпосередньо методик статистичного аналізу, таких як частотний аналіз, розрахунок статистичних характеристик, таблиць спряженості, побудови графіків, цей модуль також включає t-тести і велику кількість непараметричних тестів, а також ускладнені методи, такі як багатовимірний лінійний регресійний аналіз, факторний аналіз, кластерний аналіз, дискримінантний аналіз, дисперсійний аналіз, аналіз надійності, багатовимірне шкалування, аналіз пропущених значень, тощо [34, 279].

Регресійні моделі (Regression Models). Включає в себе різноманітні методи регресійного аналізу, такі як бінарна і мультиномінальна логістична регресія, нелінійна регресія і пробіт-аналіз [17, 26, 34, 102].

Advanced Models. Містить методи дисперсійного аналізу (багатовимірний, з урахуванням повторних вимірів), аналіз виживання, загальну лінійну модель, метод Каплана-Майєра [138], регресію Кокса [251, 252], лог-лінійні, а також логіт-лог-лінійні моделі [279].

Amos (Analysis of moment structures – аналіз моментних структур) забезпечує аналіз даних на основі лінійних структурних рівнянь. Перевіряє наявність зв'язків між різними характеристиками випадкового процесу і оцінювання їх коефіцієнтів [102]. Перевірка здійснюється у формі аналізу причин і траєкторії, вимагаючи від користувача задання графічної моделі, у яку разом з даними можуть бути включені приховані елементи.

AnswerTree (дерева рішень) містить методи автоматизованого розділення даних на окремі групи або сегменти за частотним розподілом цільової (залежної) змінної [102]. Це використовується, наприклад, для профілювання покупців при дослідженні споживчого ринку [236, 279].

Categories використовується для аналізу категоріальних даних на основі методу відповідностей, аналізу однорідності, нелінійного аналізу головних компонент, нелінійного кореляційного і класичного кореляційного аналізу.

Clementine – це додаток, який містить різні засоби побудови математичних моделей ІАД [102], можливості візуального представлення, оцінювання якості та порівняння моделей.

Conjoint застосовується для виявлення найбільш привабливих продуктів.

Data Entry використовується для швидкого введення і очистки даних. Формується вибірка питань і відповідей, які потім використовуються в якості міток змінних і значень.

Exact Tests застосовується для обчислення точного значення ймовірності похибки (величини p -value) в умовах обмеженості даних при перевірці гіпотези по критерію χ^2 та непараметричних тестах [102]. Може застосовувати і метод Монте-Карло.

GOLDMineR. Програма містить спеціальну регресійну модель для аналізу впорядкованих залежних та незалежних змінних.

SamplePower. Зручний додаток для визначення оптимального розміру вибірки для більшості моделей статистичного аналізу [279].

SPSS Missing Value Analysis. Використовується для аналізу та встановлення закономірностей щодо пропущених значень і їх відновлення.

Trends. Містить методи на основі часових рядів, інтегрованих моделей ARIMA, сезонної декомпозиції і спектрального аналізу [34].

2.4. Скорингові карти ризику

2.4.1. Застосування скорингових моделей для оцінювання фінансових ризиків

Скорингові моделі були запропоновані для оцінювання банківських ризиків, розробляються на основі статистичних даних, що характеризують об'єкт та фактори, які спричиняють ризики [85, 103, 108 – 110, 282, 345].

У процесі скорингу використовуються наступні характеристики: порогову або «бар'єрну» оцінку, оцінку банкрутства, скоринговий бал, ймовірності певних подій і т.ін. Оскільки кількість клієнтів банку зазвичай є доволі великою, а виникає необхідність у періодичній перевірці їх

платоспроможності (скоринг поведінки), то потрібний наочний інструмент для швидкого ранжування клієнтів банку за платоспроможністю [86, 110]. Для цього банком будується скорингова матриця, де в залежності від сумарного скорингового балу банку, встановленого бар'єрного балу та балу банкрутства швидко приймається рішення щодо платоспроможності клієнта.

Скорингова матриця може бути трьохвимірною, для її побудови робиться збалансований вибір між високою оцінкою одного виміру та низькою оцінкою іншого; формується як невідома або «сіра» зона і залежить від пріоритету та постановки задачі (табл. 2.1).

Таблиця 2.1

Скорингова матриця

		Відток				
		0-549	550-619	620-649	650-699	700 +
Оцінка	0-189	Відмовити	Переглянути	повторно Схвалити		Відмовити
	190-209					
	210-229					
	230-249					
	250 +					

Кредитний скоринг – це система оцінювання кредитоспроможності особи, що основана на чисельних статистичних методах [33, 106, 203]. На основі певних соціально-демографічних та кредитних параметрів позичальника розробляють скорингову математичну модель оцінки повернення кредитів [24]:

$$S = F(w^j, x_i^j), \quad (2.1)$$

де w^j – ваги параметрів x_i^j .

За скоринговою моделлю для кожного набору параметрів конкретного клієнту та кредитного продукту визначаються скорингові бали [108] і встановлюють відповідний поріг відсікання, вище якого заявку на кредит схвалюють, а нижче якого – відхиляють:

$$score > Cut - off(ApprovalRateMin), \quad (2.2)$$

де $score$ – скоринговий бал, обрахований за характеристиками позичальника та параметрами кредиту; $Cut - off(ApprovalRateMin)$ – поріг відсікання,

встановлений в банку, тобто мінімальне значення скорингового балу для ухвалення кредитної заявки.

Формально існує декілька критеріїв для вибору оптимального порогу відсікання [287, 343], наприклад на основі максимізації значення статистики Колмогорова-Смірнова, мінімізації вартості неправильної класифікації клієнтів, максимізації істинно позитивних випадків, максимізації точності, кумулятивної вигоди [343].

Для зручності роботи кредитних спеціалістів розробляють спеціальні скорингові карти, які є узагальненням скорингових оцінок (2.1), отриманих за існуючою статистикою у банку та експертних знань, що базуються на досвіді експертів в галузі кредитування та рекомендацій НБУ.

Визначення. Скорингова карта (СК) є візуальним представленням роботи логічних алгоритмів побудови скорингових моделей [110]. Серед методів для побудови скорингових карт, зокрема, використовуються: логістична регресія, нейронні мережі, дерева рішень, мережі Байєса, нечіткі підходи, тощо [103, 137, 213, 237, 240, 270, 287]. Скорингова карта є модифікацією експертного підходу [162], де в якості експертних оцінок використовуються скорингові бали, обчислені за допомогою скорингових моделей. В тих випадках, коли скорингова карта розробляється для оцінювання шансу настання ризику при визначених сумах балів, скоринговий бал розраховується наступним чином [106]:

$$\begin{aligned}
 score &= \log(odds) * factor + offset = -\left(\sum_{j,i=1}^{k,n} (woe_j * \beta_i) + a\right) * factor + offset = \\
 &= -\left(\sum_{j,i=1}^{k,n} \left(woe_j * \beta_i + \frac{a}{n}\right)\right) * factor + offset = \\
 &= \sum_{j,i=1}^{k,n} \left(-\left(woe_j * \beta_i + \frac{a}{n}\right) * factor + \frac{offset}{n}\right),
 \end{aligned} \tag{2.3}$$

де $odds$ – розподіл шансів, тобто співвідношення «хороших» і «поганих» випадків, $factor$ – фактор або коефіцієнт моделі, $offset$ – зміщення, β_i – коефіцієнти регресії; woe_j – оцінки WOE для атрибутів моделі; n – кількість

характеристик в моделі; a – відрізок, що відсікається на осі логістичною регресією; k – кількість груп (атрибутів) для кожної характеристики.

За результатами оцінювань формується таблиця скорингових балів, які обраховуються за кожною характеристикою клієнту та кредитного продукту в залежності від реального значення, які вони приймають, а потім підраховується сумарний скоринговий бал для кожного конкретного клієнта (рис. 2.6).

		Group	Scorecard Points	Weight of Evidence	Event Rate BAD = 1	Percentage of Population	Coefficient
age_oldest_tr	age_oldest_tr < 41	1.00	18	-0.76	10.39	9.29	-0.32
	41 <= age_oldest_tr < 143	2.00	23	-0.19	6.14	38.64	-0.32
	143 <= age_oldest_tr < 196	3.00	27	0.22	4.18	19.57	-0.32
	196 <= age_oldest_tr	4.00	32	0.79	2.41	29.30	-0.32
	MISSING	5.00	19	-0.58	8.83	3.20	-0.32
bureau_score	_MISSING_	5.00	10	-0.65	9.41	4.77	-0.78
	bureau_score < 630	1.00	-1	-1.14	14.55	9.13	-0.78
	630 <= bureau_score < 662	2.00	9	-0.67	9.62	13.99	-0.78
	662 <= bureau_score < 723	3.00	29	0.21	4.20	38.23	-0.78
	723 <= bureau_score	4.00	58	1.47	1.23	33.89	-0.78
ltv	_MISSING_	5.00	-92	-4.01	100.00	0.01	-1.00
	ltv < 91	1.00	44	0.67	2.71	28.37	-1.00
	91 <= ltv < 102	2.00	26	0.04	4.97	30.82	-1.00

Рис. 2.6 Приклад скорингової карти

Наприклад, для змінної «age_oldest_tr» (давність останньої угоди) позичальник з угодою, що здійснюється менше ніж 41 місяць (3,5 роки) отримує 16 скорингових балів, від 41 до 123 – 23 бали, 143..196 – 27 балів, більше 196 – 32 бали, і у випадку пропущених значень (наприклад, не було жодної угоди) – 19 балів.

Аналогічно розподіляються скорингові бали в залежності від значень, що приймають інші змінні [106]. Таким чином, скорингова таблиця є зручним, наочним і універсальним інструментом, дозволяє автоматизувати процес аналізу заявок і зменшити час на прийняття рішення щодо видачі кредиту.

Залежно від галузі підприємства, для якого необхідно аналізувати ризики, розроблюються скорингові карти: для оцінки ризиків, CRM або маркетингові скорингові карти та комбіновані [106,108].

Найбільш поширеними є скорингові карти для оцінювання кредитних ризиків [166, 287, 302, 328]. Це пов'язано з великою кількістю задач, які постають перед банками і потребують періодичного вирішення, та великою кількістю клієнтів та продуктів, для яких необхідна періодична перевірка та відслідковування. Прикладами таких задач є:

- оцінка кредитоспроможності претендентів на отримання нових кредитів (application scoring),
- оцінки ймовірності повернення вже виданих кредитів (behavioral scoring),
- відслідковування можливості повернення кредиту клієнтом при порушенні ним строків та умов погашення кредиту (collection scoring),
- оцінка ймовірності шахрайства з боку існуючого клієнта банку (fraud scoring).

2.4.2. Скорингова карта поведінки

Скорингова карта поведінки – це математична модель, яка характеризується сукупністю вхідних факторів (характеристик) клієнту та продукту (наприклад, кредиту), які змінюються в часі і впливають на цільову характеристику – змінну, яка характеризує можливість поведінки клієнта в майбутньому (наприклад, сплати за кредитом у поточному місяці). Така поведінкова скорингова карта [120] будується для різних «типових» позичальників і дозволяє оцінювати ймовірність щомісячної сплати заборгованості. Паралельно з цим банки зацікавлені в об'єктивній інформації щодо можливих втрат, які виникнуть внаслідок несплати по кредитах. У відповідності до Базеля II для оцінювання кредитних ризиків банків використовується IRB-підхід (Internal Ratings-Based Approach) з урахуванням внутрішніх рейтингів позичальників, тобто рейтингів, встановлюваних самими банками [212, 213, 307]. Такий підхід передбачає можливість розрахувати суму,

сплачену за кредитом, і суму, непокриту кредитом, для кожного конкретного кредиту у конкретний момент часу.

Введемо поняття «цикл успішного обслуговування кредиту», яке визначається як кількість місяців або днів, коли здійснюється сплата кредиту без прострочок, тобто $0 < t_{goodcredit} < 12$ при умові, що затримка $delay = 0$ днів, $P(delay = 0) = 1$.

Поява прострочки більш 3 днів, але менше 30 днів, визначається спеціальним маркером «підозріла» поведінка $I(behavioral_debts) = 1$ і вважається приводом для включення таких клієнтів у періодичний (більш частий) моніторинг з використанням скорингових карт поведінки [120] і можливих засобів захисту або протидії появі подальших прострочок за кредитом.

2.4.3. Методологічні аспекти розробки та впровадження скорингових карт

Розробка скорингових карт передбачає виконання певних вимог до вхідних даних, стандартів формування вибірок даних та обов'язкового набору учасників.

Ефективність використання розробленої скорингової карти суттєво залежить від наступних факторів [108]:

- ✓ якість доступних даних;
- ✓ тип очікуваного результату, тобто двійковий («хороший»/«поганий») або неперервний (прибуток/збитки в національній валюті);
- ✓ розмір доступних вибірок;
- ✓ платформа для впровадження (тобто чи можливе впровадження певної скорингової карти в системі обробки заявок);
- ✓ можливість інтерпретації результатів, наприклад зручність обслуговування бальної системи оцінки скорингових карт, побудованих на основі регресії;
- ✓ відповідність методології правовим нормам;

✓ можливість відслідковування і діагностування ефективності скорингових карт [106].

Під час розробки скорингових карт банк формує команду фахівців для коректної розробки з урахуванням останніх аналітичних досліджень, використанням новітніх інформаційних технологій для основних напрямів діяльності банку [343]. Команда фахівців складається з розробника, ризик-менеджера продукту чи портфелю карт, менеджерів продукту, менеджерів операційної діяльності, керівника проекту, менеджерів інформаційних технологій, співробітників відділу управління ризиками та юридичного відділу. Вони взаємодіють, узгоджують спірні питання з керівником проекту та передають інформацію розробнику скорингових карт.

Етапи розробки скорингових карт наведені нижче.

Етап 1. Підготовка і планування (укладання бізнес-плану, визначення організаційних цілей та ролі скорингової карти, вибір між внутрішньою та зовнішньою розробкою і визначення типу скорингової карти, розробка плану проекту; визначення ризиків проекту, складу учасників проектної групи та їх обов'язків) [109].

Етап 2. Аналіз даних та параметрів проекту (аналіз доступності і якості даних, збір даних для визначення параметрів проекту, оцінювання параметрів проекту, визначення показового періоду та «вікна вибірки», встановлення категорій поведінки (цілі), виключення, сегментація, методологія, перегляд плану впровадження).

Етап 3. Створення бази даних для розробки моделі (формулювання вимог до вибірки, формування вибірки, збір та побудова навчальної вибірки, поправка на апріорні ймовірності (факторинг)) [108].

Етап 4. Розробка скорингової карти (дослідження даних, визначення пропущених значень та викидів, виявлення кореляції, аналіз вихідних характеристик, створення попередньої скорингової карти, аналіз відхилених заявок, фінальний етап створення скорингової карти, визначення шкали скорингової карти, вибір і контроль скорингової карти) [109].

Етап 5. Управлінські звіти за скоринговою картою (таблиця виграшу (gains tables), звіти по характеристикам).

Етап 6. Впровадження скорингових карт (контроль перед впровадженням, розробка стратегії, стратегії оцінювання, встановлення рівнів відсікання, правила кредитних політик, коригування).

Етап 7. Дії після впровадження (звіти для моніторингу ефективності скорингових карт та кредитного портфелю, управлінські звіти по скорингових картах, звіти для моніторингу ефективності кредитного портфелю) [108].

Послідовність розробки та налаштування скорингової карти представлена на рис. 2.7.

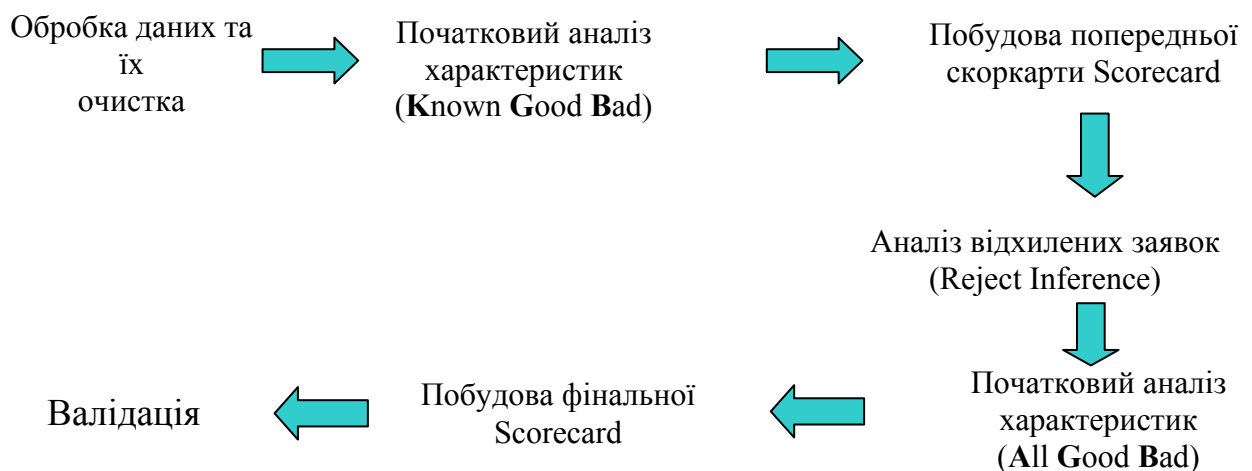


Рис. 2.7 Послідовність розробки скорингової карти

Початковий аналіз характеристик передбачає створення зв'язків на основі бізнес-логіки шляхом групування атрибутів, які перевищують мінімальний критерій інформативності. Альтернативний, суто статистичний підхід, передбачає встановлення тільки таких зв'язків, які збільшують інформативність чи інші показники [343].

Урахування особливостей бізнесу є коректним і надає певні переваги:

➤ логічні зв'язки гарантують, що результуючі ваги після регресії будуть мати сенс. Також гарантується, що після присвоєння атрибутам балів в процесі створення скорингової карти ці бали будуть логічними (наприклад, люди більш старшого віку зазвичай набирають більше балів, ніж молодші) [106];

➤ логічні зв'язки забезпечують участь у процесі кінцевих користувачів та співробітників операційного відділу, а скорингова карта, підтверджена досвідом різних підрозділів, забезпечує більш високий рівень достовірності;

➤ логічні зв'язки підтверджують бізнес-досвід, тому вони, як правило, адекватніші за суто статистичну оцінку. Використання бізнес-досвіду для удосконалення прогнозного моделювання робить модель більш придатною для бізнес-практики [108];

➤ найважливіша роль процесу узагальнення зв'язків шляхом їх логічного групування – це зниження ризику перенавчання. В результаті не буде необхідності включати в модель кожен випадкову зміну даних, призначаючи необмежену кількість ваг незгрупованим атрибутам. Тепер можна ранжувати ризики, моделювати тренди, застосовувати скорингову карту до вхідної популяції з певною еластичністю (є можливість деякої зміни в популяції) і забезпечити її стабільність протягом довшого періоду.

Існують спеціальні автоматичні засоби для розробки та обрахунку скорингових карт. Так, в інформаційній технології SAS Enterprise Miner існує спеціальна компонента Scorecard, яка відповідає за автоматичний розрахунок скорингових карт за результатами регресійної моделі, побудованої за даними початкової вибірки [337, 343]. Вона надає сукупність звітів за статистичними показниками за якістю (предикативною здатністю) побудованої скорингової карти і дозволяє визначити оптимальний бал відсікання: діаграму розподілу скорингових балів (Score Distribution), таблицю зі статистикою за групами скорингових балів (Gains Table), групу звітів Strength, звіти про якість моделі: графік зі статистикою Колмогорова-Смірнова для різних балів відсікання показує предикативну точність моделі (ROC Plot), графік, що показує процентну кількість прогнозованих «дефолтів», накопичених в сумі для різних балів відсікання (Captured Event Plot), табличний звіт зі статистичними показниками, що характеризують предикативну точність [164].

Характеристика *WOE* (weight of evidence, або вага спостереження) визначає силу атрибутів змінної-характеристики у розділенні хороших і поганих

випадків кредитування, тобто наскільки коректно сформовані інтервали для даної характеристики. Формально вона порівнює частку позитивних і негативних випадків для кожного рівня (інтервалу) значень атрибутів:

$$woe_j = \ln\left(\frac{Distr\ Good_i}{Distr\ Bad_i}\right). \quad (2.4)$$

На основі обчислених значень *WOE* обраховується ще одна характеристика, так зване інформаційне значення, або *Information Value (IV)*. *IV* оцінює силу характеристики стосовно передбачення, тобто здатність розділяти приклади на позитивні та негативні; воно формально визначає предикативну силу характеристик і обчислюється за формулою [164]:

$$\begin{aligned} IV &= \sum_{i=1}^L (DistrGood_i - DistrBad_i) * woe = \\ &= \sum_{i=1}^L (DistrGood_i - DistrBad_i) * \ln\left(\frac{Distr\ Good_i}{Distr\ Bad_i}\right), \end{aligned} \quad (2.5)$$

де L – це кількість атрибутів (рівнів або проміжків) для даної характеристики. Існують наступні правила вибору найбільш впливових характеристик за *IV*:

- ❖ < 0.02 – сила передбачення відсутня;
- ❖ $0.02 - 0.1$ – мала сила передбачення;
- ❖ $0.1 - 0.3$ – середня сила передбачення;
- ❖ > 0.3 – велика сила передбачення.

Для аплікаційного скорингу, на відміну від біхевіоріального, на етапі побудови скорингових карт застосовується етап аналізу відхилених заявок (*Reject Inference*) [164], що передбачає доповнення існуючої навчальної вибірки за рахунок даних по претендентам, яким було відмовлено у видачі кредиту, і водночас автоматично проводиться розмежування претендентів на позитивні та негативні випадки.

2.4.4. Розробка нейро-нечіткого методу доповнення відхиленими заявками при моделюванні фінансових ризиків

При розробці скорингових карт важливо використовувати як навчальну вибірку не лише дані, що характеризують появу ризиків, а й здійснювати аналіз відхилених заявок, тобто присвоєння елементам вибірки, по яким невідоме бінарне значення цільової змінної, деякого значення та подальше їх включення у навчальну вибірку для стійкості фінальної моделі відносно генеральної сукупності [343].

Існуючі методи аналізу відхилених заявок можна класифікувати як методи: присвоєння негативного результату відхиленим заявкам; доповнення у вигляді пропорції; рівного співвідношення для прийнятих заявок; повного ігнорування відхилених заявок; тимчасового погодження всіх заявок для збору даних; використання аналогічних даних банку або кредитних бюро; простого доповнення («жорстке відсікання»); доповнення для експертного процесу прийняття рішень [162]; доповнення на основі ймовірностей погодження заявок; розбиття груп ризику відхилених заявок; нечіткого доповнення. У дисертаційній роботі запропоновано удосконалити метод нечіткого доповнення, зокрема запропоновано використати нейронну мережу як модель оцінювання множини прийнятих заявок.

Для кожного інтервалу балів розраховується певний коефіцієнт F_k доповнення для прийнятих заявок (A_k – accepted), що визначається як відношення кількості всіх заявок до кількості прийнятих заявок (в межах одного інтервалу): $F_k = \frac{A_k + R_k}{A_k}$, де R_k – відхилені заявки (rejected). Формування нової доповненої вибірки здійснюється шляхом вибору з оригінальної вибірки лише прийнятих заявок з їх цільовим поведінковим результатом (good/bad) n_i разів, де кількість входжень дорівнює або пропорційна коефіцієнту доповнення. Тоді залежність відношення шансів доповненої вибірки є випуклою комбінацією оригінального відношення шансів прийнятих заявок і отриманого

відношення відхилених заявок ($\alpha \in [0;1]$): $\frac{AG}{AB} = \alpha \frac{G}{B} + (1-\alpha) \frac{AG-G}{AB-B}$. Зазвичай

$$\frac{\text{odds}_{\text{accepted}}}{\text{odds}_{\text{rejected}}^*} = \frac{AG(AB-B)}{AB(AG-G)} \in [1,5; 4,0].$$

Далі пропонується побудувати нейронну модель на множині прийнятих заявок для визначення ваги кожної прийнятої заявки $w_i = p_i(NN(\text{accepted} | \mathbf{x}))$. Доповнена вибірка складається з прийнятих заявок, включених пропорційно ненормованій вазі $w_j(\mathbf{x}) = p_j^{-1}(NN(\text{accepted} | \mathbf{x}))$, обчисленої як w_i окремо для кожної прийнятої заявки, що включається з даною вагою у нову доповнену вибірку [343].

Тоді кожна прийнята заявка оцінюється на основі нейронної мережі з одиничним ваговим коефіцієнтом, а кожна відхилена заявка входить у вибірку *два рази*: з вагою рівною ймовірності «хорошого» індикатора $p_{i1} = p(y_i^{NN} = 1 | \mathbf{x})$ з *одиничним цільовим результатом*, а також з вагою рівною ймовірності негативного цільового індикатора $p_{i0} = p(y_i^{NN} = 0 | \mathbf{x}) = 1 - p(y_i^{NN} = 1 | \mathbf{x}) = 1 - p_{i1}$ з *нульовим цільовим результатом*. Прогнозні ймовірності (ваги) отримуються за допомогою моделі, побудованої лише на прийнятих заявках (known good/bad). Основні етапи нейро-нечіткого методу доповнення наведені на рис. 2.8.

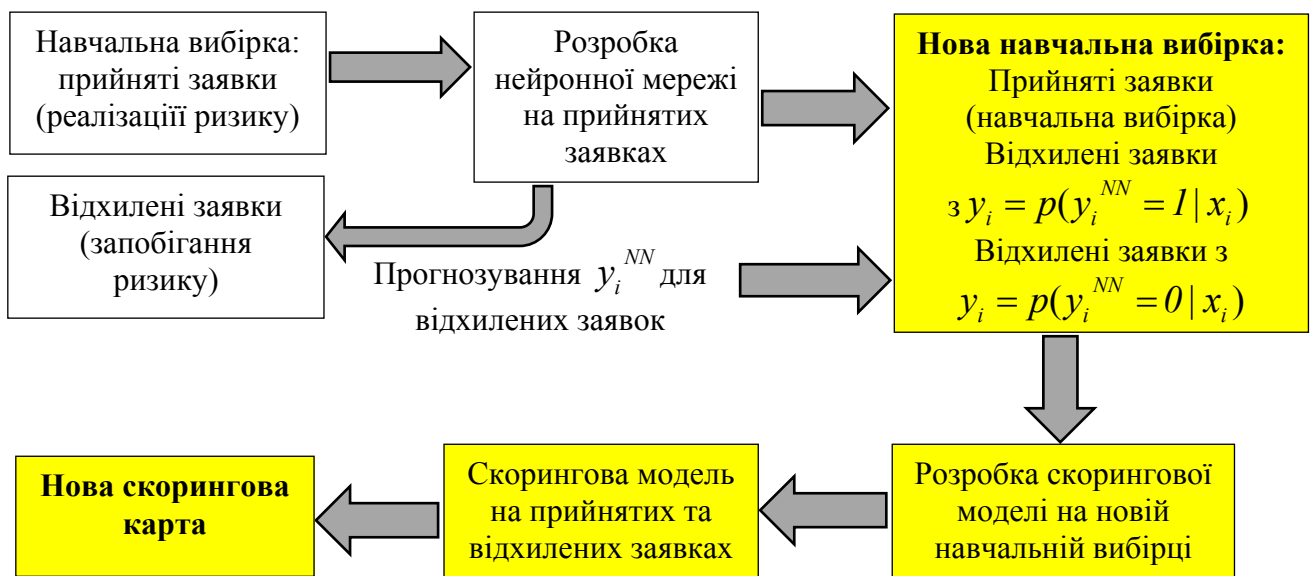


Рис. 2.8 Суть вдосконалення нейро-нечіткого методу доповнення

Модифікацією методу є перерахунок ваг з урахуванням співвідношення відхилених і погоджених заявок:

$$r = \frac{N_R}{N_A + N_R},$$

де N_R – кількість відхилених заявок, N_A – кількість погоджених заявок.

Для цього вводиться спеціальна корекція ваг:

$$W = \frac{r}{1-r} \frac{N_A}{N_R^*} = \frac{N_R / (N_A + N_R)}{1 - (N_R / (N_A + N_R))} \frac{N_A}{\sum_{i: \text{Rejects}} (p_{i0} + p_{i1})}.$$

Нові значення переоцінених ваг визначаються наступним чином:

$$w_{i1}^* = p_{i1} \cdot W \cdot \text{EventRateIncrease},$$

$$w_{i0}^* = p_{i0} \cdot W.$$

Отже, основним вдосконаленням нейро-нечіткого методу є застосування нейронної мережі, яка будується на навчальній вибірці, для прогнозування значення ймовірності для відхилених заявок. Ваги для Good/Bad визначаються за нечітким підходом. Використання такого підходу для задачі оцінювання кредитних ризиків дозволило отримати точніші оцінки прогнозів для вибірки відхилених заявок та скорингових оцінок в цілому для всієї вибірки.

Для оцінки якості побудованої скорингової карти використовуються наступні характеристики: помилки першого та другого роду, ROC-крива або крива Лоренца, вартість неправильної класифікації (втрати внаслідок видачі кредитів «поганим» клієнтам та втрачені можливості внаслідок невидачі кредитів «добрим» клієнтам) тощо [164]. Суттєвим на етапі розробки та у подальшому застосуванні скорингової карти є правильний вибір порогу відсікання моделі, тобто такої оцінки, яка сумарно розраховується по всіх характеристиках клієнта.

Етап валідації скорингової карти передбачає порівняння статистик для розроблених версій карт та порівняння розподілів позитивних та негативних випадків для розроблених версій, обрання кращої з них для подальшого використання [109]. Особлива увага звертається на процент неправильної

класифікації, «силу передбачення» карти – інформаційний критерій Акайке, Байєсовий критерій Шварца, статистику Колмогорова-Смірнова і т. ін. [26].

Слід зазначити, що навіть після впровадження скорингової карти процес моніторингу ефективності скорингових карт продовжується, оскільки в процесі її роботи можуть змінитися фактори та характеристики, що були включені, змінитися законодавство або ключові фактори оцінки [108]. На практиці використання скорингових карт є ефективним інструментом для ризик-менеджерів, оскільки дає наочне представлення щодо рівня ризикованості того чи іншого клієнту чи продукту.

У деяких випадках є сенс встановити контрольні точки у відповідності з правилами кредитної політики. Наприклад, якщо політика компанії вимагає, щоб на розгляд передавались кредити з коефіцієнтом обслуговування більше 42%, то необхідно здійснити групування коефіцієнту обслуговування боргу і встановити поріг на рівні 42%. Перевага подібного групування полягає в тому, що воно мінімізує спотворення скорингової карти, викликане таким правилом кредитної політики, і дозволяє виділити клієнтів, на яких це правило впливає [109]. Крім того, таке групування дозволяє перевірити існуюче представлення і ті правила політики, які діяли досі. Наприклад, воно дозволяє зрозуміти, чи має сенс встановлювати контрольну точку на рівні 42% або перенести її на більш високий рівень для збільшення диференціації рівнів.

Розглянемо застосування описаної вище методології розробки скорингових карт на реальному практичному прикладі.

2.4.5. Приклад побудови скорингової карти для аплікаційного скорингу

Нехай наявна інформація по автокредитам, які були видані банком протягом певного періоду та строк кредитування по яким закінчився, тобто наявна інформація-результат чи були такі кредити повернуті, чи ні. Вибірка складається з 5837 випадків та характеризується 22 змінними, такими як сума займу, рекомендована ціна виробника, реальна вартість авто, тривалість займу, вартість кредиту (Ltv), цільова змінна – чи був кредит повернутий чи ні,

скоринговий бал, авто знаходиться в лізингу чи власне, дохід позичальника, кількість відкритих торгів, відновлювальна лінія, відновлюваний борг, кількість відновлюваних торгів, чи була машина у використанні, чи відбулося банкрутство. Наявна інформація щодо заявок, які були відхилені існуючою скоринговою моделлю [108]. Необхідно розробити нову карту, яка дозволить швидко автоматизувати процес видачі кредиту заявникам того ж типу кредиту (автокредиту, споживчого, іпотечного тощо) [306, 309].

Вхідна вибірка розділяється в обраному процентному співвідношенні на навчальну, тестову і валідаційну за допомогою компоненти Data Partition. Далі виділяють характеристики-змінні, що найбільше впливають на результуючу змінну на основі інформаційного значення (IV), та інтервали для цих змінних на основі характеристики WOE [164] за допомогою компоненти інтерактивного групування змінних (Interactive Grouping). Після цього будується початкова скорингова карта за допомогою компоненти Scorecard.

Початкова скорингова карта формується на основі раніше використаної скорингової моделі і лише на виданих автокредитах, тому потім вибірка доповнюється інформацією, що характеризує реальних позичальників, тобто додається вибірка відхилених заявок на автокредити. Це наближує модельовану вибірку і, відповідно, карту до реальних даних та конкретних клієнтів, оскільки, клієнти, яким було видано кредити, можуть не відображати особливості і характеристики популяції в цілому. До того ж, це дозволяє врахувати інформацію, яка могла бути втрачена через відхилення заявок, що могли бути заявками добросовісних клієнтів [106]. Додаючи інформацію з відхилених заявок, слід враховувати статистичну інформацію банку, тобто долю кредитів, що не були видані, адже в іншому випадку можливе спотворення вибірки в бік відхилених заявок. Для цього використовується запропонований нейронечіткий метод, який оцінює заявки, що були відхилені, за допомогою нейромережі, а ваги встановлює за нечітким підходом у відповідності з існуючим у банку процентним співвідношенням щодо виданих та відхилених кредитів. Весь процес аналізу та побудови нової СК представлений на рис. 2.9.

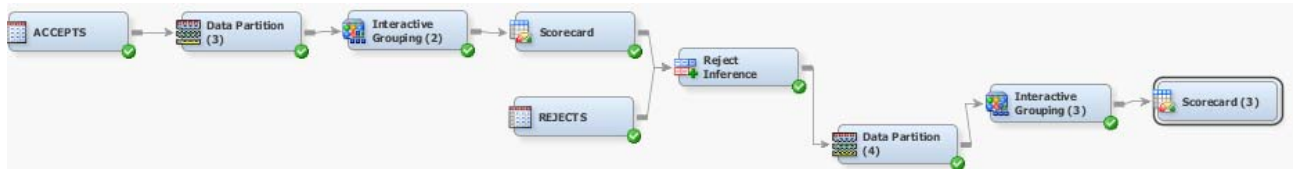


Рис. 2.9 Діаграма розробки скорингової карти для задачі автокредитування в SAS Enterprise Miner

Після додавання відхилених заявок (Reject Inference) і повторному перерозподілу вибірки (Data Partition) виконується повторно інтерактивне групування (Interactive Grouping). Далі формується фінальна СК, яка враховує інформацію щодо відхилених заявок, а тому значення скорингових балів для неї є уточненими і більш коректними для модельованої задачі.

При розрахунку скорингових балів за описаною вище формулою (2.3) обов'язково враховується вплив (як позитивний, так і негативний) значень певної характеристики на результат (рис. 2.10).

Scorecard		Scorecard Points
age_oldest_tr	age_oldest_tr < 38	16
	38 <= age_oldest_tr < 111	19
	111 <= age_oldest_tr < 172	21
	172 <= age_oldest_tr	26
	MISSING	17
bureau_score	_MISSING_	6
	bureau_score < 654	3
	654 <= bureau_score < 701	21
	701 <= bureau_score < 759	40
	759 <= bureau_score	69
ltv	_MISSING_	-86
	ltv < 75	52

Рис. 2.10 Фінальна скорингова карта для автокредитів

Для характеристики ltv наявність пропущених значень, оцінюється як негативна характеристика з вагою – 86 балів. Тому, при наявності пропусків у характеристиці ltv сукупний скоринговий бал клієнта буде знижений на 86 балів, і така заявка, скоріш за все, не буде автоматично ухвалена [106].

Порівняння скорингових карт, отриманих до та після врахування відхилених заявок, показує розбіжність як у розподілі скорингових балів, так і у значеннях за окремими характеристиками.

Звіти дозволяють аналітикам банку визначати оптимальний поріг відсікання [106], визначати характеристики, що найбільше впливають на цільову змінну, враховувати їх при прогнозуванні кредитних ризиків та розробці нових скорингових карт (рис. 2.11).

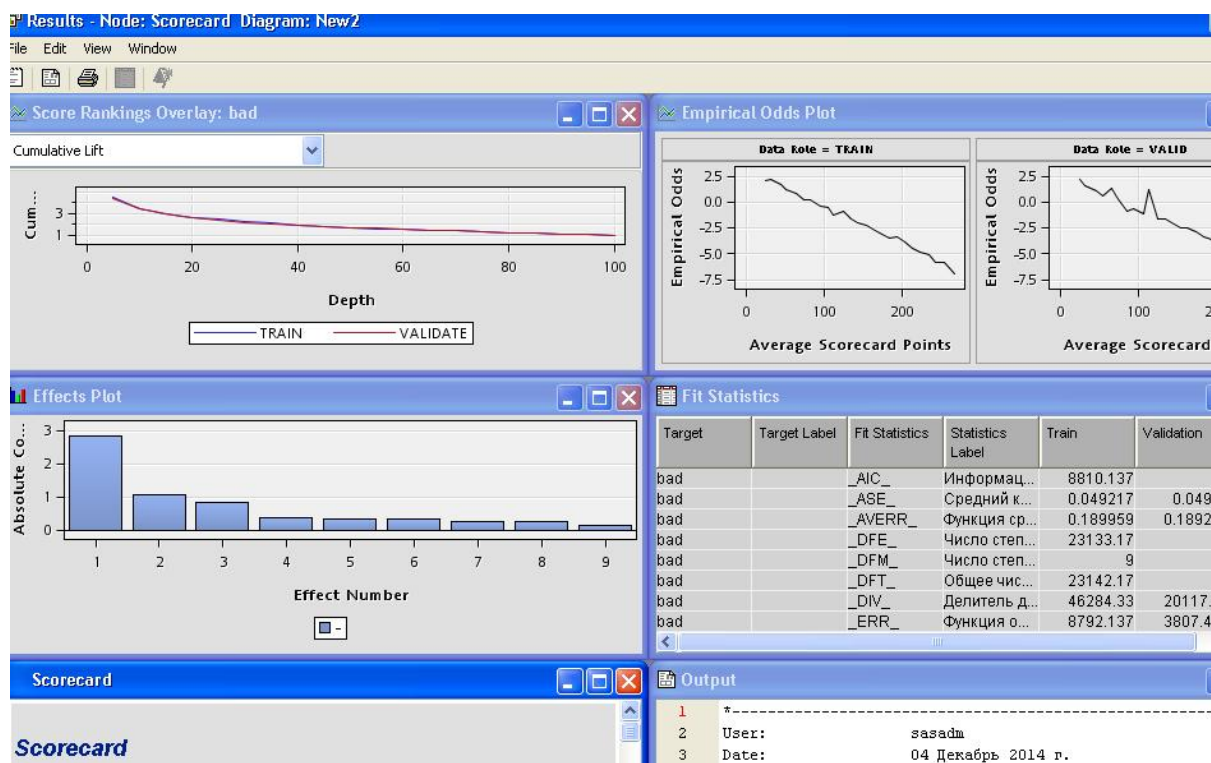


Рис. 2.11 Звіти, сформовані в SAS Enterprise Miner для нашої задачі

2.4.6. Скорингова карта як узагальнена міра ризику

Використання скорингової карти дає можливість автоматизувати процес підрахунку скорингових балів та скорингових оцінок для клієнтів-позичальників. В той же самий час, продовжуючи моніторити роботу існуючих скорингових моделей та варіюючи порогове значення в разі необхідності, банк може коригувати допустимий ступінь ризику (ймовірність його появи).

Узагальнимо підхід до моделювання скорингових моделей і скорингових карт для фінансових ризиків.

Нехай, у фінансової організації зберігається статистична інформація щодо її діяльності у попередні періоди (кредитні історії, трейдингові дні на біржі з акціями, трейдингові операції, тендерні закупівлі на платформі ProZorro, заявки на транзакції, тощо). Існує інформація щодо фінансових ризиків, які виявились раніше. Випадки, у яких фінансова організація несе збитки, характеризуються як «Погані» (“Bad”), випадки, у яких ризик не спостерігався, відносяться до категорії «Хороші» (“Good”). Фактори, за якими характеризують скорингову карту, визначаються на початковому етапі за допомогою експертного підходу або з використанням мереж Байєса, які встановлюють причинно-наслідкові зв’язки між причинами появи ризику та змінною, що характеризує, чи ризик відбувся. Далі визначається поріг відсікання, що встановлюється фінансовою (кредитною) організацією як ймовірність (ступінь ризику) [108]. Якщо ймовірність настання фінансового ризику для заданої сукупності факторів:

$$P = F(w^j, x_i^j) > cut - off,$$

перевищує допустимий рівень (прийнятний), то подія характеризується як ризикова (Bad) і відхиляється автоматично. Таким чином, можемо підрахувати статистику по транзакціям/угодам/заявкам, скільки з них виявилось поганих/хороших, який при цьому скоринговий бал був виставлений системою. В результаті отримаємо скорингову таблицю (рис. 2.12) розподілу «хороших» / «поганих» подій в залежності від ранжування скорингових балів. За нею можна отримати статистичні дані про розподіл, сукупний процент «поганих», який зростає автоматично при зниженні скорингових балів.

$$\mu_R = cut - off_i = \frac{N_{Bad}}{N}.$$

Встановлений поріг відсікання є мірою ризику і фактично показує толерантність фінансової системи до ризику, тобто визначає обсяг (частку) негативних випадків, який є допустимим для фінансової системи. Зменшуючи поріг відсікання, фінансова система свідомо погоджується прийняти на себе більший ризик і допускає більшу кількість ризикових подій.

Score Range	Count	Cumulative Count	Number of Goods	Cumulative Number of Goods	Number of Bads	Cumulative Number of Bads	Marginal Badrate	Cumulative Badrate	Approval Rate
273<=Score<279	842	842	840	840	2	2	0.24	0.24	1.81
267<=Score<273	511	1353	510	1350	1	3	0.20	0.22	2.91
262<=Score<267	574	1927	570	1920	4	7	0.70	0.36	4.14
256<=Score<262	2087	4014	2070	3990	17	24	0.81	0.80	8.63
250<=Score<256	1756	5770	1740	5730	16	40	0.91	0.89	12.41
245<=Score<250	2338	8108	2310	8040	28	68	1.20	0.84	17.44
239<=Score<245	2917	11025	2880	10920	37	105	1.27	0.95	23.71
233<=Score<239	3774	14799	3720	14640	54	159	1.43	1.07	31.83
228<=Score<233	2766	17565	2700	17340	66	225	2.39	1.28	37.77
222<=Score<228	3366	20931	3300	20640	66	291	1.96	1.39	45.01
216<=Score<222	4492	25423	4380	25020	112	403	2.49	1.59	54.67
211<=Score<216	4210	29633	4080	29100	130	533	3.09	1.80	63.73
205<=Score<211	3455	33088	3360	32460	95	628	2.75	1.90	71.16
199<=Score<205	4419	37507	4260	36720	159	787	3.60	2.10	80.66
194<=Score<199	1549	39056	1440	38160	109	896	7.04	2.29	83.99
188<=Score<194	2006	41062	1890	40050	116	1012	5.78	2.46	88.31

Рис. 2.12 Скорингова таблиця ранжування балів

Ця частка може виявитись меншою, оскільки не реалізується очікуваний ризик (неповернення кредиту або втрати клієнта) і тоді фінансова система отримає додатковий виграш (за рахунок виданих кредитів, які були повернуті).

Висновки до другого розділу

У розділі проаналізовано основний сучасний інструментарій, програмно-апаратні платформи реалізації засобів дослідження фінансових систем та оцінювання фінансових ризиків. Узагальнюючи, необхідно визначити такі основні моменти:

1. Існують різноманітні рішення та інструментальні засоби для дослідження фінансових процесів, аналізу і оцінювання фінансових ризиків, проте не вся послідовність кроків вирішення сучасних задач моделювання і прогнозування фінансових ризиків може бути виконана на одній платформі. Недоліком існуючих рішень є те, що вони часто занадто універсальні, призначені для вирішення широкого кола проблем, проте для конкретних задач вимагають додаткових застосувань, спеціальних продуктів і ліцензій. Виникає

необхідність у розробці зручного інструментарію для оцінювання та моделювання ризиків, який буде доступним в реальному часі фахівцям предметної області.

2. Ефективним інструментом трансформації експертного знання та досвіду оцінювання ризиків фінансових систем у формалізовані характеристики для підтримки прийняття оперативних рішень є скорингові карти. Подальше доопрацювання скорингових карт та періодичне їх оновлення дозволяє швидко реагувати в разі різких змін на ринку банківських послуг, нормативів НБУ, корегування внутрішньої стратегії банку, збільшення конкуренції, відтоку клієнтів, різкого збільшення неповернень позик, тощо.

3. Система оцінки ризиків (скорингу), як і інші прогностні моделі, – це засіб оцінки рівня ризику, пов'язаного з клієнтами. Доцільно застосувати скорингові карти для поведінкової оцінки клієнтів, коли необхідно запропонувати нові кредитні продукти, послуги, сформувані ліміти для клієнтів різного віку, цілей та уподобань.

4. Методика скорингу пропонує об'єктивний спосіб і єдиний підхід до оцінки ризиків. Показано, що скорингові карти є зручним інструментом для кредитних фахівців, і визначено доцільність їх застосування для опрацювання різних видів фінансових ризиків для таких прикладних задач, як прогнозування ризиків неплатежів (банкрутства), шахрайства, претензій (у страхуванні), стягнення боргів, тощо.

5. У розділі удосконалено методологію застосування скорингової карти як інструменту оцінювання фінансового ризику, в якій запропоновано використання розробленого нейро-нечіткого методу доповнення вибірки відхиленими заявками при моделюванні фінансових ризиків. Нейро-нечіткий метод доповнення відхиленими заявками передбачає застосування нейронної мережі для прогнозування значень ймовірності відхилених заявок та нечіткого підходу для визначення їх ваг. Використання такого підходу для задачі оцінювання кредитних ризиків дозволило отримати точніші прогнози для вибірки відхилених заявок та скорингових оцінок в цілому для всієї вибірки.

РОЗДІЛ 3

РОЗРОБКА СИСТЕМНОЇ МЕТОДОЛОГІЇ МЕНЕДЖМЕНТУ ФІНАНСОВИХ РИЗИКІВ

3.1. Основні складові системної методології менеджменту фінансових ризиків

Методологія дослідження фінансових ризиків з точки зору системного підходу (рис. 3.1) має включати вивчення причин появи фінансових ризиків та опрацювання невизначеностей різної природи, здійснювати попередню оцінку за відомими статистичними, математичними та економічними методами для виявлення найважливіших і найкритичніших рівнів ризику та можливих втрат. У системній методології мають бути задіяні функціональні, економічно-фінансові, ймовірнісно-статистичні та інші критерії, які дозволять обрати коректний метод для опрацювання конкретного виду фінансового ризику, оцінити повноту його дослідження та вибрати кращу модель, що описує ймовірність його появи та можливих втрат[76, 126].

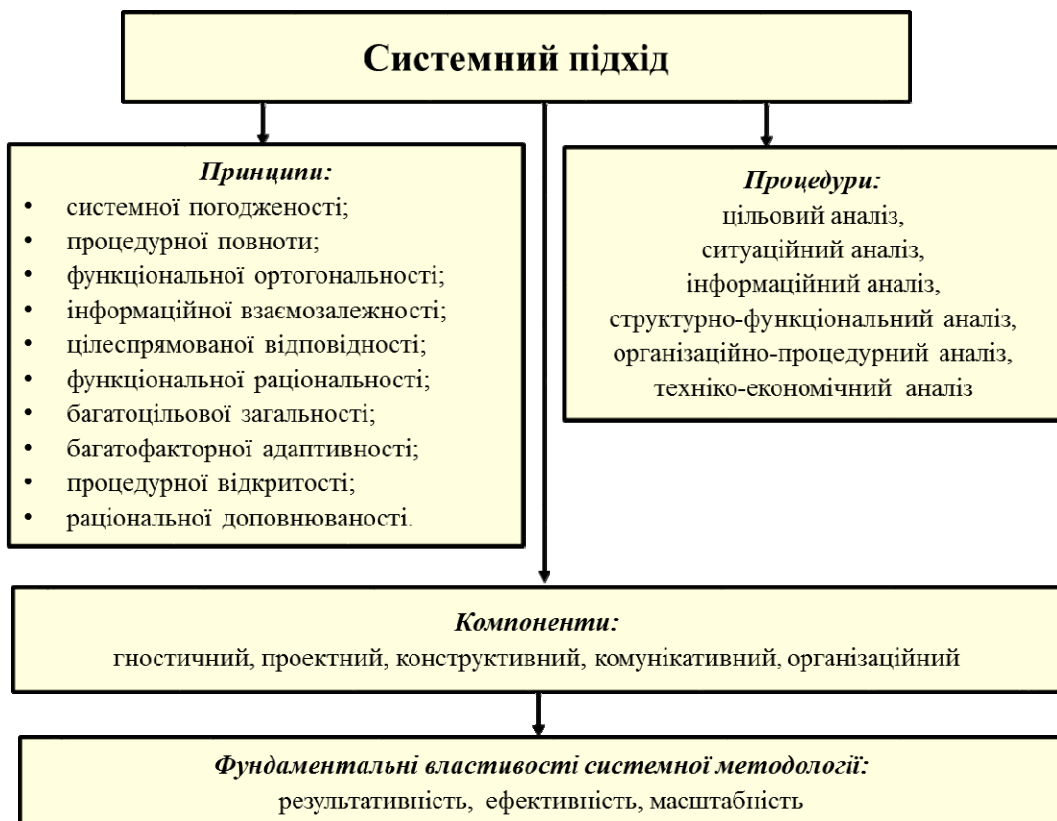


Рис. 3.1 Складові системного підходу

3.1.1. Принципи менеджменту ризиків

Разом з принципами системного підходу необхідно сформулювати принципи, що характеризують специфіку оцінювання та менеджменту ризиків. Вони стануть основою для розробки нових і удосконалення існуючих методів і моделей для оцінювання фінансових ризиків.

Основними принципами менеджменту ризиків [84, 142, 165, 167, 312] є:

- ✓ принцип існування ризику;
- ✓ принцип усвідомленого прийняття ризиків [126];
- ✓ принцип керованості прийнятих ризиків;
- ✓ принцип незалежного менеджменту окремих груп ризиків;
- ✓ принцип відповідності рівня прийнятих ризиків рівню прибутковості операцій, що проводяться;
- ✓ принцип відповідності рівня прийнятих ризиків можливим втратам.

У дисертаційній роботі розробляються та пропонуються додаткові принципи менеджменту ризиків, такі як: принцип адаптивного менеджменту ризиків, принцип урахування інформаційного ризику та принцип інтегрованого динамічного урахування *фактору часу* в аналізі ризиків. Принцип адаптивного менеджменту ризиків передбачає необхідність неперервного моніторингу ризику, корегування та адаптації розробленої раніше моделі, що оцінює обраний вид фінансового ризику з урахуванням того, що як мінімум один з параметрів оцінювання ризику, ймовірність або втрати, змінюється. Цей принцип реалізований у вигляді методу структурно-параметричної адаптації, розробленому у п'ятому розділі. На основі принципу інтегрованого динамічного урахування фактору часу, ступеня та рівня ризику у фінансовому ризику-менеджменті розроблений метод динамічного оцінювання ризиків та динамічні моделі, викладені у четвертому розділі.

Всі описані принципи формують системну методологію менеджменту ризиків, яка реалізується у вигляді розроблених інформаційних технологій та систем підтримки прийняття рішень для опрацювання ризиків [142].

3.1.2. Принцип урахування інформаційного ризику під час оцінювання фінансового ризику

Невизначеності, які зустрічаються під час оцінювання фінансових ризиків, обумовлюють необхідність використання для оцінювання інформаційних ризиків такі методи, що дозволяли б комбінувати експертний досвід та статистичні дані, при цьому встановлювати взаємозв'язки між окремими факторами [91, 96]. Це можливо при застосуванні мереж Байєса, особливості яких обговорювались у роботах [75, 123, 125].

Оцінювання інформаційних ризиків за допомогою мереж Байєса [96] є сенс здійснювати для найсуттєвіших факторів ризику, виявляючи найбільш уразливі ІТ-активи, взаємозв'язки між ними та засоби, що дозволяють зменшити їх вразливість.

На етапі побудови моделі потрібно виявити всі фактори ризику, що можуть зустрічатися під час роботи компанії, щоб отримати коректну початкову модель у вигляді мережі Байєса. Далі виконується оцінка включених до неї «суттєвих» змінних та визначення початкових апріорних ймовірностей. Для певних характеристик визначаються рівні контролю та бальна оцінка, для ризикових подій здійснюється оцінка ймовірності їх реалізації та наслідків, пов'язаних з ними втрат [75]. Наслідки класифікують на двох рівнях. Перший рівень – зміни в технологічних та інформаційних активах (для інформаційних активів – порушення цілісності, доступності та конфіденційності; для матеріальних – від повної втрати активу до збою на несуттєвий проміжок часу). Другий рівень класифікації – з точки зору можливих втрат, таких як втрачена вигода, пеня, втрати робочого часу, втрата репутації тощо.

Така дворівнева класифікація [79, 80] дозволяє співвіднести витрати з вартісною оцінкою, яка необхідна для управління ризиками. Величина прямих фінансових витрат може бути оцінена експертами або на основі даних операційних втрат. Вартість нематеріальних витрат оцінити важче, але тут також можливі експертні оцінки, такі як відтік клієнтів, зниження темпів

відкриття нових рахунків, зміна ринкової вартості компанії/акцій компанії тощо.

Для банку, наприклад, основними факторами інформаційного ризику є ризик витоку інформації про клієнтів, неповернення кредиту, некоректна робота інформаційної системи, збої при роботі з рахунками, перерахуваннями грошових систем, хакерські атаки, помилки при нарахуванні процентів[75].

Побудова мережі Байеса може здійснюватись комплексно, у розрізі всіх цих категорій ризиків для оцінки сукупного інформаційного ризику інформаційної системи, підсумовуючи розподіл втрат по декільком ризиковим подіям, потенційно здатним реалізуватись у системі [75]. Сума розподілів втрат за окремими ризиковими подіям необхідна також для визначення очікуваних та неочікуваних втрат, які отримують на основі розрахунку математичного очікування і VaR-агрегованого розподілу IT-ризиків.

Оскільки якість вхідних даних також є причиною інформаційного ризику, як і реалізація інформаційної загрози [96], то пропонується оцінити фінансовий ризик через оцінку неякісної, недостовірної, неповної інформації або зовнішнього інформаційного впливу за наступною формулою [112]:

$$EL(FR | IR) = P(IR) \times EL(IR) \times EF(IR),$$

де $EL(FR | IR)$ – очікувані втрати ФП від реалізації інформаційного ризику (загрози),

$P(IR)$ – ймовірність реалізації інформаційної загрози,

$EL(IR)$ – очікувані втрати від реалізації інформаційного ризику (загрози),

$EF(IR)$ – міра уразливості ресурсу до інформаційної загрози.

Ймовірність фінансового ризику через появу інформаційного ризику визначається так [112]:

$$P(FR | IR) = \frac{P(IR | FR) \times P(FR)}{P(IR)}.$$

На основі принципу урахування інформаційного ризику можна врахувати вплив неповної, неякісної, неточної інформації у вхідних даних.

3.1.3. Основні критерії у задачах аналізу та оцінювання фінансових ризиків

Задачі аналізу, оцінювання та прогнозування фінансових ризиків в різних постановках використовують специфічні функціональні, економічні і фінансові критерії, критерії якості даних та адекватності моделей, критерії оцінювання оптимальності та якості прийнятих рішень. Вони дозволяють визначити повноту статистичних даних, коректність застосованих методів та розроблених моделей для певних категорій ризику, точність та якість побудованих прогнозів, визначити стратегію мінімізації ризику, наскільки знайдене рішення є близьким до оптимального, оцінити ймовірність реалізації ризику та витрати на подолання ризику. Перелік основних критеріїв, які використовуються в задачах ризик-менеджменту, наведений на рис. 3.2.

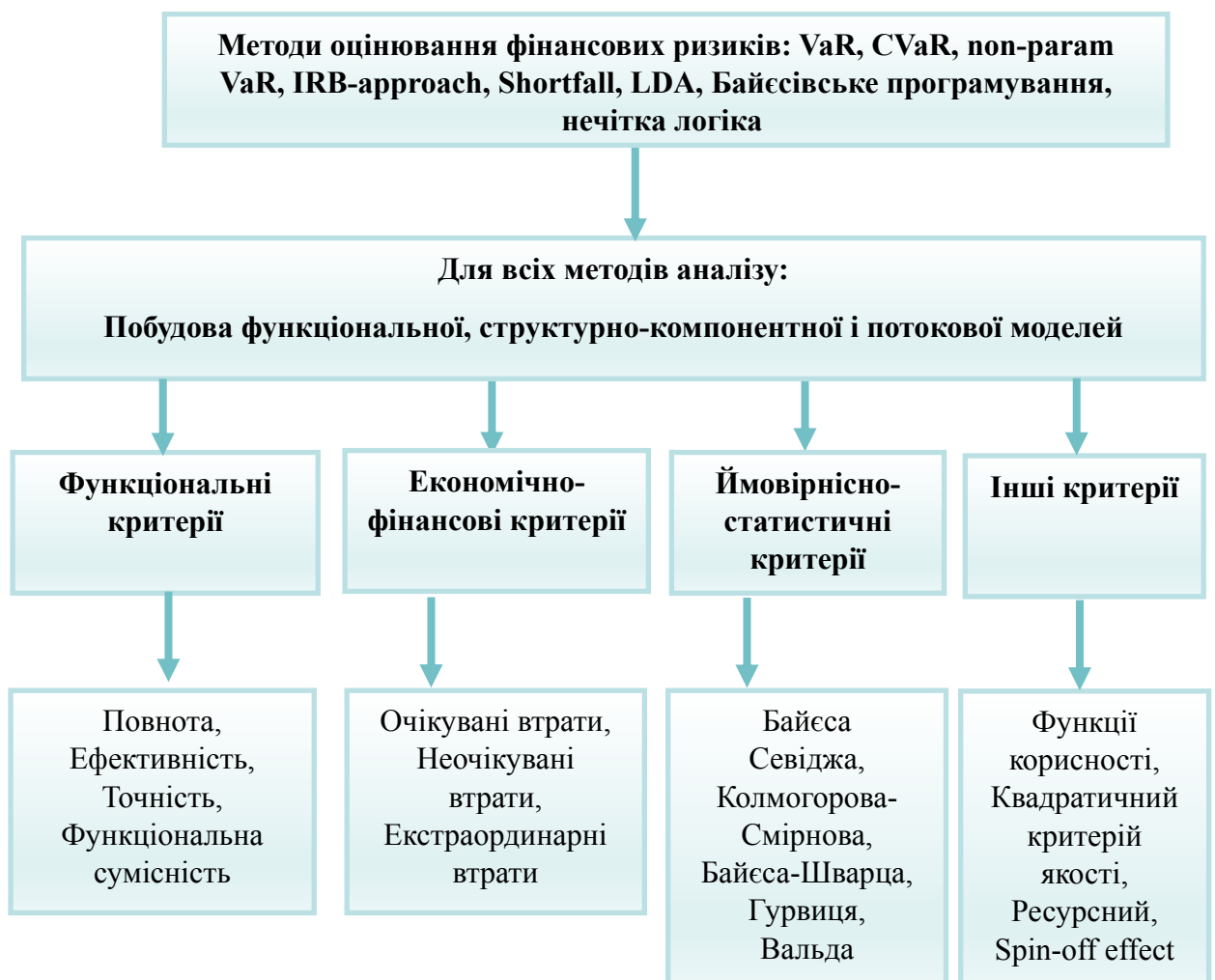


Рис. 3.2 Основні групи критеріїв системної методології менеджменту фінансових ризиків

Функціональні критерії [78] передбачають перевірку повноти опису ризику, ефективність застосування методу оцінки до даної категорії ризику, точність оцінювання ризику та функціональну сумісність, наприклад при застосуванні комбінованих методів оцінювання ризику чи одночасному урахуванні різних видів ризиків.

3.1.4. Економічно-фінансові критерії

Економічні або фінансові критерії оцінюють можливі збитки внаслідок реалізації ризику та перевіряють, чи задовольняє існуюча методологія ризик-менеджменту нормативам щодо необхідного економічного капіталу як джерела покриття збитків. До них відносяться оцінки ризиків у грошовому еквіваленті. Зазвичай оцінюють очікувані, неочікувані та екстраординарні втрати і формують резерви капіталу на їх покриття. За IBR-підходом [107] очікувані втрати внаслідок реалізації ризиків можуть бути обчислені так:

$$EL = \sum_{i=1}^N P(R_i) \cdot CE_i \cdot LGD_i, \quad (3.1)$$

де $P(R_i)$ – ймовірність (очікувана частота) прояву i -го виду ризику (наприклад, ризику зниження фінансової стабільності), що набуває значення на відрізку $[0,1]$; CE – загроза внаслідок реалізації ризику – сума втрат (заборгованості внаслідок реалізації даного ризику)[107]; LGD – покриття ризику страховкою (в разі її наявності), заставою або ефективність запобіжних заходів, що приймає значення від 0 (ризик, повністю покритий заставою) до 1 (ризик, не покритий заставою); N – кількість типів ризиків.

За методологією VaR (Value at Risk) верхня межа втрат обчислюється у довірчому інтервалі [10]:

$$P(Loss_t(k) < VaR_t(k)) = (100 - \alpha)\%, \quad (3.2)$$

де $Loss_t(k)$ – фактичні втрати на момент часу t за період k днів, $VaR_t(k)$ – прогнозовані втрати на момент часу t за період k днів, α – довірчий рівень.

Conditional Value at Risk (CVaR), також відомий як очікуваний дефіцит, визначає кількість ризику або «товщину хвоста» для інвестиційного портфеля.

CVaR розраховується через середньозважене значення «екстремальних» втрат у хвості, що виходять за межі граничного значення VaR.

$$CVaR = E(X | X > VaR) ,$$

$$\text{тобто } CVaR = \frac{1}{1-c} \int_{-1}^{VaR} xp(x)dx ,$$

де $p(x)$ – це щільність розподілу втрат, c – точка відсікання на розподілі, встановлена аналітиком як поріг VaR, VaR – погоджена верхня межа VaR.

Параметричний VaR розраховується як [124]:

$$VaR = \alpha * \sigma * ВП * \sqrt{N} , \quad (3.3)$$

де α – квантиль довірчого інтервалу; σ – волатильність (норма мінливості); ВП – величина відкритої позиції; N – період прогнозування.

Тобто очікувані втрати – це середнє значення VaR. Тоді неочікувані втрати (UL) розраховуються як дисперсія очікуваних втрат в часі:

$$UL = \max(VAR) - EL . \quad (3.4)$$

Економічний капітал, який має покривати втрати, розраховується як:

$$EC = UL \times RR \times CM ,$$

де RR – залишковий ризик, тобто частка неочікуваних втрат, які неможливо усунути за рахунок диверсифікації, а CM – коефіцієнт капіталу, який визначається як кількість стандартних відхилень, необхідна для досягнення довірчого рівня (який визначається експертним шляхом на основі рейтингу міжнародної фінансової групи ING). Може бути 1, 2, 3, 5, 7 і т.д. ING використовує 7 (для планового рейтингу AA).

3.1.5. Ймовірно-статистичні критерії оцінювання оптимальності прийнятих рішень

Для оцінювання оптимальності прийнятих рішень використовують різні критерії, які ґрунтуються на пошуку оптимального компромісу та раціональної корисності для особи, що приймає рішення. Вони можуть бути отримані шляхом розв'язання оптимізаційних задач в умовах мінімаксу, максиміну тощо.

Найпоширенішими критеріями в умовах невизначеності є критерії Севіджа, Байєса, Лапласа, Гурвиця, Вальда) [132].

Критерій Севіджа оцінює, чи можна за будь-яких умов уникнути великого ризику при прийнятті рішень. Це критерій крайнього песимізму, де гіршим є не мінімальний виграш, а максимальний ризик – максимальна втрата виграшу порівняно з тим, що можна було б досягнути за таких умов [132].

Для матриці виграшів $A = |a_{ij}|$ у кожному стовпці обчислюється максимальне значення по стовпцю $\max a_{ij}$ і формується нова матриця ризиків:

$$1 \leq i \leq n$$

$$r_{ij} = \max_{1 \leq i \leq n} a_{ij} - a_{ij}.$$

Матриця ризиків $R = |r_{ij}|$ дозволяє обрати таку стратегію, за якою величина ризику приймає найменше значення у найбільш несприятливій ситуації:

$$F_{opt} = F(X_{opt}, S) = \min \max_{1 \leq i \leq n} r_{ij} = \min \max_{1 \leq i \leq n} (\max_{1 \leq j \leq m} a_{ij} - a_{ij}).$$

Критерій Байєса мінімізує середні втрати. Передбачається, що відомий апіорний розподіл ймовірностей ситуацій $P(s_j)$, тоді ризик обчислюється через математичне сподівання [132]:

$$r(x_i) = \sum_{j=1}^m \bar{a}_{ij} P(s_j).$$

Визначається найкраща стратегія, яка мінімізує ризик:

$$r(x^*) = \min \sum_{j=1}^m \bar{a}_{ij} P(s_j), 1 \leq i \leq n.$$

Критерій Лапласа. Передбачається, що ймовірності всіх можливих результатів s_j рівні між собою [126]. Тоді для кожного рядка матриці виграшів $A = |a_{ij}|$ підраховується середнє значення по рядку, а оптимальною буде стратегія, яка максимізує це середнє:

$$F_{opt} = F(X_{opt}, S) = \max_{1 \leq i \leq n} \left[\frac{1}{m} \sum_{j=1}^m a_{ij} \right].$$

Критерій Гурвиця передбачає певний компроміс між песимістичною і оптимістичною стратегією [132] і розраховує деякий середній результат при виборі оптимальної стратегії в умовах невизначеності:

$$F_{opt} = F(X_{opt}, S) = \max[\alpha \max a_{ij} + (1 - \alpha) \min a_{ij}], 1 \leq i \leq n, 1 \leq j \leq m,$$

де α – коефіцієнт оптимізму, $0 \leq \alpha \leq 1$. При $\alpha = 1$ критерій Гурвиця перетворюється в критерій крайнього песимізму Вальда; при $\alpha = 0$ – критерій крайнього оптимізму, коли рекомендується обирати стратегію, за якої в найкращих умовах виграш максимальний [132]; при $0 < \alpha < 1$ – середня оцінка між крайнім песимізмом та крайнім оптимізмом. Чим більш небезпечна ситуація, тим більше особа, що приймає рішення, намагається «підстрахуватись» і $\alpha \rightarrow 1$.

Критерій Вальда є критерієм крайнього песимізму, що орієнтує ОПР на найгірші результати та рекомендує обирати стратегію, для якої в гірших умовах виграш максимальний. Оптимальною вважається стратегія [132], де мінімальний виграш є максимальним, тобто гарантується виграш, не менший, ніж максимум:

$$F_{opt} = F(X_{opt}, S) = \max_{1 \leq i \leq n} \min_{1 \leq j \leq m} a_{ij}.$$

3.1.6. Критерії оцінювання якості даних та адекватності моделей

Для перевірки коректності і доцільності використаних методів в системній методології пропонується використовувати множину критеріїв перевірки адекватності розроблених моделей, якості вхідних даних, точності оцінювання прогнозів. Умовно всю критеріальну базу можна розділити на:

❖ критерії *перевірки адекватності моделей*: (R^2 , Дарбіна-Уотсона, інформаційний критерій Акайке, критерій Байєса-Шварца)[26]:

– коефіцієнт детермінації R^2 :

$$R^2 = \frac{Var[\hat{\sigma}]}{Var[\sigma]},$$

– статистика Дарбіна-Уотсона:

$$DW = 2 - 2\rho,$$

$$\text{де } \rho = \frac{\frac{1}{N-1} \sum_{i=2}^N [e(i) - \bar{e}][e(i-1) - \bar{e}]}{\text{Var}[e]};$$

– інформаційний критерій Акайке [17]:

$$AIC = N \ln \left(\sum_{k=1}^N e^2(k) \right) + 2n;$$

– критерій Байєса-Шварца:

$$BSC = N \ln \left(\sum_{k=1}^N e^2(k) \right) + n \ln(N),$$

де $n = p + q + 1$ – кількість параметрів моделі, які оцінюються за допомогою статистичних даних (p – кількість параметрів авторегресійної частини моделі; q – кількість параметрів ковзного середнього; „1” з’являється тоді, коли оцінюється зміщення (або перетин, тобто a_0), N – довжина вибірки. Критерії Акайке та Байєса-Шварца пов’язані з квадратом похибок, а тому мають бути мінімальними для кращої моделі. Причому, введення нового регресора приводить до збільшення критерію (оскільки збільшується n), але одночасно зменшується сума квадратів похибок і критерій в цілому зменшується [26]. Якщо регресор не покращує модель, тобто похибка не зменшується, то критерій збільшується. Асимптотичні властивості для довгих вибірок кращі у критерію Байєса-Шварца, тому його рекомендують застосовувати при відносно великих значеннях N ($N > 100$).

❖ критерії **роздільної здатності моделі** в задачах класифікації ризиків (загальна точність, помилки 1-го та 2-го роду, індекс GINI, статистика Колмогорова-Смірнова, доля неправильної класифікації тощо) [108].

$$\text{Загальна точність моделі: } CA = \frac{\text{Вірноспрогнозовані}}{N},$$

де *Вірноспрогнозовані* – кількість коректно визначених випадків, а N – загальна кількість випадків.

В залежності від встановленого порогу відсікання обчислюють *помилки I-го та II-го роду* (табл.3. 1).

Таблиця 3.1

Помилки I-го та II-го роду

	Прогноз моделі: 0	Прогноз моделі: 1
Фактично: 0	Вірно класифіковані (TP)	Помилки II-го роду (FP)
Фактично: 1	Помилки I-го роду (FN)	Вірно класифіковані (TN)

ROC-крива (Receiver Operation Characteristic) показує залежність кількості вірно класифікованих позитивних прикладів від кількості невірно класифікованих негативних прикладів [123].

Індекс GINI, розроблений італійським статистиком Коррадо Джіні [287], – це площа області між діагоналлю і ROC-кривою, поділена на площу усієї області під діагоналлю. Формально визначається наступним чином:

$$GINI = \left(\int_{x \in X} F_B(x) dF_G(x) - \frac{1}{2} \right) / \left(\frac{1}{2} \right) = 2 \cdot AUC - 1,$$

де $F_B(x)$ та $F_G(x)$ – це емпіричні кумулятивні функції розподілів негативних («bad») та позитивних («good») елементів навчальної вибірки за бінарною цільовою змінною та відносно дійсних значень досліджуваної змінної.

Критерій Колмогорова-Смірнова визначається як:

$$KS = \max_{x \in X} |F_B(x) - F_G(x)|.$$

Діапазон значень індексу GINI складає $0 \leq G \leq 1$, а моделі з найвищою розділювальною здатністю, тобто моделі, які роблять високоякісну класифікацію, отримують найвищі коефіцієнти [164].

Доля неправильної класифікації (Misclassification Rate) обчислюється як відношення помилково спрогнозованих значень по відношенню до загальної кількості значень N [112]:

$$Misclassification Rate = \frac{\text{кількість хибно спрогнозованих значень}}{N}. \quad (3.5)$$

❖ критерії *якості вхідних даних* (статистичні характеристики ряду – середнє, дисперсія, ексцес, асиметрія, потужність вибірки, процент пропущених

значень, статистика Жак-Бера, наявність екстремальних значень, інформативність).

Критерій оцінки кількості пропусків обчислюється за кожною j -ю характеристикою:

$$I_{j(\text{missing})} = \frac{N_{j(\text{missing})}}{N_j} \times 100\%, \quad (3.6)$$

де $N_{j(\text{missing})}$ – кількість пропущених значень за j -ю змінною,

N_j – кількість значень для j -ї змінної (довжина вибірки).

Насамперед цей критерій дозволяє оцінити, чи слід включати цю змінну-характеристику в модель [127], адже будувати модель на нерепрезентативній або навіть зумисно неповній вибірці не має сенсу. Якщо $I_{j(\text{missing})} > p_{\text{cut-off}}\%$, то цю характеристику слід виключити з аналізу.

❖ критерії **якості оцінок прогнозів**.

Для аналізу спрогнозованих значень застосовують такі показники [26]:

- середня похибка $ME = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)$,
- середня абсолютна похибка $MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$,
- середня відсоткова абсолютна похибка $MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{|y_i|} \times 100\%$,
- середня відсоткова похибка $MPE = \frac{1}{N} \sum_{i=1}^N \frac{(y_i - \hat{y}_i)}{y_i} \times 100\%$,
- середньоквадратична похибка $RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$,
- середній квадрат похибок $MSE = E((y - \hat{y})^2) = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}$,
- сума квадратів похибок: $\sum_{k=1}^N e^2(k) = \sum_{k=1}^N [\hat{y}(k) - y(k)]^2 \rightarrow \min_{\theta}$,
- коефіцієнт Тейла:

$$U = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i)^2} + \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i)^2}}.$$

❖ критерії *якості характеристик-змінних* (кореляція, статистика Стьюдента, інформаційне значення (IV), значущість кожної окремої групи значень WOE) [108, 109, 164].

Кореляція характеризує наявність (відсутність) лінійної або нелінійної залежності між змінними. Кореляція може бути лінійною або нелінійною, залежно від типу взаємозв'язку, який фактично існує між змінними. Вибірковий коефіцієнт кореляції між двома змінними обчислюється за формулою [26]:

$$r_{yx} = \frac{\sum_{k=1}^N \{[y(k) - \bar{y}][x(k) - \bar{x}]\}}{\sqrt{\sum_{k=1}^N [y(k) - \bar{y}]^2} \sqrt{\sum_{k=1}^N [x(k) - \bar{x}]^2}},$$

де $-1 < r_{yx} < +1$; σ_x, σ_y – стандартні відхилення для змінних x і y , відповідно.

Статистика Стьюдента [17] використовується для перевірки значущості коефіцієнта в статистичному сенсі:

$$t = \frac{\hat{a} - a^0}{SE_{\hat{a}}},$$

де \hat{a} – оцінка коефіцієнта моделі, a^0 – нуль-гіпотеза, $SE_{\hat{a}}$ – стандартна похибка оцінки.

Характеристика *WOE*, описана формулою (2.4), визначає силу атрибутів змінної-характеристики, тобто наскільки коректно сформовані інтервали для даної характеристики. Інформаційне значення Information Value (IV) визначає передбачуючу силу характеристики [342] та описана формулою (2.5).

3.1.7. Критерії оцінки якості рішень

Для оцінювання *якості прийнятого рішення* можуть використовуватись сукупності критеріїв якості: різновиди функцій корисності, квадратичні критерії якості, тощо.

Квадратичний критерій якості оцінюється як [112]:

$$I = \sum_{k=1}^N [u^T(k) \times R \times u(k) + x^T(k) \times Q \times x(k)]^2 \rightarrow \min_{u,x}, \quad (3.7)$$

де $u(k)$ – вектор входів, $x(k)$ – вектор виходів, u^* – оптимальний вхідний вектор, який визначається в результаті розв’язання задачі оптимального керування.

Перший доданок передбачає мінімізацію втрат, другий – мінімізацію відхилень від заданого реального завдання.

Автором запропоновано такий *квадратичний критерій якості опрацювання ризику* [112]:

$$I_1 = \sum_{k=1}^N [x^T(k) \times R \times x(k) + (y(k) - y_\phi(k))^T Q (y(k) - y_\phi(k)) + z^T(k) P z(k)] \rightarrow \min_{y,x,z} \quad (3.8)$$

де $x^T(k) \times R \times x(k)$ – витрати на інвестиції,

$y(k) - y_\phi(k) = \tilde{y}(k)$ – реальні втрати,

$z^T(k) \times P \times z(k)$ – компенсаційні витрати.

Третій доданок в цьому критерії – це мінімізація втрат на боротьбу з ризиком (внаслідок дій конкурентів, зовнішніх невизначеностей). Детальніше можливості застосування критерію якості до практичних задач менеджменту ризиків буде описано у п’ятому розділі.

До категорії «інші критерії» зазвичай відносять найрізноманітніші критерії, що дозволяють перевірити коректність використання ресурсів в існуючій методології, корисність застосування методології, супровідні ефекти, які дозволяє отримати розроблена методологія, критерії якості оцінок та гіпотез. Ці критерії не завжди є формалізованими, можливо, ґрунтуються на досвіді і знаннях експертів, можуть бути сформульовані не одразу, а після запровадження даної методології на практиці.

3.2. Розробка системної методології аналізу, оцінювання і прогнозування фінансових ризиків

Методологія опрацювання фінансових ризиків має узагальнювати важливі складові, пов'язані з необхідністю опрацювання невизначеностей та неструктурованості вхідних даних, розробки і побудови моделей інтелектуального аналізу даних, використання інтегрованих моделей [293] та комбінованих методів для отримання вищих прогнозних оцінок. Перевірка описаних моделей здійснюється за сукупністю статистичних критеріїв з метою визначення кращої з моделей-кандидаток.

Розробка методології оцінки невизначеностей фінансових даних на основі системного підходу передбачає розв'язання таких задач [126]:

- ❖ опис невизначеностей в окремих елементах, що використовують фінансові дані;
- ❖ агрегування невизначеностей окремих елементів в загальну невизначеність сукупності фінансових даних;
- ❖ встановлення значущості поточних відмінностей і довгострокових тенденцій з урахуванням інформації про невизначеності;
- ❖ визначення видів використання фінансових даних, кількісне визначення невизначеностей різних видів в задачах аналізу фінансових даних;
- ❖ встановлення невизначеностей якісного характеру, які неможливо опрацювати статистичними засобами.

Фундаментальними властивостями системної методології є результативність, ефективність і масштабність [78]. Фундаментальні принципи, яких необхідно дотримуватися під час формування системної методології та її практичної реалізації у вигляді сукупності конкретних підходів, методів, методик, алгоритмів, пакетів прикладних програм є принципи системної погодженості, процедурної повноти, функціональної ортогональності, інформаційної взаємозалежності, цілеспрямованої відповідності, функціональної раціональності, багатоцільової загальності, багатофакторної адаптивності, процедурної відкритості, раціональної

доповнюваності [144]. Всі описані вище принципи, методи, критерії та властивості мають сформувати єдину методологію, що розробляється як основа інформаційної СППР (ІСППР), і представлена на рис. 3.3.

Урахування багатофакторності ризику пов'язане з виявленням можливих причин і зовнішніх факторів, які спричиняють його появу, та можуть викликати появу ризиків іншої природи (наприклад інформаційних). В системній методології визначено такі показники ризику: ступінь ризику як ймовірність його появи, та рівень ризику як відносні (можливі) втрати, які з'являються внаслідок реалізації ризику [126].

Однією з основних причин появи ризику будь-якої природи є невизначеність природи, дій учасників, зовнішніх факторів, інформації тощо. Тому оцінювання якості та перевірка коректності вхідної інформації та даних є важливим елементом системної методології, що розробляється.

Неповнота, нечіткість, спотвореність, неоднозначність інформації може проявлятися як ризик у певному розумінні [126]. Так, невизначеність інформації, яка зберігається у базах знань і даних, може унеможливити отримання коректної статистичної вибірки і застосування кількісного методу оцінювання ризику. Неточність інформації призводить до побудови неадекватних моделей і некоректних оцінок фінансового ризику. Нечіткість нормативів, постановок задач, неможливість однозначного трактування вхідної інформації призводить до прийняття суб'єктивного або навіть хибного рішення при оцінюванні ризику експертами [125].

Несвоєчасність отримання інформації може стати причиною некоректного оцінювання або неусвідомлення катастрофічності ситуації і, відповідно, призвести до пізнього реагування та застосування механізмів і засобів зниження ризиків. Суперечливість вхідної інформації або оцінок експертів може призвести до неможливості розробки адекватного механізму і своєчасного реагування на появу ризику [91].

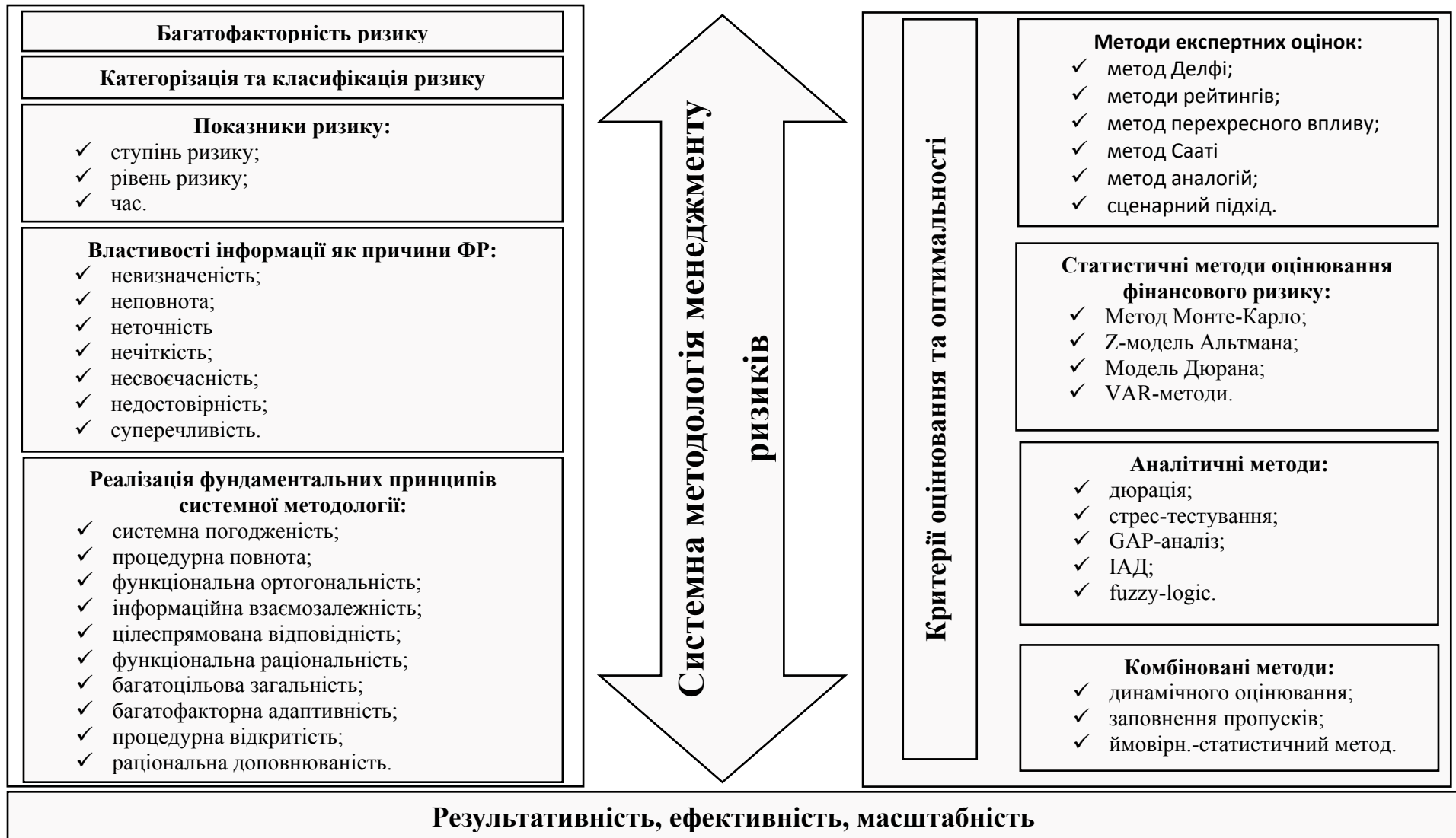


Рис. 3.3 Системна методологія менеджменту фінансових ризиків

Принцип системної узгодженості [76] передбачає використання у розробленій системній методології різноманітних методів, які є взаємопов'язаними і функціонально залежними, та надає чіткий алгоритм реалізації та послідовність викликів відповідних модулів для розроблених методів у ІСППР. Принцип процедурної повноти методології передбачає, що розроблена методологія охоплює всі етапи аналізу ризиків, починаючи від класифікації, категоризації ризику, аналізу причин його появи, і завершуючи оцінкою його рівня і рейтингу, та перевіркою побудованих моделей і оцінок [125]. Це дозволяє виконати повний цикл підготовки даних, моделювання та аналізу ризиків. Принцип функціональної ортогональності дозволяє незалежну реалізацію окремих процедур, модулів, що реалізують методи, розрахунок оцінок. Вони реалізуються, наприклад, у вигляді окремих застосувань в інформаційній технології і можуть викликатись незалежно один від одного. Принцип цілеспрямованої відповідності передбачає наявність різних критеріїв, оцінок, які взаємопов'язані і дозволяють отримати інформацію щодо точності прогнозу, якості моделі та відповідно якості прийнятого рішення [125].

Принцип інформаційної взаємозалежності передбачає, що всі результати, отримані на кожному етапі, реалізованому у вигляді окремих процедур або функцій, мають бути узгодженими з тими методами і процедурами [76], які є взаємозалежними (наприклад, при застосуванні комбінованих методів ймовірнісно-статистичного або динамічного оцінювання).

З точки зору принципу багатоцільової залежності [78], розроблені та запропоновані раніше методи можуть використовуватись для різних видів фінансових ризиків, різних типів системних задач. Багатофакторна адаптивність передбачає, що використані методи моделювання не мають суттєвих обмежень і мають високий ступінь гнучкості, характеризуються високим ступенем повноти вихідної інформації і реалізуються, наприклад, у вигляді контуру адаптації у структурно-параметричному методі.

При розробці системної методології накладалось обмеження щодо недопустимості дублювання функцій, що узгоджується з принципом функціональної раціональності. Для цього були сконструйовані спеціальні модулі, що реалізували окремі процедури, методи, моделі та могли викликати один одного, а не реалізовувати одні й ті самі функції. Принцип процедурної відкритості передбачає, що реалізована у вигляді ІСППР, методологія дозволяє вдосконалювати, видаляти, агрегувати існуючі та додавати нові дані, знання, методи, алгоритми, критерії [125]. Принцип раціональної доповнюваності встановлює, що методологія має можливість розширення сфери застосування, наприклад для інших видів ризиків, за рахунок доповнення її додатковими методами, моделями, критеріями та принципами, але лише за умови їхньої несуперечливості між собою та з вихідною методологією.

У системній методології обов'язково реалізується сукупність методів для опрацювання невизначеностей та формалізації експертного знання. На рис. 3.3 подано множину методів для встановлення експертних оцінок для ризиків, таких як метод Делфі, метод рейтингів (зокрема шкали для оцінювання ризиків за методологією S&P, Moody's, Fitch, НБУ), метод перехресного впливу, метод Сааті, метод пошуку аналогій серед історичних даних, сценарний підхід [76, 125]. Вони використовуються на підготовчому етапі, зокрема, коли відсутні статистичні дані (наприклад, для виняткових, нових або малоймовірних ризиків), і доводиться залучати експертів і фахівців з даної предметної області. Експертний підхід може використовуватись як в явному виді (наприклад, у мережах Байєса, нечітких моделях), так і у неявному вигляді, визначаючи сукупність характеристик, необхідних для формування статистичних вхідних вибірок [126]; обираючи критерії оптимальності; при організації процесу моделювання і формуванні навчальних та перевірочних вибірок.

Статистичні методи оцінювання ризиків базуються на використанні статистичних даних та використанні функцій, визначених емпіричним шляхом (різноманітні варіації моделей Альтмана, VAR-методів тощо) [10, 15, 198-200,

216, 258]. Аналітичні методи умовно розділяють на 2 категорії. До першої категорії відносяться дюрація, стрес-тестування, GAP-аналіз, тобто вони базуються на принципі відслідковування змін показників ризику (рівня та рейтингу) при змінах вхідних параметрів і змінних. Такі методи використовуються здебільшого при аналізі ФР юридичних осіб, банків, наприклад, з метою визначення їх платоспроможності в разі одночасного виникнення декількох несприятливих ситуацій, і перевірці достатності забезпечення і мінімальних резервів [125]. Друга категорія, що ґрунтується на використанні методів інтелектуального аналізу даних та нечіткої логіки [71, 220, 370, 371], передбачає розрахунок показників фінансових ризиків, тобто прогнозування ймовірності ризику та можливих втрат на основі попередньої статистичної інформації, застосуванні експертного знання. Методи другої групи передбачають розробку скорингових моделей на навчальній вибірці [203, 213, 237, 270, 287], їх тестування та удосконалення, повторну перевірку (цей процес має бути автоматизованим і може повторюватись ітераційно) та надання інструменту для поточного оцінювання фінансового ризику на робочі місця, наприклад, у вигляді скорингових карт [106, 108-111, 164].

Комбіновані методи, запропоновані і розроблені у дисертації (описані у третьому та п'ятому розділах), передбачають використання кількох методів (статистичних, імовірнісних, експертних, аналітичних) з метою поглибленого аналізу фінансових ризиків, урахування невизначеностей різної природи [107] та отримання вищих оцінок якості для моделей та якісніших прогнозних оцінок.

3.3. Опрацювання невизначеностей в системній методології аналізу ризиків

Серед можливих невизначеностей, що мають місце в процесі обробки даних, моделювання і прогнозування можуть бути [93, 107, 156]: невизначеність даних; невизначеності структури і параметрів моделі;

невизначеність, пов'язана з обчислювальним процесом, наприклад, систематичні обчислення і помилки методу [74, 156]. Невизначеність даних викликана наступними причинами: неповні вимірювання, короткі вибірки і відсутність даних; вплив зовнішніх випадкових збурень на функціонування системи при дослідженні; похибка вимірювань. Існує велика кількість різних методів заповнення відсутніх даних, такі як алгоритм максимізації очікування (EM), ковзне середнє і прогнозування на основі регресії, методи оптимізації, заповнення за відповідним розподілом і т. д. [93, 168, 169, 245, 254, 311, 340]. Негативний вплив зовнішніх випадкових збурень і похибок вимірювань може бути частково усунений шляхом застосування фільтра Калмана, еліптичної фільтрації, експоненціального згладжування, тощо [26, 74, 217].

Узагальнюючи огляд стандартних підходів до відновлення різних видів відсутніх даних, запропоновано відповідну методика [297] для глибинного аналізу причин появи відсутніх даних і вибору методів їх заповнення.

Методика аналізу причин появи невизначеностей і заповнення відсутніх значень

Крок 1. При виявленні факту відсутності даних при аналізі вибірки слід з'ясувати характер і, за можливістю, провести класифікацію відсутніх даних. Це можуть бути категоріальні або числові відсутні значення [297]. У разі категоріальних змінних один з можливих методів обробки є виділення їх в окрему категорію. Для числових відсутніх значень і категоріальних відсутніх значень, які все ще потребують додаткового аналізу, слід перейти на крок 2.

Крок 2. Провести поглиблений аналіз причин появи відсутніх даних і джерел простору даних: для оцінки відсутніх значень в першу чергу необхідно вивчити можливість в цілому і оцінити необхідні зусилля для відновлення відсутніх значень. Якщо немає можливостей для уточнення відсутніх значень, мають бути виявлені і виправлені ключові причини відсутності даних. Якщо причиною відсутності даних є той факт, що ці дані не були зібрані до або не збираються зараз, це означає, що існує деяка зміна процесу збору даних і ця

інформація про зміни фіксується в архівах для певних продуктів із зазначенням дати, коли ці зміни відбулися [297]. Якщо немає інформації про такі зміни, то необхідно з'ясувати характер відсутніх значень: систематичний (наприклад, в якійсь області персонал систематично не збирає деякі характеристики про клієнтів) або випадковий. Якщо ця інформація не пов'язана з особливістю системи або співробітників, або ми певними збоями в програмному забезпеченні, то необхідно визначити, чи є відсутні дані реальною загрозою якості аналізу.

Крок 3. Аналіз пропусків, пов'язаних з системними відсутніми даними через збої у роботі персоналу або системи. Причинами появи таких пропусків можуть бути наступні: 1) випадковий пропуск інформації (наприклад, не є обов'язковими для заповнення клієнтом ІПН через релігійні переконання), 2) клієнт не хоче надати певну інформацію; 3) клієнт робить помилки в наданні деякої інформації, і є реальна можливість їх встановити і відновити (умисне спотворення інформації). Останні дві причини вельми істотні при аналізі і можуть бути підставою для віднесення цих клієнтів до ненадійних і відмові їм у наданні певних послуг або товарів [126]. Тому такі випадки повинні включатися окрему категорію в наступному кроці.

Крок 4. Глибинний аналіз причин появи відсутніх значень і визначення найгіршого випадку (причини 2 і 3 на кроці 3) з заміною відсутнього значення найменшим / найбільшим значенням, яке є найгіршим варіантом для цієї змінної.

Крок 5. Використання спеціальних інструментів і методів для відновлення відсутніх значень. Для різних інтелектуальних платформ існує досить багато методів заповнення відсутніх даних. Алгоритми, такі як Zet, Zetbraid і т.д., встановлюють пропущене значення, знаходячи найбільш подібні характеристики інших показників у наборі даних, або замінюють середнім значенням для цієї характеристики [93, 169].

Крок 6. Використання відновлених значень для подальшого моделювання.

Для застосування запропонованої методики при вирішенні практичних задач було розроблено комбінований метод, який дозволяє врахувати описані вище невизначеності та відновити втрачені дані.

3.4. Комбінований метод обробки неповних даних при моделюванні фінансових ризиків

Метод реалізується у вигляді послідовного виконання кроків, що реалізують перевірку повноти, інформативності та якості даних. Далі здійснюється аналіз систематичності втрачених або пропущених даних і здійснюється їх відновлення за допомогою побудови множини регресійних моделей та обрання кращої з них.

Крок 1. Оцінювання неповноти даних в цілому по вибірці для кожної характеристики.

Якщо критерій оцінки кількості пропусків, описаний у п. 3.1.5, $I_{j(missing)} > p_{cut-off} \%$, то змінна-характеристика виключається з моделювання і пропущені значення за цією характеристикою не має сенсу відновлювати.

Крок 2. Аналіз змінних та систематичності появи пропущених значень.

2.1. Для категоріальної змінної виділення пропущених значень в окрему категорію – заповнення пропусків значенням:

$$V_{категор} := \text{“Missing”}.$$

2.2. Для всіх числових змінних, які містять пропущені значення, здійснюємо аналіз їх появи (S-systematic):

$$S_{j_{num}} = \begin{cases} 1 - \text{для систематичних пропусків, де } I_{j(missing)} \geq p_{cut-off} \% \\ 0 - \text{для несистематичних пропусків, де } I_{j(missing)} < p_{cut-off} \% \end{cases}.$$

Крок 3. Аналіз причин та наслідків появи пропусків

Використовується мережа Байєса для встановлення причинно-наслідкових зв'язків між змінними та аналізу наслідків появи пропуску. Цільова (прогнозована) змінна для МБ – наслідки.

3.1. Для $S_{j_{num}} = 1$ необхідно проаналізувати причини та наслідки появи пропусків.

$$C_j = \begin{cases} 1 - \text{випадкові} \\ 2 - \text{критичні} \\ 3 - \text{катастрофічні} \end{cases}$$

3.2. Якщо для j -ї змінної $S_{j_{num}} = 0$, $C_j = 1$, то всі i -ті пропущені значення замінюють як:

$$v_{ji} = \begin{cases} 0 \\ \text{мода} \end{cases}, \text{ де } V_{j_{num}} = \begin{pmatrix} v_{j1} \\ v_{j2} \\ \blacktriangle \\ v_{j4} \end{pmatrix} \text{ } i\text{-те значення є пропуском.}$$

3.3. Інакше застосовуємо регресійне рівняння для прогнозування пропущених значень.

Крок 4. Побудова та вибір кращої моделі для відновлення пропущених значень.

4.1. Регресійне моделювання

Для відновлення втрачених даних пропонується використати прогнозу модель регресійного типу. Для цього можлива побудова авторегресійної моделі (АР), авторегресії з ковзним середнім, або множини гетероскедастичних моделей (АРУГ, УАРУГ, ЕУАРУГ, тощо).

При побудові моделі регресії першого порядку AR(1) [17]:

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k), \quad E[\varepsilon(k)] = 0, \quad (3.9)$$

прогнозування на один крок вперед:

$$y(k+1) = a_0 + a_1 y(k) + \varepsilon(k+1),$$

Якщо коефіцієнти a_0, a_1 відомі, то прогноз втраченого значення знаходять як умовне математичне сподівання [26]:

$$\begin{aligned}\hat{y}(k+1, k) &= E_k[y(k+1)] = E_k[y(k+1) | y(k), y(k-1), \dots, \varepsilon(k), \varepsilon(k-1), \dots)] = \\ &= a_0 + a_1 E_k[y(k)] = a_0 + a_1 y(k),\end{aligned}$$

Використовуючи знову той самий підхід, можна знайти прогноз на два кроки вперед (3.9) [26]:

$$\begin{aligned}y(k+2) &= a_0 + a_1 y(k+1) + \varepsilon(k+2), \\ \hat{y}(k+2, k) &= E_k[y(k+2)] = a_0 + a_1 E_k[y(k+1)] = a_0 + a_1 E_k[a_0 + a_1 y(k)] = \\ &= a_0 + a_0 a_1 + a_1^2 y(k).\end{aligned}$$

За індукцією можна виконати прогноз на три кроки вперед в такий спосіб:

$$\hat{y}(k+3, k) = E_k[y(k+3)] = a_0 + a_0 a_1 + a_0 a_1^2 + a_1^3 y(k).$$

На s -кроків вперед прогноз може бути обчислений за функцією [297]:

$$\hat{y}(k+s, k) = E_s[y(k+s)] = a_0 \left(\sum_{i=0}^{s-1} a_1^i \right) + a_1^s y(k) = a_0 \sum_{i=0}^{s-1} a_1^i + a_1^s y(k). \quad (3.10)$$

Рівняння (3.10) називається функцією прогнозування для довільного числа кроків-вперед. Послідовність прогнозів є збіжним процесом, якщо виконується умова: $|a_1| < 1$, тобто

$$\lim_{s \rightarrow \infty} E_k[y(k+s)] = \frac{a_0}{1 - a_1}, \quad |a_1| < 1, \quad (3.11)$$

де a_1 – знаменник геометричної прогресії в правій частині (3.10). Вираз (3.11) показує, що для будь-якого стаціонарного АР процесу або процесу АРКС послідовність умовних прогнозних оцінок асимптотично збігається $s \rightarrow \infty$, до безумовного середнього значення [26].

Далі виконується розширення функції прогнозування в процесі АР (р) у вигляді рекурсії:

$$\hat{y}(k+s, k) = a_0 + \sum_{i=1}^p a_i \hat{y}(k+s-i),$$

$$\text{де } \hat{y}(k+s-i) = E_k[y(k+s-i)].$$

За аналогічним підходом формується множина функцій прогнозування регресійного типу, які застосовують для відновлення пропущених значень, оцінюється точність прогнозування.

4.2. Застосування альтернативних методів відновлення пропусків (наприклад, заповнення середнім значенням, нулем або EM-алгоритмом). Оцінка якості прогнозування за критеріями оцінки точності прогнозів (MAPE, MSE, RMSE, тощо).

4.3. Обрання кращої моделі за сукупністю критеріїв якості прогнозів та використання її у подальшому для відновлення втрачених або пропущених значень.

Крок 5. Застосування відновлених даних для моделювання.

Відновлені дані включаються у вибірку даних і застосовуються для подальшої розробки моделей оцінювання фінансових ризиків.

Твердження про комбінований метод обробки неповних даних. Застосування комбінованого методу заповнення пропусків надає вищі оцінки з точки зору оцінювання якості даних та їх подальшого моделювання, оскільки виконується глибинний аналіз причин появи пропусків та використовуються ідеологічно різні методи для відновлення втрачених даних (регресійні, статистичні, ймовірнісні і ймовірнісно-статистичні).

Це твердження проілюструємо прикладом. Якщо відновлення втрачених даних здійснюється за допомогою регресійного моделювання, то це однозначно зсуває вибірку в бік регресійного рівняння, що використовувалось для заповнення пропусків (наприклад, рівняння $AR(2)$). Якщо доля пропущених значень становила 10%, то в подальшому, при застосуванні цієї самої вибірки для моделювання ризиків різної природи, приблизно на 10% збільшується ймовірність обрання в якості кращої моделі регресійного рівняння типу $AR(2)$. Для навчальної вибірки ця модель дійсно дає кращі результати, оскільки частина даних згенерована за законом регресійного рівняння другого порядку, проте на тестовій вибірці розподіл даних може мати деякі відмінності. Однак, у

загальному випадку це не критично для оцінювання структури і параметрів моделі.

Слід зазначити, що обмеженням для використання даного методу є необхідність достатньо потужного набору даних, у якому є можливість вручну видалити пропущені значення та їх відновити для того, щоб перевірити коректність розробленої моделі. Тоді буде можливість встановити і точність прогнозу відновлення, бо будуть наявні початкові значення вхідних змінних. За такого підходу є можливість побудови множини моделей-кандидаток на тестовому (видаленому вручну та відновленому) наборі, і обрання для нього кращої моделі серед кандидаток. Далі ця модель буде використовуватись для відновлення реально пропущених/втрачених даних. Були виконані обчислювальні експерименти із застосуванням даного методу, де на четвертому кроці використовувалась побудова регресійних моделей, застосовувався ЕМ-алгоритм, тощо. Були отримані реальні результати відновлення штучно видалених даних з досить високою точністю. Ілюстрація прикладу такого моделювання на реальних даних наведена нижче.

Ілюстрація роботи методу заповнення пропусків.

Для прикладу моделювання і відновлення пропущених значень було обрано вибірку, що складається з 500 значень (ціни акцій). Спочатку модель прогнозування була побудована для всієї вибірки даних. Потім 25 точок даних були штучно видалені (усунені значення з 301 до 325), а кілька альтернативних методів були використані для оцінювання цих пропущених значень (результати наведені у таблиці 3.2).

Для даного набору даних найкращою моделлю за критерієм адекватності виявилась авторегресійна модель першого порядку $AR(1)$ з трендом третього порядку (t^3), а кращими методом відновлення пропущених даних виявився ЕМ-алгоритм. Середня абсолютна помилка у відсотках (MAPE) для одного кроку прогнозування близько 3,21%, що цілком можна порівняти з 3,19%, обчисленою для повного зразку [297].

Таблиця 3.2

Результати відновлення даних та прогнозування за комбінованим методом

Умови побудови моделей	Найкраща модель	Якість моделі		Якість прогнозу	
		R^2	DW	$RMSE$	$MAPE$
Повна вхідна вибірка	AR(1) + t3	0.993	2.12	25.92	3.19
Втрачені дані заповнені нулем (реальні пропуски)	AR(1) + t3	0.31	0.10	118.14	6.51
Заміна пропусків середнім значенням вибірки	AR(1) + t3	0.990	1.97	27.97	3.45
Заміна пропусків прогнозом	AR(1) + t3	0.991	2.18	26.78	3.27
Заміна пропусків за EM-алгоритмом	AR(1) + t3	0.992	2.12	25.97	3.21

Адекватність моделі перевірялась за коефіцієнтом детермінації та статистикою Дарбіна-Уотсона, і виявилась майже однаковою в обох випадках. Найгірший результат був отриманий при заміні відсутніх значень нулями. В цьому випадку $MAPE = 6,51\%$, що в два рази більше в порівнянні з повною вибіркою. А коефіцієнт детермінації зменшився більш ніж в три рази (з 0,993 до 0,31) [297]. Оскільки, як правило, важко передбачити, який метод дасть кращий кінцевий результат, бажано використовувати альтернативні обчислювальні процедури і обирати кращу модель для конкретного випадку, використовуючи відомі статистичні критерії якості.

Враховуючи значні обсяги даних, які потрібно обробляти під час аналізу фінансових ризиків, слід намагатися автоматизувати процес аналізу даних [155], спрямований на поліпшення його якості із заповненням відсутніх значень. Автоматична процедура передбачає застосування різних методів, що забезпечують не лише відновлення пропущених даних, а й попереднє згладжування, боротьбу з шумами та збуреннями даних, дають різні проміжні результати, такі як короткострокове прогнозування та комплексування.

3.5. Статичне оцінювання фінансових ризиків

Алгоритм роботи для статичного оцінювання фінансових ризиків наведений на рис. 3.4. На першому етапі здійснюється оцінювання фінансового ризику, виявляються притаманні даній фінансовій системі ризики і розробляються моделі їх кількісної оцінки. Оцінка фінансових ризиків дає можливість визначити можливі втрати при коливаннях ринку і розмір капіталу, який необхідно резервувати для покриття цих втрат.

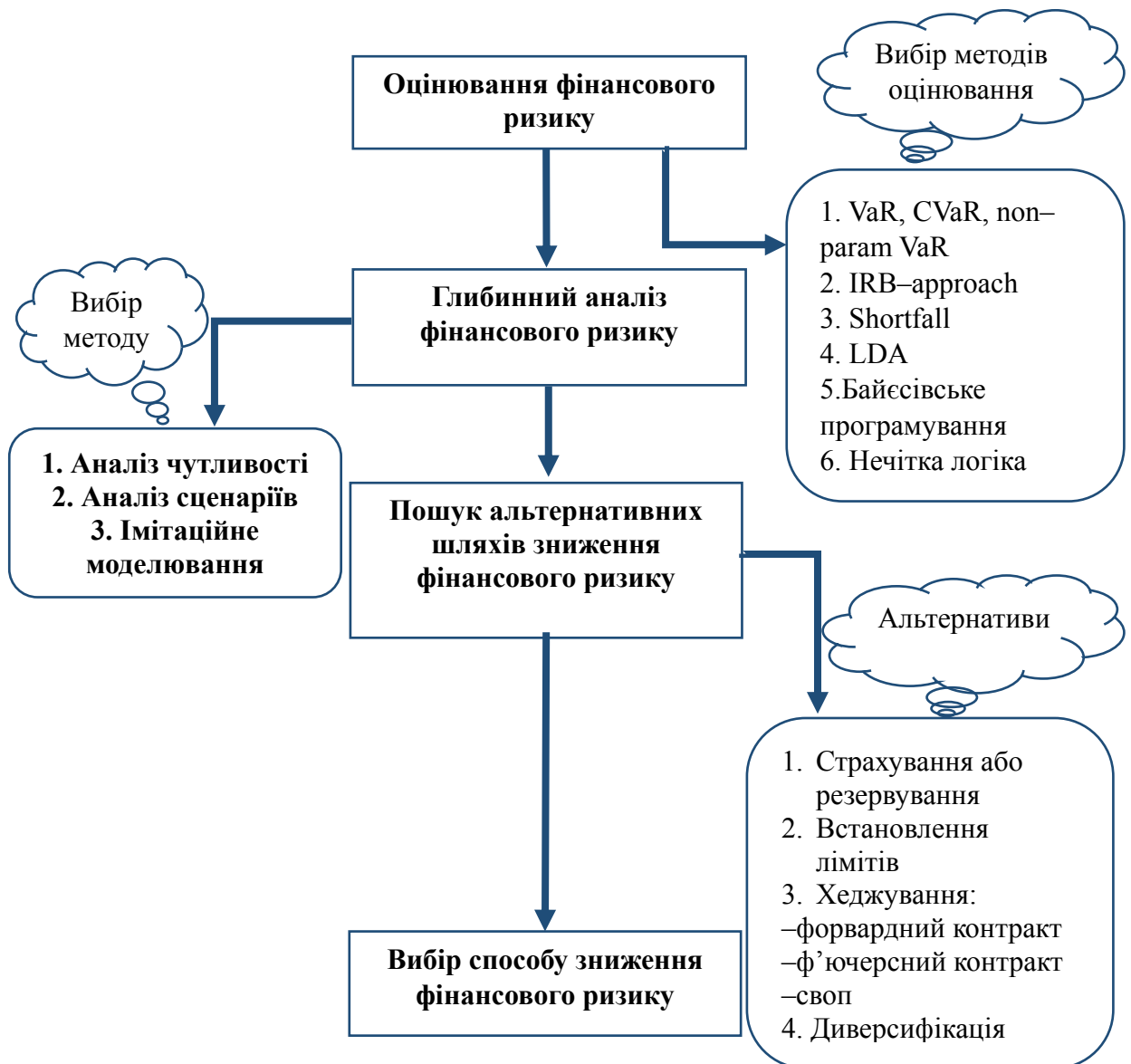


Рис. 3.4 Статичне оцінювання фінансових ризиків

На сьогодні розроблено множину методів оцінювання фінансових ризиків, зокрема різноманітні варіації VaR (Value-at-Risk), методи на основі

IRB-підходу, Shortfall, LDA, методи з використанням байєсівського програмування та нечіткої логіки, докладно розглянуті у першому розділі. Методи VaR оцінюють ризик як очікуваний максимальний збиток протягом встановленого періоду часу із встановленим рівнем ймовірності; Shortfall є більш консервативним, ніж VaR метод оцінювання ризику, тобто оцінювання ризику того, що фактичний прибуток інвестиції буде меншим, ніж очікуваний дохід; за методом LDA виконується оцінювання розподілу втрат для розрахунку суми капіталу банку для операційного ризику [10].

Методи байєсівського програмування та нечіткої логіки є більш універсальними і можуть застосовуватись до різних видів фінансового ризику, зі встановленням правил та критеріїв і формуванням висновку щодо рівня ризику (нечітка логіка) [71, 77]. Байєсівське програмування включає сукупність таких методів, як мережі Байєса для прогнозування ймовірності ризику, байєсівська регресія для оцінювання рівня ризику і можливих втрат, гранулярна фільтрація та байєсівський класифікатор для обробки вхідних даних і моделювання багатовимірних розподілів.

При виконанні наступного етапу здійснюють глибинний аналіз фінансових ризиків [115], зокрема визначення значень ключових параметрів, які можуть поставити під сумнів успіх бізнесу (аналіз чутливості), аналіз різноманітних сценаріїв, розглядаючи альтернативні набори вхідних даних, які можуть з'явитись у реальній ситуації; імітаційне моделювання, тобто імітаційні комп'ютерні прогони моделей, що оцінюють фінансові показники.

На третьому етапі відбувається пошук шляхів зниження фінансового ризику: за рахунок страхування або резервування (без зменшення ймовірності появи ризиків, а лише з орієнтацією на відшкодування матеріальних збитків від прояву ризиків) [115]; встановленням лімітів операцій; хеджуванням (мінімізація цінового ризику з метою фіксації певного рівня цін) та диверсифікацією (зменшення сукупної схильності до ризику за рахунок розподілу коштів між різними активами, ціна або прибутковість яких слабо

корельовані між собою) [126]. Диверсифікація полягає у зниженні значення максимально можливих втрат за одну подію, однак при цьому одночасно зростає кількість видів ризику, які необхідно контролювати. Для хеджування можна використовувати біржові (товарні ф'ючерси та опціони) та позабіржові інструменти (форвардні контракти, свопи тощо).

На останньому, четвертому етапі приймається управлінське рішення щодо фінансового ризику [115]. На основі всіх проаналізованих альтернатив, методів зниження збитків обирається та альтернатива (або навіть комбінація альтернатив), яка дозволяє отримати найменші фінансові втрати і тим самим збільшити прибутковість.

3.6. Динамічне оцінювання ризиків

Узагальнюючи описане вище статичне оцінювання, можна зазначити необхідність реалізації окремих методів обробки невизначеностей, пошуку кращої альтернативи оцінювання ймовірності та рівня ризику, тобто можливих втрат. Крім того, у багатьох випадках важливим є визначення моменту настання ризику.

Принцип динамічного оцінювання ризиків передбачає, що для оцінювання рівня втрат слід визначити очікувані моменти настання допустимого, критичного і катастрофічного ризиків, ймовірність настання таких ризиків, і розмір можливих втрат. Саме момент, в який переходить і змінюється рівень ризику від допустимого до катастрофічного, є найбільш важливим для оцінювання рівня втрат. Ймовірність прояву ризиків також оцінюється і змінюється в часі, і момент, в який ймовірність прояву ризику різко зростає, також може бути оцінений [115]. На рис. 3.5 подана характеристика основних зон ризику через ймовірність прояву ризику та обсяг його можливих втрат.

Визначення. *Динамічне оцінювання ризиків* – це оцінювання ризиків за ймовірністю, втратами та прогнозування часу (моменту) переходу ризику на більш високий (критичний) рівень з точки зору ймовірності або втрат

$DE = \langle PR, Losses, t, S(t|x), \lambda(t|x) \rangle$, де PR – ймовірність настання ризику; $Losses$ – рівень максимально можливих втрат; t – час; $S(t|x)$ – функція умовного виживання, тобто подальше функціонування фінансової системи навіть після прояву ризику; $\lambda(t|x)$ – умовний рівень небезпеки, тобто рівень втрат в момент часу t .

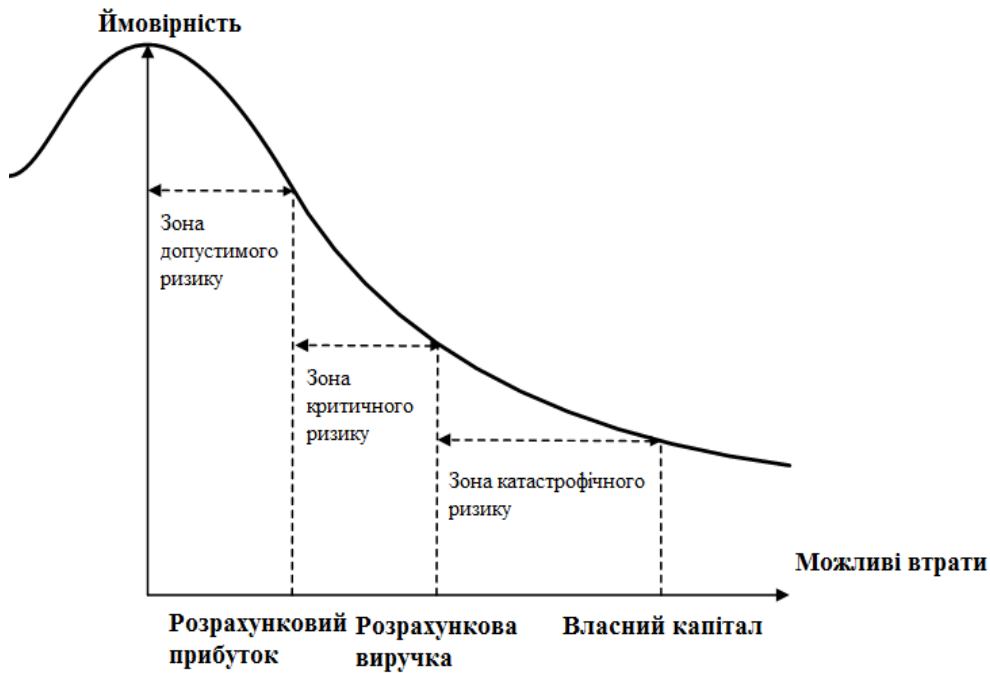


Рис. 3.5 Характеристика основних зон ризику

Динамічне оцінювання відрізняється від статичного передбаченою можливістю оцінювання ризиків у явному вигляді в динаміці, тобто прогнозування функції втрат та ймовірності ризику (переходу на вищий ступінь: критичний, катастрофічний) – як функцій часу.

Стратегія, яку компанія обере в подальшому для своєї роботи, буде суттєво залежати від її фінансових можливостей та толерантності до ризику, тобто того, який рівень ризику і, відповідно, можливих втрат, вона може на себе прийняти [115]. Формально це множина кривих – функцій виживання.

$S(t_1|x)$ – функція умовного функціонування фінансової системи на рівні допустимого ризику, в якому вона успішно працює до часу t_1 . В цей момент розмір можливих втрат оцінюється через недоотримання розрахункового прибутку.

$S(t_2 | x)$ – функція умовного функціонування фінансової системи на рівні критичного ризику (в момент часу від t_1 до t_2). Можливі втрати вираховуються з доходу компанії.

$S(t_3 | x)$ – функція умовного функціонування ФС на рівні катастрофічного ризику (в момент часу від t_2 до t_3), коли система несе втрати з власного капіталу.

$\lambda(t | x)$ – умовний рівень небезпеки [115] (перехід у зону катастрофічного рівня), що призводить до великого відтоку клієнтів і великих фінансових втрат.

Ключові моменти, які потребували розробки та доопрацювання для статичного та динамічного оцінювання, наведені у таблиці 3.3.

Таблиця 3.3

Статичне та динамічне оцінювання ризиків

Статичне оцінювання	
Проблеми та обмеження	Методи та способи подолання
Неповнота вхідних даних	Адаптація для коротких вибірок, комбінований метод обробки неповних даних, урахування інформаційних ризиків
Неструктурованість вхідних даних	Критерії для формування структури моделі R^2, χ^2, IV, WOE
Недостатня ефективність існуючих методів ІАД	Оцінювання ризиків інтегрованими та комбінованими моделями (нейро-нечіткі методи, дерева рішень, регресійні та байєсові моделі)
Визначення міри ступеня ризику	Скорингові карти
Критерії якості	$GINI, CA, BS$ та розроблені власні критерії (квадратичний критерій та критерій ефективності)
Динамічне оцінювання	
Момент настання ризику	Визначення часу з використанням параметричних, напівпараметричних та непараметричних моделей
Визначення ймовірності настання фінансового ризику	Використання функції ризику (hazard function)
Оцінювання втрат на конкретний момент часу	1) Функції виживання; 2) Ймовірісно-статистичний метод оцінювання втрат

Встановлення рівня небезпеки і ключових моментів часу, які характеризують допустимий, критичний та катастрофічний рівень ризику, є задачею системного аналізу, яку необхідно вирішувати в рамках кожного виду

ризиків незалежно від типу ризику та галузі, в якій він спостерігається [115]. У дисертації пропонується підхід, що базується на визначенні втрат компанії як допустимих $\lambda(t_1 | x) = c_1$, критичних $\lambda(t_2 | x) = c_2$ та катастрофічних $\lambda(t_3 | x) = c_3$, де c_1, c_2, c_3 – певні константи, які визначаються компанією в залежності від її фінансових обертів, потужностей, тощо (наприклад, обсягу власного капіталу).

Для визначення моментів часу t_1, t_2, t_3 запропоновано метод динамічного оцінювання ризиків на основі динамічних моделей виживання та алгоритми визначення моментів часу на основі допустимих втрат та ймовірності, які детально описані у розділі 4.

3.7. Ймовірнісно-статистичний метод оцінювання ризику фінансових втрат

Для математичного моделювання ризиків фінансових втрат застосовують різні методи залежно від типу ризику, даних стосовно розвитку фінансових процесів та методів їх обробки, наявності необхідних інструментальних засобів для моделювання, оцінювання і прогнозування фінансових процесів та досвіду використання тих чи інших методів і підходів. Широке застосування для розв'язання задач моделювання і оцінювання ризиків фінансових втрат знаходять ймовірнісно-статистичні методи і моделі, які ґрунтуються на комбінуванні різних підходів [24, 121, 312], що дає можливість використовувати переваги ідеологічно різних методів та отримувати оцінки можливих втрат у формі точкових оцінок та ймовірностей настання ризикових ситуацій.

Слід відмітити, що наявні статистичні дані потребують, як правило, застосування методів попередньої обробки, спрямованих на заповнення пропусків вимірів, нормування, фільтрації даних з метою зменшення впливу похибок вимірів та випадкових зовнішніх збурень, бутстреп-аналізу і т. ін [244, 245]. Математичні моделі, які будуються на основі статистичних даних,

потребують адаптації у процесі надходження нових вимірів і застосування нетрадиційних оптимізаційних процедур для оцінювання їх параметрів [115].

Пропонується побудова комбінованої моделі на основі оптимального фільтра, регресійної моделі і байєсівської мережі для оцінювання ризику можливих фінансових втрат. Структура ймовірно-статистичного методу оцінювання ризику фінансових втрат подана на рис. 3.6.

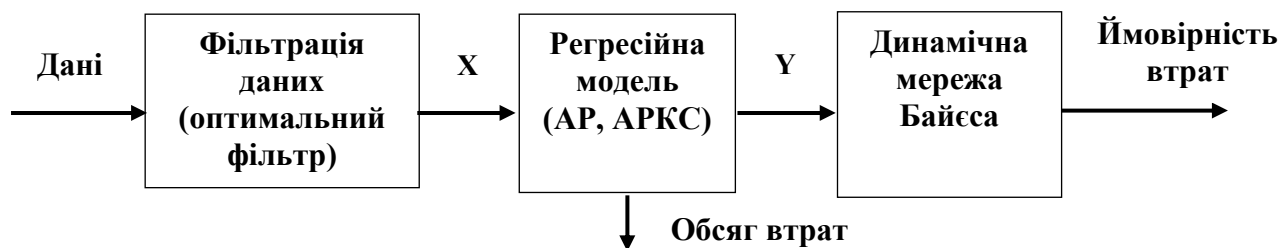


Рис. 3.6 Структурна схема ймовірно-статистичного методу оцінювання ризику фінансових втрат

Ймовірно-статистичний метод оцінювання ризику фінансових втрат реалізується у вигляді кроків, поданих нижче [19].

Крок 1. Вхідні дані, які характеризують можливу появу фінансових втрат, можуть бути різного типу і різних форматів. Статистичні дані характеризуються, як правило, неточністю, неповнотою, нечіткістю інформації, наявністю похибок вимірів та впливом випадкових зовнішніх збурень. Тому виникає необхідність здійснити фільтрацію даних. Для цього можна застосувати цифрові або оптимальні фільтри, наприклад, фільтр Калмана [74]. Застосування цифрових фільтрів передбачає знання смуги частот, у якій знаходяться корисні складові даних, і смуги, де знаходяться згадані шкідливі випадкові впливи, які фільтр не повинен пропускати. На основі цієї інформації проектується цифровий фільтр у вигляді лінійних рівнянь авторегресії (АР) або авторегресії з ковзним середнім (АРКС), які будуть використовуватись для попередньої обробки даних згідно схеми, поданої на рис. 3.6.

Статистичні дані, що описують фінансовий ризик за різними показниками (ліквідності, обслуговування боргів і т. ін.), можуть бути представлені у табличному вигляді, а змінна або вектор змінних стану, що характеризують

можливі фінансові втрати в часі, може бути описаний рівнянням динаміки у просторі станів [19]:

$$\mathbf{x}(k+1) = \mathbf{F}\mathbf{x}(k) + \mathbf{w}(k), \quad (3.12)$$

$$\mathbf{z}(k) = \mathbf{H}\mathbf{x}(k) + \mathbf{v}(k), \quad (3.13)$$

де $\mathbf{x}(k)$ – вектор стану досліджуваного процесу; \mathbf{F} – перехідна матриця станів; $\mathbf{w}(k)$ – векторний процес зовнішніх випадкових збурень з коваріаційною матрицею $E[\mathbf{w}(k)\mathbf{w}^T(j)] = \mathbf{Q}(k)\delta_{kj}$; $\mathbf{z}(k)$ – вектор вимірів; \mathbf{H} – матриця коефіцієнтів вимірів; $\mathbf{v}(k)$ – вектор похибок вимірів з коваріаційною матрицею $E[\mathbf{v}(k)\mathbf{v}^T(j)] = \mathbf{R}(k)\delta_{kj}$. У класичній постановці задачі [74] випадкові процеси $\mathbf{w}(k)$ і $\mathbf{v}(k)$ некорельовані між собою та вектором стану і мають нульове середнє та постійні скінченні коваріації. Система матричних рівнянь (3.12), (3.13) – це модель даних у просторі станів, яка для спрощення аналізу не враховує можливих керуючих впливів. Початковим станом системи \mathbf{x}_0 будемо вважати випадкові змінні з відомими статистиками:

$$E[\mathbf{x}_0] = \bar{\mathbf{x}}_0; \quad E[\mathbf{x}_0\mathbf{x}_0^T] = \mathbf{M}; \quad E[\mathbf{w}(k)\mathbf{x}_0^T] = 0, \quad \forall k.$$

Оптимальна оцінка стану $\hat{\mathbf{x}}(k)$ повинна обчислюватися як найкраща за мінімумом середнього значення суми квадратів оцінок похибок. Іншими словами, оцінка повинна бути такою, щоб

$$E[(\hat{\mathbf{x}}(k) - \mathbf{x}(k))^T (\hat{\mathbf{x}}(k) - \mathbf{x}(k))] = \min_{\mathbf{K}}, \quad (3.14)$$

де $\mathbf{x}(k)$ – точне значення вектора стану, яке може бути обчислене за допомогою детермінованої складової математичної моделі процесу; \mathbf{K} – оптимальний матричний коефіцієнт фільтра [74], який необхідно обчислити в результаті розв'язування оптимізаційної задачі. Основне рівняння фільтрації має такий вигляд:

$$\hat{\mathbf{x}}(k) = \mathbf{F}(k)\hat{\mathbf{x}}(k-1) + \mathbf{K}(k)[\mathbf{z}(k) - \mathbf{H}(k)\mathbf{F}(k)\hat{\mathbf{x}}(k-1)]. \quad (3.15)$$

На основі рівняння (3.12) можна записати функцію прогнозування на один крок як умовне математичне сподівання вектора стану [19]:

$$\hat{\mathbf{x}}(k+1|k) = E_k[\mathbf{x}(k+1)] = \mathbf{F} \mathbf{x}(k), \quad (3.16)$$

де $\hat{\mathbf{x}}(k+1)$ – прогноз на один крок на основі інформації на момент k включно. Функцією (3.16) можна скористатись для обчислення прогнозу на довільну кількість кроків. Так, прогноз на два кроки має вигляд [19]:

$$\hat{\mathbf{x}}(k+2) = \mathbf{F} \hat{\mathbf{x}}(k+1) = \mathbf{F} \cdot \mathbf{F} \mathbf{x}(k) = \mathbf{F}^2 \mathbf{x}(k),$$

і на довільну кількість кроків s :

$$\hat{\mathbf{x}}(k+s) = \mathbf{F}^s \mathbf{x}(k).$$

Очевидно, що дисперсія похибки прогнозу буде зростати пропорційно кількості кроків s . Так, похибка прогнозу на один і два кроки складає [26]

$$\mathbf{e}_f(1) = \mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1) = \mathbf{F} \mathbf{x}(k) + \mathbf{w}(k) - \mathbf{F} \mathbf{x}(k) = \mathbf{w}(k),$$

$$\begin{aligned} \mathbf{e}_f(2) &= \mathbf{x}(k+2) - \hat{\mathbf{x}}(k+2) = \mathbf{F} \mathbf{x}(k+1) + \mathbf{w}(k+1) - \mathbf{F}[\mathbf{F} \mathbf{x}(k)] = \\ &= \mathbf{F} \mathbf{x}(k+1) + \mathbf{w}(k+1) - \mathbf{F}[\mathbf{x}(k+1) - \mathbf{w}(k)] = \mathbf{w}(k+1) + \mathbf{w}(k). \end{aligned}$$

Тобто, дисперсію оцінок прогнозів на s кроків можна обчислити так:

$$\begin{aligned} \text{Var}[\mathbf{e}_f(s)] &= E\{[\mathbf{w}(k+s-1) + \mathbf{w}(k+s-2) + \dots + \mathbf{w}(k)] \times [\mathbf{w}^T(k+s-1) + \\ &+ \mathbf{w}^T(k+s-2) + \dots + \mathbf{w}^T(k)]\} = s \mathbf{Q}_w. \end{aligned}$$

Оцінювання компонент вектора стану [74], які неможливо виміряти за допомогою приладів (або зареєструвати), виконується наступним чином. У випадку наявності невимірюваних компонент розмірність вектора вимірів $\mathbf{z}(k)$ є меншою вимірності вектора стану $\mathbf{x}(k)$, тобто, $\dim[\mathbf{z}] < \dim[\mathbf{x}]$.

Оптимальний коефіцієнт фільтра обчислюється за рівнянням [19]:

$$\mathbf{K}(k) = \mathbf{P}'(k) \mathbf{H}^T [\mathbf{H} \mathbf{P}'(k) \mathbf{H}^T + \mathbf{R}]^{-1},$$

де $\dim[\mathbf{P}'(k)] = [n \times n]$; $\dim[\mathbf{H}^T] = [n \times r]$ за визначенням; n – розмірність вектора стану і $\dim[\mathbf{H} \mathbf{P}'(k) \mathbf{H}^T + \mathbf{R}]^{-1} = [r \times r]$. Таким чином, $\dim[\mathbf{K}(k)] = [n \times r]$. Вектор нев'язок $\gamma(k)$ в рівнянні оцінювання (3.15)

$$\gamma(k) = \mathbf{z}(k) - \mathbf{H} \mathbf{F} \hat{\mathbf{x}}(k)$$

має розмірність $[r \times 1]$, а розмірність добутку $\dim[\mathbf{K}(k)\mathbf{v}(k)] = [n \times 1]$. Наприклад, якщо $\dim[\mathbf{x}] = [3 \times 1]$, а $\dim[\mathbf{z}] = [2 \times 1]$, то добуток $\mathbf{K}(k)\boldsymbol{\gamma}(k)$ має вигляд:

$$\mathbf{K}(k)\boldsymbol{\gamma}(k) = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \\ K_{31} & K_{32} \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}$$

і оцінка вектора стану визначається за рівняннями:

$$\begin{aligned} \hat{x}_1(k) &= \hat{x}_1(k, k-1) + K_{11}\gamma_1 + K_{12}\gamma_2, \\ \hat{x}_2(k) &= \hat{x}_2(k, k-1) + K_{21}\gamma_1 + K_{22}\gamma_2, \\ \hat{x}_3(k) &= \hat{x}_3(k, k-1) + K_{31}\gamma_1 + K_{32}\gamma_2. \end{aligned}$$

Тобто, невимірювана третя компонента вектора стану може бути оцінена, якщо матриці $\mathbf{P}'(k)$, $\mathbf{P}(k)$ і $\mathbf{K}(k)$ мають ненульові відповідні елементи [17]. Таким чином, фільтр дає можливість обчислити оцінки змінних, які є оптимальними у сенсі мінімуму функціоналу (3.16) і знаходяться під значно меншим впливом випадкових збурень стану і похибок вимірів.

Крок 2. На наступному кроці будується регресійна модель певної структури, яка здійснює прогнозування показників фінансового ризику втрат (фінансової стійкості) на наступний момент часу [19]. Як правило, існує можливість побудови деякої множини моделей-кандидаток у вигляді авторегресії (векторної авторегресії) або авторегресії з ковзним середнім з різними параметрами для авторегресії та ковзного середнього, які будуть здійснювати прогнозування значень фінансових змінних на 1, 2, 3, ... кроки (періоди дискретизації) вперед [26]. Цим самим забезпечується оцінювання прогнозних значень можливих втрат. При цьому для зручності обчислення прогнозів доцільно будувати і використовувати функції прогнозування на основі розв'язків рівнянь [25]. Так, для моделі авторегресії першого порядку функція прогнозування втрат має вигляд:

$$\hat{y}(k+s) = E_k[y(k+s)] = a_0 \left(\sum_{i=0}^{s-1} a_1^i \right) + a_1^s y(k) = a_0 \sum_{i=0}^{s-1} a_1^i + a_1^s y(k),$$

де $\hat{y}(k+s)$ – оцінка прогнозу основної (цільової) змінної на s кроків; E_k – оператор умовного математичного сподівання стосовно моменту часу k ; a_0, a_1 – параметри регресійної моделі. Для моделі АРКС(2,1) побудована функція прогнозування на три кроки має вигляд:

$$\begin{aligned} \hat{y}(k+3) = E_k[y(k+3)] &= a_0 + a_1 E_k[y(k+2)] + a_2 E_k[y(k+1)] = a_0(1 + a_1 + a_1^2 + a_2) + \\ &+ (a_1^3 + 2a_1 a_2)y(k) + (a_1^2 a_2 + a_2^2)y(k-1) + \beta_1(a_1^2 + a_2)\varepsilon(k). \end{aligned}$$

Табулювання функцій прогнозування дає можливість використовувати їх повторно.

Похибка прогнозування $f_k(k)$ для моделей типу АР та АРКС визначається випадковою складовою $\varepsilon(k)$ і оцінюється так:

$$f_k(s) = \varepsilon(k+s) + a_1 \varepsilon(k+s-1) + a_1^2 \varepsilon(k+s-2) + \dots + a_1^{s-1} \varepsilon(k+1). \quad (3.17)$$

Враховуючи те, що $E[f_k(s)] = 0$, оцінка прогнозу, яка обчислюється за виразом (3.17), є незміщеною. Дисперсія похибки прогнозування визначається за виразом

$$Var[f_k(s)] = \sigma^2 [1 + a_1^2 + a_1^4 + a_1^6 + \dots + a_1^{2(s-1)}],$$

тобто дисперсія є функцією кількості кроків прогнозування s . Асимптотичне значення дисперсії похибки прогнозу для стаціонарного процесу прямує до сталої величини:

$$\lim_{s \rightarrow \infty} Var[f_k(s)] = \frac{\sigma^2}{1 - a_1^2},$$

де a_1^2 – знаменник геометричної прогресії.

У випадку наявності пояснюючих змінних лінійна регресійна модель може мати таку структуру:

$$y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) + b_1 x_1(k-d_1) + \dots + a_m x_m(k-d_m) + \varepsilon(k), \quad (3.18)$$

де $x_i, i=1, \dots, m$ – пояснюючі змінні (предиктори); $d_i, i=1, \dots, m$ – дискретний час запізнення для пояснюючих змінних [19]. Зазначимо, що незалежні змінні, які входять у модель з фактичним часом запізнення d , називають провідними індикаторами. Очевидно, що структура моделі (3.18) може містити складові, що описують тренд, сезонні ефекти, ковзне середнє та нелінійні ефекти. Оцінювання структури моделі здійснюється на основі кореляційного аналізу наявних даних, застосування статистичних тестів на інтегрованість, гетероскедастичність, наявність нелінійності та експертних оцінок [17]. Як правило, для подальшого аналізу використовують кілька моделей-кандидаток.

Далі аналізується якість моделі, тобто виконується перевірка оцінених кандидаток на адекватність процесу. Діагностика побудованих моделей може здійснюватися наступними способами: візуальне дослідження графіка похибок, аналіз характеристик похибок [26], перевірка, чи корелюють похибки моделі між собою (критерій Дарбіна-Уотсона), чи коефіцієнт множинної детермінації R^2 прямує до 1 ($R^2 \rightarrow 1$ коли дисперсії вимірів змінної та оцінок цієї ж змінної, отриманих за моделлю, збігаються), і чи є мінімальною сума квадратів похибок для вибраної моделі, тобто

$$\sum_{k=1}^N e^2(k) = \sum_{k=1}^N [\hat{y}(k) - y(k)]^2 \rightarrow \min_{\theta}.$$

При використанні методів рекурсивного оцінювання [350] для лінійної моделі 2-3-го порядку оцінки параметрів мають збігатися до усталених значень після 30-40 ітерацій алгоритму оцінювання. Якщо кількість ітерацій набагато перевищує вказані значення, то це свідчить про те, що процес може бути нестационарним та/або нелінійним. Узагальнена методика побудови регресійних моделей та оцінювання прогнозів описана у роботах [14, 25].

В результаті виконання цього кроку отримаємо прогноз величини втрат та рівняння регресійної моделі.

Крок 3. На третьому кроці здійснюється ймовірнісне оцінювання втрат на основі адаптивної динамічної мережі Байєса (МБ). Сьогодні МБ – це потужний інструмент ймовірнісно-статистичного моделювання, який ґрунтується на спеціальних чисельних методах, які забезпечують оцінювання альтернативних структур і параметрів моделей, а також формування точного або наближеного ймовірнісного висновку – остаточного результату застосування моделі [77].

Адаптація МБ полягає у тому, що структура і параметри моделі оцінюються повторно з надходженням нових даних. При цьому часовий інтервал, через який виконується повторне оцінювання, визначається емпірично для конкретної динаміки досліджуваного процесу і статистичних даних конкретного типу.

У комбінованій моделі для оцінювання втрат пропонується застосовувати динамічну мережу Байєса [314]. Для побудови цієї моделі використовуємо виходи обраних на попередньому кроці кращих моделей авторегресії (або інших структур), які подаються на вхід стаціонарного шару ймовірнісної моделі.

Динамічні мережі Байєса – це розширення статичних байєсівських мереж для моделювання спільних розподілів ймовірностей на множині випадкових змінних Z_1, Z_2, \dots [66]. Як правило, змінні розбиваються на трійки $Z_t = (U_t, X_t, Y_t)$, що позначають множину змінних вхідного, прихованого та вихідного шарів моделі у просторі станів [314]. Надалі будемо розглядати лише моделі стохастичних процесів з дискретним часом, тобто індекс t буде зростати з появою кожного нового спостереження.

Оскільки динамічна мережа Байєса визначається як пара МБ (B_I, B_{\rightarrow}) , де B_I – МБ, що визначає апіорну ймовірність $P(Z_I)$, а B_{\rightarrow} – двошарова МБ, що визначає ймовірності $P(Z_t | Z_{t-1})$ таким чином [19]:

$$P(Z_t | Z_{t-1}) = \prod_{i=1}^N P(Z_t^i | Pa(Z_t^i)),$$

де Z_t^i – i -й вузол в момент часу t , що може бути компонентою X_t, Y_t або U_t , а $Pa(Z_t^i)$ – батьківські вершини Z_t^i на графі. Вузли першого шару двошарової МБ не мають жодних параметрів, що з ними асоціюються, але кожен вузол другого шару має зв'язаний з ним розподіл умовної ймовірності, що визначають $P(Z_t^i | Pa(Z_t^i))$ для всіх $t > 1$. Батьки вершини $Pa(Z_t^i)$ можуть бути або ж у тому самому, або у попередньому часовому шарі. Проте, не існує суворих математичних обмежень стосовно того, що батьківські вершини знаходяться не далі, ніж в сусідньому шарі, і не можуть знаходитися, скажімо, через один шар. Дуги між шарами спрямовуються зліва направо, що позначає напрям протікання часу. Якщо існує дуга від Z_{t-1}^i до Z_t^i , то ця вершина називається сталою [19]. Дуги в межах одного шару є спрямованими, оскільки динамічна мережа є спрямованим ациклічним графом. В межах одного часового шару, як виняток, дозволяється використання неспрямованих дуг, що позначають сильні кореляції між змінними або деякі обмеження.

Для формалізації процесу адаптації моделі формулюється теорема про адаптацію динамічної байєсівської мережі (ДБМ), яка полягає у тому, що нова мережа має кращу якість за рахунок додавання нових і пропущених значень, а також модифікації структури і параметрів нової мережі [19]. Як критерій якості вибрано байєсівський інформаційний критерій (критерій Байєса-Шварца).

Теорема 1. Якщо адаптована динамічна байєсівська мережа має вищу очікувану адекватність за критерієм Байєса-Шварца (BSC), ніж очікувана (у сенсі математичного сподівання) адекватність попередньої мережі

$$\begin{aligned} & BSC((B'_0, B'_\rightarrow): D^+) - BSC((B_0, B_\rightarrow): D) > \\ & E[BSC((B'_0, B'_\rightarrow): D^+): D, (B_0, B_\rightarrow)] - \\ & E[BSC((B_0, B_\rightarrow): D^+): D, (B_0, B_\rightarrow)] \end{aligned}$$

то фактична якість нової мережі також буде вищою, ніж попередньої моделі (мережі) за рахунок заповнення пропусків, модифікації її структури і

параметрів. Крім того, різниця між очікуваними оцінками є нижнім граничним значенням покращення фактичних оцінок.

Тут: $E[BSC((B'_0, B'_{\rightarrow}): D^+): D, (B_0, B_{\rightarrow})]$ – очікувана оцінка ДБМ (B'_0, B'_{\rightarrow}) за критерієм Байєса-Шварца на основі всіх можливих доповнень даних D^+ , тобто заповнення пропущених значень і, можливо, використання значень прихованих змінних. Очікування береться по відношенню до $\Pr(D^+ | D, (B_{(0)}, B_{\rightarrow}))$, тобто ймовірність віднесена до доповнення даних старої ДБМ. Формально БМ для змінних \mathbf{X} можна визначити парою $B = (G, \Theta)$ – спрямований граф і параметри. ДБМ, що визначається парою (B_0, B_{\rightarrow}) , відповідає напів-нескінченній мережі для змінних $\mathbf{X}[0], \dots, \mathbf{X}[\infty]$ [60]. Але на практиці розглядається скінченний часовий інтервал $0, \dots, T$; $B' = (B'_{(0)}, B'_{\rightarrow})$ – це модифікована (або адаптована) байєсівська мережа.

Оскільки на попередньому кроці запропонованого методу було отримано авторегресійну модель (або авторегресію з ковзним середнім), то перший (статичний) шар динамічної мережі Байєса (рис. 3.7) буде складатись зі змінних, що входять в рівняння авторегресії [19]:

$$Z(t) = a_0 + a_1 U_1(t-1) + a_2 U_2(t-2) + \dots + a_p U_p(t-p) + \beta_1 \varepsilon_1(t-1) + \beta_2 \varepsilon_2(t-2) + \dots + \beta_q \varepsilon_q(t-q),$$

де p – порядок авторегресії, отриманої на другому кроці, а q – порядок ковзного середнього [26]. Якщо порядок ковзного середнього становить 1, то модель вироджується у звичайну авторегресію порядку p .

На виході будемо отримувати ймовірність набуття вихідними змінними значення у часовому просторі [26]. Для оцінювання фінансових ризиків це може бути ймовірність настання певного ризику або ймовірність фінансових втрат в часі (наприклад, через 30 днів). Мережа Байєса також може бути використана для короткострокового прогнозування [17] ймовірності втрат у певному діапазоні та певному інтервалі значень на момент часу t .

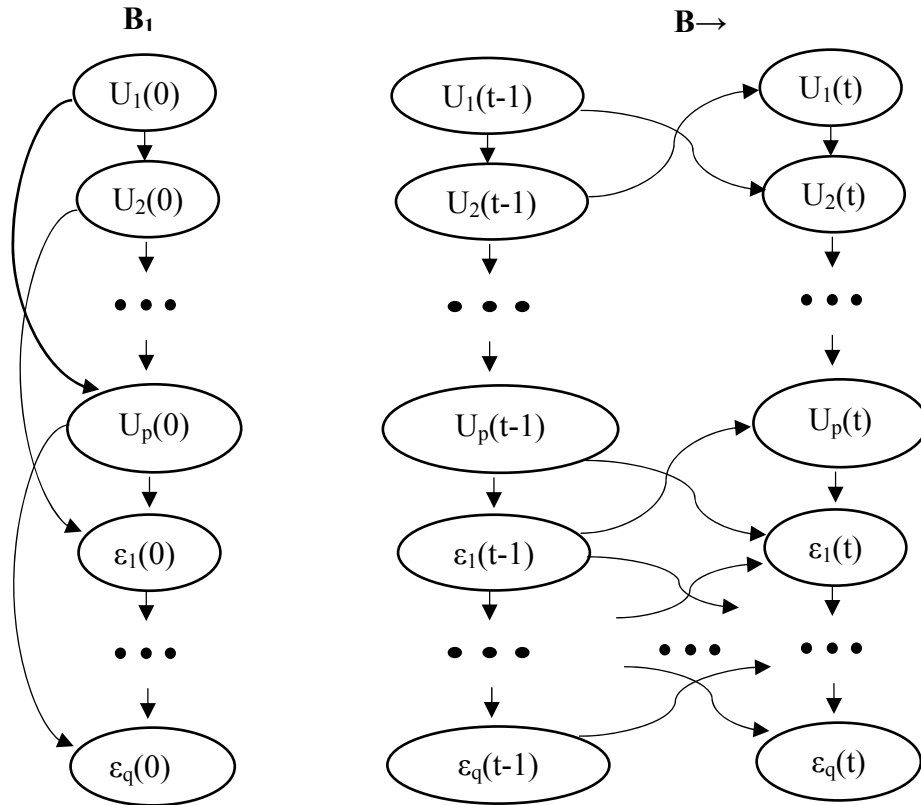


Рис. 3.7 Структура динамічної мережі Байєса

Метод може бути узагальнений на різні типи можливих розподілів. Основною перевагою цього методу є можливість одночасного оцінювання як ймовірності ризику (завдяки використанню мереж Байєса) [13], так і обсягів можливих втрат (на основі регресійних моделей). Окремим випадком є можливість його застосування під час статичного оцінювання ризиків, застосував замість динамічної звичайну дискретну або гібридну мережу Байєса [19], і прогнозування втрат та ймовірності настання ризику в статистиці.

3.7.1. Приклад оцінювання ринкового ризику на основі ймовірнісно-статистичного методу

Розглянемо задачу оцінювання ринкового ризику, який виникає при виконанні операцій з фінансовими активами. Вхідними даними комбінованої моделі є ціни акцій різних компаній, зокрема, Microsoft, Google і Cisco.

Часові ряди даних, які характеризують ціни акцій при закритті торгів на фондовій біржі, взяті з історичної бази даних фінансової системи YahooFinance.

Послідовність виконання обчислювальних операцій стосовно обробки даних, побудови математичних моделей, прогнозування волатильності (умовного стандартного відхилення) та оцінювання можливого ризику втрат відбувається за ієрархічною схемою зверху-вниз [19]. На першому кроці здійснювалось моделювання з використанням фільтра Калмана (ФК) та без нього. Далі на другому кроці здійснювалось короткострокове прогнозування волатильності прибутку за акціями Microsoft різними видами регресійних моделей, описаних у розділі 1 та додатку А.2.

В таблиці 3.4 подані статистичні показники точності короткострокового прогнозування волатильності на навчальній та перевірочній вибірках для досліджених моделей АРУГ, УАРУГ, ЕУАРУГ і МСВ [224, 259, 268, 328, 355]. Значення середньої абсолютної похибки (САП) і середньої абсолютної похибки у відсотках (САПП), отримані на навчальній вибірці для моделей УАРУГ НВ, ЕУАРУГ НВ, МСВ НВ [26], показують менші значення похибок прогнозування, ніж оцінки прогнозів моделей на перевірочній вибірці (УАРУГ ПВ, ЕУАРУГ ПВ та МСВ ПВ), що і очікувалось. Зазначимо, що модель АРУГ демонструє велику міру неточності оцінок прогнозів як на навчальній вибірці, так і поза її межами, що пояснюється простотою її структури [19], яка не відображає реальної взаємодії змінних.

Таблиця 3.4

Прогнозування волатильності прибутку акцій Microsoft

Модель	САП без ФК	САПП без ФК	САПП з ФК
АРУГ НВ	0,000359	9454,4	9188,7
УАРУГ НВ	0,0000791	45,960	36,270
ЕУАРУГ НВ	0,4353	4,9400	3,7530
МСВ НВ	0,64	7,5000	4,9023
АРУГ ПВ	0,00041	5123,5	2494,4
УАРУГ ПВ	0,00013	51,993	28,396
ЕУАРУГ ПВ	0,5053	5,93	4,0710
МСВ ПВ	0,84	10,90	6,9590

На третьому кроці здійснювалось ймовірнісне оцінювання втрат. Для цього була побудована динамічна мережа Байєса (рис. 3.8), яка прогнозувала на наступні кроки ймовірність фінансових втрат [19].

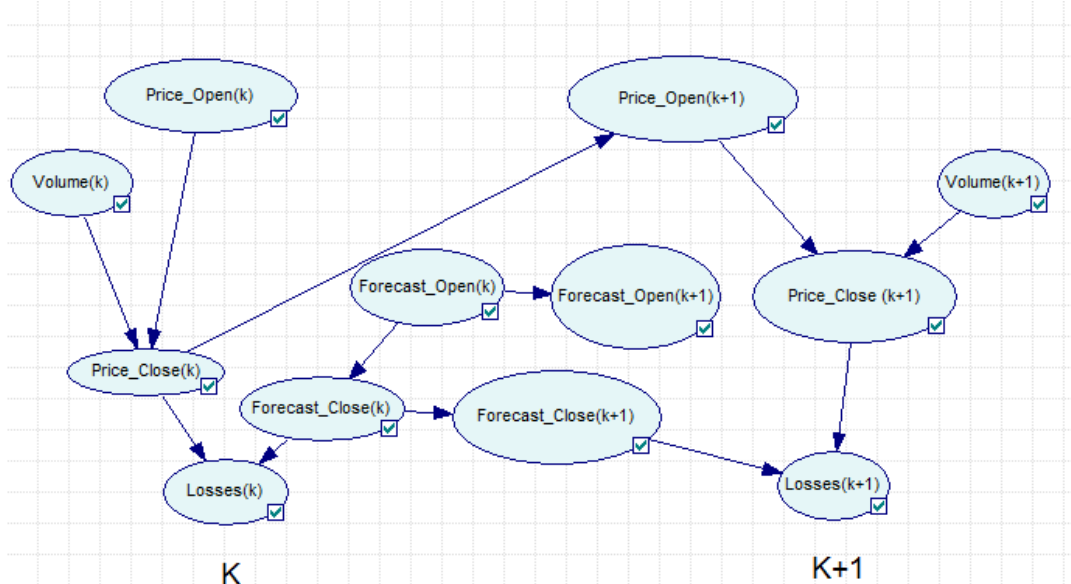


Рис. 3.8 Динамічна мережа Байєса для ілюстративного прикладу

Фінансові втрати за акціями в цілому за день оцінювались за формулою:
 $Losses(k) = (Forecast_Close(k) - Price(k)_Close) * Volume(k)$.

Застосування фільтра Калмана для попередньої фільтрації даних дало можливість зменшити похибки оцінок прогнозів волатильності (без моделі АРУГ) від 7,1% аж до 45,4%. Подібні ж результати моделювання і прогнозування волатильності отримані для цін акцій компаній Google і Cisco. Оцінки прогнозів волатильності використані для обчислення величини можливих втрат за методикою Value-at-Risk (VaR). Оскільки методика VaR дає значення можливих втрат у вибраному довірчому інтервалі [10], то уточнення оцінки ймовірності настання втрат виконано за допомогою мережі Байєса, яка побудована на таких змінних: значення оцінок можливих втрат, обчислені на попередніх трьох інтервалах (періодах дискретизації даних); прогнози значень волатильності і фактичні втрати на попередніх десяти інтервалах [19]. Ймовірності можливих втрат, отримані за допомогою байєсівської мережі, практично повністю співпали з фактичними ймовірностями втрат, обчисленими частотним методом на основі наявних історичних даних. Характерні значення

обчислених ймовірностей знаходяться у діапазоні: 0,01 – 0,13. Застосування запропонованої комплексної моделі дозволило підвищити якість оцінок прогнозів волатильності (на використаних статистичних даних) і уточнити ймовірність настання ризикових ситуацій.

Висновки до третього розділу

У розділі описано уперше запропоновану системну методологію менеджменту фінансових ризиків як узагальнення процедур статичного і динамічного оцінювання ризиків, застосування розроблених методів обробки пропущених та втрачених даних, адаптації моделей ризиків до нових даних і оцінювання адекватності їх структур.

Описано фундаментальні властивості і принципи системної методології, характеристики та обмеження вхідних даних, категорії та показники фінансових ризиків, критерії оцінювання якості даних, рішень, прогнозів, стратегій, що входять до єдиної системної методології менеджменту ризиків. Обґрунтовано і описано методи статичного та динамічного оцінювання ризиків, запропоновано комбінований метод обробки неповних даних та ймовірнісно-статистичний метод оцінювання ризику фінансових втрат.

Особливістю статичного оцінювання є розширення застосування скорингової карти до аналізу ризиків різної природи, розроблений нейронечіткий метод урахування відхилених раніше даних, застосування розробленого критерію урахування інформаційного ризику як прояву невизначеностей різних типів та зовнішніх впливів і інформаційних загроз.

Динамічне оцінювання ґрунтується на динамічному методі оцінювання ризиків з можливістю прогнозування рівня та ступеню ризику і ключових моментів часу (докладно описаний у розділі 4), та ймовірнісно-статистичному методі оцінювання ризику фінансових втрат. Метод передбачає комбінування різних підходів, використовує переваги ідеологічно різних методів та дозволяє оцінювати можливі втрати у формі точкових оцінок та ймовірностей настання

ризикових ситуацій. Запропоновано комбіновану модель на основі оптимального фільтра, регресійної моделі і байєсівської мережі для оцінювання ризику можливих фінансових втрат.

Методологія базується на принципах та процедурах системного аналізу, враховує основні принципи менеджменту ризиків та запропоновані принципи динамічного оцінювання і адаптивного менеджменту ризиків (докладно описаний у розділі 5). Системна методологія може бути застосована для прогнозування ризиків фінансового ринку, фінансової стабільності, кредитних та операційних ризиків.

Розроблена системна методологія і запропоновані нові методи практично реалізовані і впроваджені. Приклади реалізації у вигляді інформаційних технологій та ІСППР для компаній Artcom Venture GmbH, «САС Інстїт्यूт ЕЛ.ЕЛ.СІ.» описані у шостому розділі дисертаційної роботи, акти впровадження наведені у додатку Д.

РОЗДІЛ 4

РОЗРОБКА МЕТОДУ ДИНАМІЧНОГО ОЦІНЮВАННЯ РИЗИКІВ НА ОСНОВІ ТЕОРІЇ ВИЖИВАННЯ

У попередньому розділі обґрунтовано необхідність розробки системної методології дослідження ризиків фінансових систем, яка узагальнює і надає чітку послідовність застосування запропонованих методів, алгоритмів та моделей на основі визначених статистичних критеріїв і дозволяє оцінювати ризики за ступенем та рівнем, враховуючи невизначеності різного характеру і неповноту даних та моделюючи ризики за допомогою скорингової карти, яка дозволяє оцінювати ризики і швидко приймати рішення на робочому місці, згідно політики фінансової установи й для визначених типів ризиків.

Цей розділ присвячено розробці динамічної складової методології менеджменту ризиків, яка передбачає прогнозування ризиків (ступеню та рівню) з можливістю визначення завчасно моменту переходу ризику на більш високий ступінь або рівень. В основу запропонованого динамічного оцінювання покладено застосування теорії виживання і розроблено динамічні моделі різних типів: лінійні, пропорційні, непараметричні; та на їх основі виконано прогнозування динаміки зміни ризиків в часі. Також у розділі представлено розроблені методи визначення моменту переходу ризиків на більш високий ступінь та показано на конкретному прикладі можливість динамічного оцінювання ризиків у реальному часі.

4.1. Теоретичні і прикладні аспекти теорії виживання

Класична модель аналізу виживання використовується для дослідження моменту загибелі деякої популяції. Час, що проходить до настання цього моменту, називається часом виживання.

Аналізу виживання передували таблиці смертності, що використовувалися в страхуванні життя та демографічних науках XVII ст. На

початку свого існування метод таблиць смертності базувався на широких часових проміжках та великих обсягах даних. У 1958 році Едвард Каплан та Пауль Мейер опублікували роботу, в якій запропонували статистичну оцінку кривої виживання [285]. Вони розробили метод для коротких часових відрізків та менших вибірок, порівняно із тими, що використовувалася в демографічних дослідженнях.

Давид Кокс у роботі [251] запропонував метод, який дозволяв додавати коваріанти до аналізу подібних даних, він зараз відомий як метод «пропорційних ризиків Кокса» (Proportional hazards) (PH). Модель на його основі використовує регресори, що не залежать від часу, або статичні змінні та припускає, що співвідношення ризиків не змінюється із плином часу. Однак в реальних даних часто виникають саме характеристики, що змінюються із часом. Такі змінні порушують припущення про постійність відношення, тому модель Кокса [251] була модифікована і доповнена. На сьогодні відомі її стратифікована та узагальнена модифікації.

Застосування теорії аналізу виживання для моделювання кредитних ризиків було запропоновано нещодавно. Так, у роботах [264, 315] досліджуються переваги методів аналізу виживання перед загальноприйнятими статистичними методами. Методологія виживання використовує більшу кількість інформації, ніж звичайні моделі, оскільки вона дозволяє деталізувати поведінку шляхом цензурування та за рахунок використання змінної часу, що неможливо застосувати напряду ані в лінійній, ані в логістичній регресіях [349]. До того ж не потрібно ніяких припущень щодо розподілу змінної виходу. Саме такі особливості стали поштовхом для більш глибокого вивчення у дисертаційній роботі методів аналізу виживання та їх удосконалення для аналізу часу настання фінансових ризиків.

4.1.1. Припущення і означення теорії виживання

Для аналізу даних використовується вибірка (популяція), яка характеризується тим, що для кожного об'єкту відомий результат події (загибель чи виживання). Для цього здійснюється один з видів цензурування (відсікання). Спостереження називаються цензурованими, якщо спостережувана залежна змінна представляє момент настання термінальної події, а тривалість дослідження обмежена за часом [307]. Можливі механізми цензурування змінних: фіксоване цензурування (спостереження відбувається протягом фіксованого проміжку часу) та випадкове цензурування (спостереження відбувається протягом проміжку часу, який настає після того часу, коли елементи вибірки пережили певну подію) [239, 251].

Визначення. *Правими цензурованими спостереженнями* називають такі спостереження, які припиняються до настання події.

Спостереження називається *цензурованим зліва*, якщо воно спостерігалось до настання початку періоду спостереження. *Інтервально цензурованими спостереженнями* називаються, якщо час їх виживання розподілений між змінними a та b ($t \in [a, b]$) [197].

Типи правого цензурування:

- 1) Суб'єкти дослідження вижили до кінця дослідження. Час цензурування є фіксованим.
- 2) Суб'єкти дослідження вижили до кінця дослідження. Час цензурування настає, коли відбудеться попередньо визначена кількість подій [197].
- 3) Випадкові спостереження припиняються з причин, які не можуть бути контрольованими дослідником.

Функція виживання визначається як $S(t) = P(T > t)$, а функція ризику

$$h(t) = \lim_{\Delta \rightarrow 0} \frac{P(t < T < t + \Delta | T > t)}{\Delta}, \quad h(t) = -\frac{dS(t)}{S(t) dt}.$$

Найпростіша функція, яка визначає, що ризик є константою в часі: $h(t) = \lambda$, або що еквівалентно $\log h(t) = \mu$ [115].

Оскільки $S(t) = \exp\left[-\int_0^t h(u)du\right]$, то після підстановки та інтегрування отримуємо: $S(t) = e^{-\lambda t}$, а $f(t) = \lambda e^{-\lambda t}$. Це функція щільності ймовірності з відомим експоненційним розподілом з параметром λ [115]. Таким чином, сталий ризик передбачає експоненційний розподіл для часу, поки не наступить подія (або час між подіями) [261, 349].

У наведених позначеннях та припущеннях може бути виконана формалізація реальної практичної задачі оцінювання ризиків.

4.1.2. Формалізація задач динамічного прогнозування часу безпроблемності обслуговування банківського кредиту

Динамічне прогнозування передбачає поглиблене розв'язання задачі обслуговування кредиту: прогнозування не самого факту настання дефолту, а оцінювання часу (моменту) настання дефолту, враховуючи кредитну інформацію клієнтів та кредитних продуктів (ендогенні коваріанти) та індикатори економічного циклу (екзогенні коваріанти). Таким чином, ризик дефолту вимірюється за допомогою умовного розподілу випадкової змінної часу до дефолту T , заданого вектором коваріант X . Оскільки за припущеннями теорії виживання здійснюється обов'язкове цензурування змінних, то змінна T не є повністю спостережуваною.

Реально в практичних задачах частка прострочених кредитів є незначною порівняно з безпроблемними кредитами, отже це обмеження може бути доволі серйозним для коректного і ефективного використання відомих статистичних методів. Необхідно значно збільшувати розмір вибірки, а потім її стратифікувати для стандартних методів інтелектуального аналізу даних [85, 251, 326].

На рис. 4.1 проілюстровано застосування цензурування для вивчення кредитного ризику та розробки моделі ймовірності настання ризику. Можливі три найпоширеніші ситуації, коли кредитна компанія дотримується «строку» кредиту. Розглянемо інтервал $[0, \tau]$ як горизонт дослідження для ймовірності дефолту. Випадок (a) ілюструє обслуговування кредиту з дефолтом, що наступив до кінцевого часу дослідження (τ) [239]. В цьому випадку час життя кредиту, який є часом до дефолту кредиту, є спостережуваною змінною. Випадки (b) та (c) показують різні ситуації, коли не спостерігається момент часу переходу кредиту у дефолт через нестачу інформації, оскільки було здійснено праве цензурування по часу.

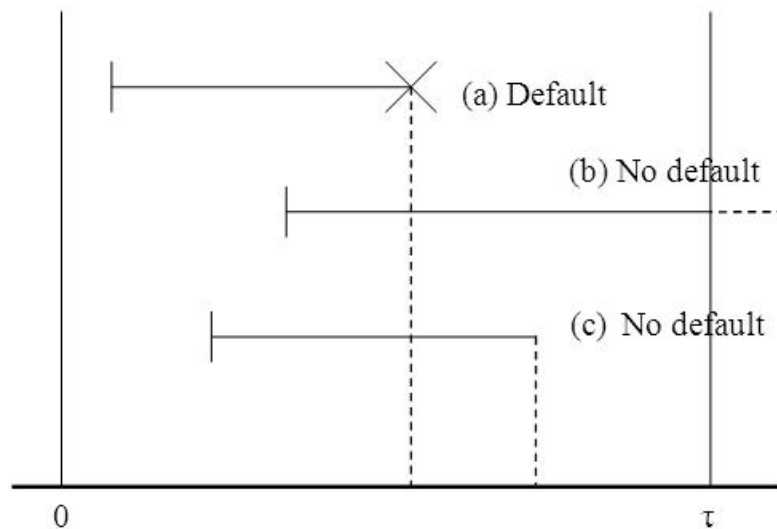


Рис. 4.1 Ілюстрація процесу цензурування для часу дефолту в споживчих кредитах

У випадку (b) це лише час від початку кредитування до кінця дослідження, в той час як (c) враховує ситуацію, коли очікується відмова від кредиту або закінчення строку обслуговування кредиту без настання дефолту [85, 239].

Доступною інформацією для моделювання ймовірності дефолту (PD) за кредитом є вибірка з n незалежних однаково розподілених випадкових величин $\{(Y_1, X_1, \delta_1), \dots, (Y_n, X_n, \delta_n)\}$, випадкового вектора $\{Y, X, \delta\}$, де $Y = \min\{T, C\}$ – спостережувана зрілість, T – час до настання дефолту, C – час до кінця дослідження або передбачуваного анулювання кредиту [129], $\delta = (T \leq c)$ –

індикатор не цензурування, X – вектор пояснюючих коваріант. У цих умовах аналізу виживання припускається, що існує певне невідоме співвідношення між T і X , а випадкові величини T і C є умовно незалежними для даного X [239].

У попередніх припущеннях можна повністю охарактеризувати умовний розподіл випадкової величини T , використовуючи деякі загальні співвідношення з теорії виживання. Функція умовного виживання, $S(t|x)$, умовний ступінь небезпеки, $\lambda(t|x)$, умовна кумулятивна функція ризику, $\Lambda(t|x)$, умовна функція розподілу $F(t|x)$ пов'язані наступним чином [115]:

$$S(t|x) = P(T > t | X = x) = \int_t^{\infty} f(u|x) du$$

$$\lambda(t|x) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t, X = x)}{\Delta t} = \frac{f(t|x)}{S(t|x)};$$

$$\Lambda(t|x) = \int_0^t \lambda(u|x) du = \int_0^t \frac{f(u|x)}{S(u|x)} du$$

$$S(t|x) = e^{-\Lambda(t|x)};$$

$$F(t|x) = 1 - S(t|x).$$

Відповідно до вимог Базеля II [304] моделі кредитного скорингу використовуються для вимірювання ймовірності дефолту в горизонті часу $t + b$ з часом зрілості t . Типове значення $b = 12$ (в місяцях). Таким чином, має бути обчислена наступна ймовірність [115]:

$$PR(t|x) = P(t \leq T < t + b | T \geq t, X = x) =$$

$$\frac{P(T < t + b | X = x) - P(T \leq t | X = x)}{P(T \geq t | X = x)} = , \quad (4.1)$$

$$= \frac{F(t + b|x) - F(t|x)}{1 - F(t|x)} = 1 - \frac{S(t + b|x)}{S(t|x)}$$

де t – спостережуваний термін погашення кредиту; x – значення коваріаційного вектору X для цього кредиту.

В контексті споживчих кредитів [306] популяція за теорією виживання складається із клієнтів-власників кредитних карт або інших позик, які діють за такими правилами [120]:

- коли користувач кредиту перестає виконувати свої зобов'язання (переходить у стан дефолту) щодо погашення заборгованості, то це вважається його загибеллю [185];
- час виживання вимірюється, починаючи із дати відкриття рахунку;
- якщо клієнт ніколи не виходить в дефолт упродовж періоду спостереження, то він підлягає цензурі в точці спостереження, тобто фіксується і не використовується при побудові моделі [129].

Для предметної області кредитних карток правила цензурування можна визначити наступним чином: кредитні карти, за якими було прострочено щонайменше 3 платежі на суму не менше 100 грн. вважаються дефолтними, тобто є «поганими» в контексті даного дослідження [129]. Усі інші цензуються, тобто вважаються такими, що не дійшли до свого логічного кінця.

Аналогічним чином для задач оцінювання часу настання ризику можна здійснити формалізацію, визначив ймовірність настання ризику як несприятливу подію, функцію виживання як функцію до настання ризику, час t як момент до визначеного часу дослідження або настання ризику T . У додатку В наведено приклади динамічного моделювання ризиків для різних фінансових систем, де було здійснено попередню постановку задачі в основних формалізмах та припущеннях з теорії виживання.

4.2. Динамічні моделі прогнозування ризиків

4.2.1. Модель пропорційних ризиків Кокса

Відома модель Кокса, запропонована в 1972 році [251], інтенсивно досліджується і використовується в різних областях діяльності, [48, 100, 109, 196, 197, 214, 239, 289, 324], особливо в медицині і страхуванні [48], для оцінки умовного ризику захворювання при заданих значеннях вихідних ознак, а робота Д.Кокса на сьогодні є найбільш цитованою публікацією по статистиці в історії.

Модель Кокса заснована на припущенні, що функцію ризику можна факторизувати, тобто представити у вигляді добутку двох функцій [197]:

$$h_i(t) = h_0(t) \cdot \psi(X_{i1}, \dots, X_{ik}),$$

де $h_0(t)$ – базова функція інтенсивності, що включає фактор часу, але не включає коваріанти, а $\psi(X_{i1}, \dots, X_{ik})$ – лінійна функція досліджуваних ознак, яка не включає фактор часу.

Досить часто модель записують у наступному вигляді [197]:

$$h_i(t) = h_0(t) \cdot e^{\{\beta_1 x_{i1} + \dots + \beta_k x_{ik}\}},$$

$$\ln h_i(t) = \ln h_0(t) + \beta_1 x_{i1} + \dots + \beta_k x_{ik},$$

де β_1, \dots, β_k – невідомі параметри.

Модель пропорційних ризиків Кокса у вигляді функції умовного виживання $S(t|x)$ передбачає оцінку сукупної умовної функції ризику, $L(t|x)$, з використанням максимальної правдоподібності. Метою є розробка умовної моделі для індивідуального $S(t|x)$, яка визначена в термінах $L(t|x)$. Для того, щоб описати PD^{PHM} (ймовірність дефолту при моделі пропорційних ризиків) ми визначимо наступні вирази по відношенню до теорії регресії Кокса [251].

Оцінка функції умовної швидкості небезпеки визначається як:

$$\hat{\lambda}(t|x) = \hat{\lambda}_0(t) \exp(x^T \hat{\beta}),$$

де $\hat{\lambda}_0(t)$ є оцінкою базової функції швидкості небезпеки $\lambda_0(t)$, а $\hat{\beta}$ – це оцінка вектора параметрів β [115].

Таким чином, при припущенні про модель пропорційних ризиків, PR оцінюється як [307]:

$$PR^{\hat{R}^{PHM}}(t|x) = \frac{\hat{F}_{\hat{\beta}}(t+b|x) - \hat{F}_{\hat{\beta}}(t|x)}{1 - \hat{F}_{\hat{\beta}}(t|x)} = 1 - \frac{\hat{S}_{\hat{\beta}}(t+b|x)}{\hat{S}_{\hat{\beta}}(t|x)},$$

де $1 - \hat{F}_{\hat{\beta}}(t|x) = \hat{S}_{\hat{\beta}}(t|x) = \exp(-\hat{\Lambda}(t|x))$.

Метод оцінювання для цієї моделі складається з двох кроків [307]. На першому етапі інтегральна функція базового ризику $\Lambda_0(t)$, оцінюється наступним чином [129]:

$$\hat{\Lambda}_0(t) = \sum_{i=1}^n \frac{I\{Y_i \leq t, \delta_i = 1\}}{\sum_{j=1}^n I\{Y_j \geq Y_i\}},$$

Тоді параметр β оцінюється як:

$$\hat{\beta}^{PHM} = \arg \max_{\beta} L(\beta),$$

а часткова функція правдоподібності задається таким виразом:

$$L(\beta) = \prod_{i=1}^n \frac{\exp(x_i^T \beta)}{\left(\sum_{j=1}^n I_{\{Y_j > Y_i\}} \exp(x_j^T \beta) \right)}.$$

Таким чином, оцінка умовної інтегральної функції ризику обчислюється за формулою [239]:

$$\hat{\Lambda}(t | x) = \int_0^t \hat{\lambda}(s | t) ds = \exp(x^T \hat{\beta}^{PHM}) \hat{\Lambda}_0(t).$$

Для оцінки вектору параметрів β достатньо розглядати функцію часткової правдоподібності як імовірність того, що відбулася загибель певного конкретного індивіда i , за умови, що відомо про його загибель [129], наприклад:

- t_i – час спостереження, тобто загибель або цензура;
- c_i – індикатор дефолту, $c_i = 1 \Leftrightarrow t_i$ – час загибелі.

Ймовірність того, що спостереження провалюється в певний час t серед інших спостережень обраховується наступним чином [239]:

$$\frac{h(t, x_i(t), \beta)}{\sum_{j \in R(t_i)} h(t_j, x_j(t), \beta)} = \frac{e^{\beta^T x_i(t)}}{\sum_{j \in R(t)} e^{\beta^T x_j(t)}}, \quad (4.2)$$

де $R(t) = \{j | t_j \geq t\}$.

Позначимо $\theta_j(t) := e^{\beta^T x_j(t)}$, $\theta_j^i := \theta_j(t_i)$. Тоді часткова функція правдоподібності та лог-перетворення від неї матимуть наступний вигляд:

$$L_p(\beta) = \prod_{i=1}^n \left(\frac{\theta_i^i}{\sum_{j \in R(t_i)} \theta_j^i} \right)^{c_i} \quad (4.3)$$

$$l_p(\beta) = \log L_p(\beta) = \sum_{i:c_i=1} \left(\beta^T x_i(t_i) - \log \sum_{j \in R(t_i)} \theta_j^i \right)$$

Звідси можна отримати вирази для градієнта та матриці Гессе, що застосовуються при знаходженні оптимального значення [239].

$$l'_p(\beta) = \sum_{i:c_i=1} \left(x_i(t_i) - \frac{\sum_{j \in R(t_i)} \theta_j^i x_j(t_i)}{\sum_{j \in R(t_i)} \theta_j^i} \right),$$

$$l''_p(\beta) = - \sum_{i:c_i=1} \left(\frac{\sum_{j \in R(t_i)} \theta_j^i x_j(t_i) x_j^T(t_i)}{\sum_{j \in R(t_i)} \theta_j^i} - \frac{\left(\sum_{j \in R(t_i)} \theta_j^i x_j(t_i) \right) \left(\sum_{j \in R(t_i)} \theta_j^i x_j^T(t_i) \right)}{\left(\sum_{j \in R(t_i)} \theta_j^i \right)^2} \right)$$

За наявності таких характеристик проводиться максимізація функції l_p для знаходження оцінки параметра β^* аналогічно формулі для логістичної регресії:

$$\beta^* = \operatorname{argmax}_{\beta} L(\beta) = \operatorname{argmax}_{\beta} l(\beta).$$

Одночасність настання подій

Концепція часткової максимальної правдоподібності при оцінюванні параметрів регресії припускає, що немає подій, які відбуваються одночасно. Однак в певних вибірках даних, таких як дані кредитної заборгованості, коли деталізація процесу невелика (місяць, квартал), просто неможливо уникнути одночасності. Тому розроблені відповідні методи, що дають можливість обходити подібні обмеження [197]:

- Точний метод. Перебирає всі можливі впорядкування одночасних подій, що вимагає інтенсивного використання обчислювальних ресурсів при великих масивах даних.
- Дискретний метод. Припускає, що події дійсно відбулися в один і той же час, замінює модель пропорційних ризиків Кокса на логістичну і обраховує імовірність того, що події відбулися до вже певної множини одночасних подій. Метод є дуже ресурсозатратним [197].
- Метод Бреслоу: апроксимує точний метод.
- Метод Ефрона: також апроксимує точний метод.

Усі вищезазначені методи є еквівалентними у випадку відсутності одночасних подій. Однак коли розмір вибірки є великим, як і кількість одночасних подій, то вважається, що метод Ефрона є більш придатним [307].

Нехай $\tau_1 < \tau_2 < \dots < \tau_k$ – упорядкована послідовність унікальних моментів загибелі. Позначимо $D_j := \{i \mid t_i = \tau_j, c_i = l\}$, $d_j := |D_j|$. При застосуванні методу Бреслоу функція правдоподібності залишається тією ж (4.3), і у введених позначеннях набуває такого вигляду:

$$L(\beta) = \prod_{i=1}^k \left(\frac{\prod_{j \in D_i} \theta_j(\tau_i)}{\sum_{j \in R(t_i)} \theta_j(\tau_i)^{d_i}} \right). \quad (4.4)$$

Використовуючи процедуру методу Ефрона [307], часткова функція правдоподібності обраховуватиметься за такою формулою:

$$L(\beta) = \prod_{j=1}^k \left(\frac{\prod_{i \in D_j} \theta_i(\tau_j)}{\prod_{l=0}^{d_j-1} \left(\sum_{i \in R(t_j)} \theta_i(\tau_j) - \frac{l}{d_j} \sum_{i \in D_j} \theta_i(\tau_j) \right)} \right) \quad (4.5)$$

Емпірична функція виживання

Усі спостереження вважаються завершеними і відсортованими за довжиною часових відрізків, починаючи з найменшого. Оцінка функції виживання обраховується за наступною формулою:

$$S_n(t) = \frac{\sum_{i=1}^n c_i(t)}{n}, \quad (4.6)$$

де $c_i(t) = \begin{cases} 1, & t_i > t, \\ 0, & t_i \leq t; \end{cases}$ n – кількість спостережень.

Крива функцій виживання $S_n(t)$ є ступінчастою функцією, кожна точка розриву якої відповідає смерті одного чи декількох суб'єктів. Оцінка математичного сподівання у такому випадку обраховується за формулою (4.7):

$$\hat{p} = \int_0^{\infty} S_n(t) dt. \quad (4.7)$$

Моделювання ризиків на основі припущення щодо їх пропорційності наразі є найбільш поширеним у кредитній галузі. У дисертації проводилось моделювання інших видів залежностей, зокрема представлення ймовірності настання ризику у вигляді узагальненої лінійної моделі або параметричної моделі, які будуть описані у наступних підрозділах.

4.2.2. Узагальнена лінійна модель

Для визначення «часу життя» до настання ризику можна сформулювати таку узагальнену лінійну модель:

$$P(T \leq t | X = x) = F_{\theta}(t | x) = g(\theta_0 + \theta_1 t + \theta^T x), \quad (4.8)$$

де $\theta = (\theta_2, \theta_3, \dots, \theta_{p+1})^T$, p – вимірний вектор, g – відома функція зв'язку, така як логістична чи пробіт-функція. Таким чином, ця модель характеризує умовний розподіл часу життя до настання ризику T [239] в термінах невідомих параметрів. Після оцінювання параметрів буде отримана оцінка функції умовного розподілу, $F_{\hat{\theta}}$ і, нарешті, оцінка ймовірності настання ризику P може бути обчислена шляхом включення цієї оцінки у рівняння (4.1), тобто

$$P\hat{R}^{GLM}(t | x) = \frac{F_{\hat{\theta}}(t + b | x) - F_{\hat{\theta}}(t | x)}{1 - F_{\hat{\theta}}(t | x)} = 1 - \frac{S_{\hat{\theta}}(t + b | x)}{S_{\hat{\theta}}(t | x)}, \quad (4.9)$$

де $\hat{\theta} = \hat{\theta}^{GML}$ є оцінкою максимальної правдоподібності вектору параметрів.

Розглянемо одновимірний випадок коваріантності. У такому випадку $\theta = \theta_2$ і умовний розподіл задається моделлю $F(t|x) = g(\theta_0 + \theta_1 t + \theta_2 x)$, зі щільністю $f(t|x) = \theta_1 g'(\theta_0 + \theta_1 t + \theta_2 x)$. Оскільки зазвичай задана випадкова цензурована справа вибірка, то умовна функція правдоподібності [239] представляє собою добуток членів, що включають умовну щільність, для нецензурованих даних та умовної функції виживання для цензурованих даних:

$$L(Y, X, \theta) = \prod_{i=1}^n f(Y_i | X_i)^{\delta_i} (1 - F(Y_i | X_i))^{1-\delta_i}, \quad (4.10)$$

де Y_i – строк обслуговування i -го клієнта до настання ризику, а δ^i є індикатором настання ризику для i -го клієнта.

Таким чином, логарифмічна функція правдоподібності визначається:

$$\begin{aligned} l(\theta) &= \ln(L(Y, X, \theta)) = \sum_{i=1}^n [\delta_i \ln(f(Y_i | X_i)) + (1 - \delta_i) \ln(1 - F(Y_i | X_i))] = \\ &= \sum_{i=1}^n [\delta_i \ln(\theta_1 g'(\theta_0 + \theta_1 Y_i + \theta_2 X_i)) + (1 - \delta_i) \ln(1 - g(\theta_0 + \theta_1 Y_i + \theta_2 X_i))] = \\ &= \sum_{i=1}^n \delta_i [\ln(\theta_1) + \ln(g'(\theta_0 + \theta_1 Y_i + \theta_2 X_i))] + \sum_{i=1}^n (1 - \delta_i) \ln(1 - g(\theta_0 + \theta_1 Y_i + \theta_2 X_i)) \end{aligned}$$

І, нарешті, оцінка знаходиться як максимізація функції логарифмічної правдоподібності:

$$\hat{\theta}^{GML} = \arg \max_{\theta} l(\theta).$$

Таким чином, використання узагальненої лінійної регресії можливе для оцінювання і прогнозування часу настання ризику при вирішенні задачі максимізації функції логарифмічної правдоподібності.

4.2.3. Розвиток моделі пропорційних ризиків

За припущенням модель Кокса [251, 252] передбачала, що всі фактори в моделі мають один і той самий час виміру, тобто не потребують урахування лагових (попередніх) значень. Крім цього передбачається лінійна залежність моделі ризику від інших факторів та адитивність функції ризику, тобто

спільний ефект від усіх предикторів є сумою ефектів окремих предикторів. В загальному випадку це не так, тому у дисертаційній роботі було розроблено моделі, що враховують й інші можливості, які будуть детальніше наведені нижче.

Модель Вейбулла

Модель є незначною модифікацією експоненційної моделі. Для неї зберігається припущення, що ε має стандартний розподіл екстремальних значень, проте послаблюємо припущення що $\sigma = 1$. Для $\sigma > 1$ ризик зменшується в часі, для $0,5 < \sigma < 1$, ризик зростає в часі зі швидкістю, що зменшується. Для $\sigma = 0,5$ функція ризику зростає лінійно, починаючи з 0. Модель називається моделлю Вейбулла (Weibull), тому що час T розподілений за розподілом Вейбулла, умовним для коваріант. Модель є досить популярною у біостатистиці через те, що функція виживання є доволі простою з математичної точки зору:

$$S_i(t) = \exp\{-[t_i e^{-\beta x_i}]^{\frac{1}{\sigma}}\},$$

де x_i – вектор значень коваріант і β – вектор коефіцієнтів.

Крім цього модель є також моделлю пропорційних ризиків. Це означає, що її коефіцієнти можуть бути представлені у вигляді відповідного співвідношення.

Існує співвідношення між лог-формою моделі функції виживання

$$\log h(t) = \alpha \log t + \beta_0^* + \beta_1^* x_1 + \dots + \beta_k^* x_k$$

і лог-моделлю для часу

$$\log T_i = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \sigma \varepsilon.$$

Для моделі Вейбулла відношення між параметрами порівняно з експоненційною моделлю визначається наступним чином:

$$\beta_j^* = \frac{-\beta_j}{\sigma} \quad \text{для } j = 1, \dots, k \text{ і } \alpha = (1/\sigma) - 1.$$

Оскільки $\beta_j = 0$ тільки тоді, коли $\beta_j^* = 0$ і тест на нуль-гіпотезу, що коефіцієнт дорівнює 0, буде таким самим, незважаючи, яка форма використовується. З іншого боку, стандартна похибка і довірчі інтервали для коефіцієнтів у логарифмічній моделі часу виживання не так легко представляються у логарифмічному форматі ризику.

Функція ризику з часовою залежністю предикторів.

$$h(t) = h_0(t) e^{\{\beta_1 X_1 + \beta_2 (X_1 \cdot f(t))\}}, \quad (4.11)$$

де $\beta_1 X_1$ – частина регресорів, що не залежить від часу,

$\beta_2 (X_1 \cdot f(t))$ – частина регресорів, залежних від часу [197].

$$\ln h_i(t) = \alpha(t) + \beta_1 x_{i1} + \dots + \beta_2 x_{i2}(t) \quad (4.12)$$

Моделювання різними функціями ризику для різних страт.

В моделі Кокса припускалось, що всі індивіди розподілені за однаковим законом, який і потрібно знайти. Проте в реальних практичних задачах оцінювання ризиків виникає необхідність моделювання різних груп клієнтів або класів, побудови для них різних моделей та скорингових карт в залежності від віку, доходу, типу клієнту, типу позики і т.д. Для того щоб врахувати різницю у поведінці таких груп (страт) і було запропоновано використання стратифікації та моделювання їх різними функціями [197]:

$$h_{ig}(t) = h_{0g}(t) e^{\{\beta_1 x_{i1} + \dots + \beta_k x_{ik}\}}, \quad (4.13)$$

причому $h_{0g}(t)$ буде різною базовою функцією для кожної страти, а параметри x_{i1}, \dots, x_{ik} будуть по-різному обраховуватись для кожної страти. Таким чином будуть формуватися різні функції виживання після реалізації ризику для кожної групи або страти.

Тоді

$$h_g(t) = h_{0g}(t) e^{\{\beta_1 X_1 + \beta_2 (g X_1)\}} \text{ або } h_g(t) = h_{0g}(t) e^{\{\beta_{g1} X_1\}}, \quad (4.14)$$

$$L = \prod_{i=1}^g L_i,$$

де L_i – часткова функція правдоподібності страти i , а g – номер страти.

Для оцінювання параметрів моделі методом часткової правдоподібності необхідно виконати наступні кроки [197]:

1. Побудувати різні функції виживання для кожної групи або страти.
2. Знайти добуток цих функцій.
3. Обрати значення параметрів β , які максимізують функцію часткової правдоподібності.

4.2.4. Непараметрична модель оцінювання фінансових ризиків

Непараметрична модель, описана у [239], передбачає застосування запропонованого Бераном [215] оцінювача для умовної функції виживання:

$$\hat{S}_h(t|x) = \prod_{i=1}^n \left(I - \frac{I_{\{Y_i \leq t, \delta_i = 1\}} B_{ni}(x)}{I - \sum_{j=1}^n I_{\{Y_j < Y_i\}} B_{nj}(x)} \right),$$

де Y_i – спостережуваний час для i -го клієнту, δ_i – це індикатор, що показує настання ризику для i -го клієнту і X_i – вектор пояснюючих коваріант для i -го клієнту. Терми B_{ni} – це непараметричні ваги Надарая-Вотсона [197]:

$$B_{ni}(x) = \frac{K((x - X_i)/h)}{\sum_{j=1}^n K((x - X_j)/h)}, 1 \leq i \leq n,$$

і $h \equiv h_n$ є згладжуючим параметром, який наближається до нуля, якщо розмір популяції наближається до нескінченності.

Для оцінювання ймовірності настання ризику в момент t для заданого набору параметрів виконують підстановку, тоді теоретичне значення умовної функції виживання [197] за його оцінкою \hat{S}_h :

$$P(\hat{R}^{NPM}(t|x)) = \frac{\hat{F}_h(t+b|x) - \hat{F}_h(t|x)}{1 - \hat{F}_h(t|x)} = I - \frac{\hat{S}_h(t+b|x)}{\hat{S}_h(t|x)}.$$

4.3. Оцінки для порівняння динамічних моделей

Оцінка Каплан-Майєра.

Оцінка Каплан-Майєра (КМ) є узагальненням емпіричної функції виживання і враховує відцензуровані спостереження. Формула Каплан-Майєра [285] для ймовірності виживання у певний час t_j обмежується добутком характеристик, що відповідають особам, що залишилися живими після настання часу t_j . Тому часто таку оцінку також називають Product-limit estimator [253].

Для обрахунку оцінки КМ усі спостереження сортують у порядку зростання часу їхнього життя. Перше входження починається з нуля. Ймовірність виживання для цього часу дорівнює 1. Подальші спостереження виключаються в момент часу їхньої загибелі (можливо, у результаті цензурування) [285]. Множина ризику, що позначається $R(t_j)$, – це кількість усіх індивідів, що дожили хоча б до часу t_j .

Основна ідея оцінки – представлення її у вигляді:

$$\hat{S}(t_i) = \prod_{j=1}^i \hat{P}(T > t_j | T \geq t_j). \quad (4.15)$$

Враховуючи те, що

$$\hat{P}(T > t_j | T \geq t_j) = \frac{n_j - d_j}{n_j}. \quad (4.16)$$

Тоді оцінка Каплан-Майєра:

$$\hat{S}(t) = \prod_{t_j < t} \left(1 - \frac{d_j}{n_j}\right), \quad (4.17)$$

де d_j – кількість зразків, що вибули під час дослідження (загибель в час t_j ($d_j = |D_j|$)), n_j – кількість зразків, що знаходяться під загрозою вибування (number at risk).

За обчисленою оцінкою Каплан-Майєра можливе порівняння функцій виживання для різних груп, що моделюються при оцінюванні фінансових

ризиків. Тобто є реальний інструмент порівняння і визначення часу життя та часу переходу до більш високого ступеня ризику для різних страт, клієнтів, груп.

Оцінка LogRank.

$$\text{LogRank} = \frac{(\sum_{j=1}^r (d_{1j} - e_{1j}))^2}{\text{var}(\sum_{j=1}^r (d_{1j} - e_{1j}))}, \quad (4.18)$$

де e_{1j} – кількість випадків 1-го виду на j -му часовому інтервалі, r – кількість часових проміжків [197].

Цей критерій подібний до χ^2 і використовується, коли функції ризику однакові, а рівень ризику є константою в часі.

Оцінка Wilcoxon.

Оцінка Вілкоксона (Wilcoxon) обчислюється за виразом:

$$\text{Wilcoxon} = \frac{(\sum_{j=1}^r n_j (d_{1j} - e_{1j}))^2}{\text{var}(\sum_{j=1}^r n_j (d_{1j} - e_{1j}))}. \quad (4.19)$$

Критерій порівняння функцій виживання, який є чутливим до зразків, що цензуються. Його рекомендовано [115, 197] використовувати на ранніх періодах дослідження.

4.4. Розробка алгоритмів прогнозування часу настання ризику

Візуальне порівняння кривих виживання для різних страт або груп, яке часто використовується, при прогнозуванні часу настання ризику є можливим, але не дуже зручним. Для задач, де зрізи даних надходять у часовому проміжку щоденно, погодинно, щохвилино, і точність прогнозування часу є критичною, необхідно розробити алгоритми встановлення часу.

Для цього можна використати кілька різних підходів. Якщо функція ризику визначається у процесі моделювання через параметричний,

непараметричний розподіл, то можливе обчислення часу через похідну функції ризику.

Алгоритм 1 розрахунку моменту переходу на вищий ступінь ризику.

Для визначення моменту часу t виконаємо такі кроки:

1. Задамо вид початкової функції $\hat{\Lambda}_0(t)$. Нехай $\hat{\Lambda}_0(t) = \exp(-a \cdot t)$, де $a > 0$, але має невелике значення, щоб $\hat{\Lambda}_0(t) = \exp(-a \cdot t)$ не швидко спадала.
2. Підставивши базову функцію небезпеки для пропорційних ризиків, отримуємо: $\hat{\Lambda}(t) = \exp(x^T \cdot \beta^{PHM}) \cdot \exp(-a \cdot t)$.
3. Взявши похідну від функції небезпеки по часу, отримаємо:

$$\frac{\partial \Lambda(t | x)}{\partial t} = \exp(x^T \cdot \beta^{PHM}) \cdot \exp(-a \cdot t) \cdot (-a) = -a \cdot \exp(x^T \cdot \beta^{PHM} - a \cdot t).$$

Для можливості диференціювання функції необхідно існування її похідної, а це не завжди можна гарантувати. Тому можливо адаптувати алгоритм без прямого обрахунку похідної. Виходячи з визначення похідної як швидкості зміни певної функції, можна обчислити швидкість зміни ймовірності, тобто переходу її до критичної ймовірності настання ризику $P_{крит}(t)$:

1. Задається ймовірність, яка є критичною, $P_{крит}$.
2. Розраховується значення $P_{поточн}(t) - P_{крит} = \Delta P$ як «запас по ймовірності».
3. Визначається момент переходу ризику до критичного як:

$$t_{крит} = \frac{\Delta P}{\frac{\partial P(t)}{\partial t}}.$$

Якщо встановити ймовірність ризику, яка є критичною, є неможливим, тоді пропонується розробити алгоритм розрахунку часу через критичний (або катастрофічний) рівень ризику, тобто критичні (або катастрофічні) втрати. Це завжди є можливим, оскільки фінансова система або підприємство функціонує

для отримання певного прибутку, а тому може визначити, які втрати за ризиками є більшими, ніж отриманий прибуток.

Алгоритм 2 визначення моменту настання критичного (катастрофічного) рівня ризику за втратами:

1. Задати інтервал часу, на якому буде здійснюватись пошук критичного часу $\Delta T = (0, T)$.
2. Задати крок збільшення часу $t := \Delta t$.
3. $t := 0$.
4. Обчислити $Losses_{ноточ}(t)$.
5. Якщо $Losses_{ноточ}(t) \geq Losses_{крит}$, то $t_{крит} := t$ і STOP.
6. $t := t + \Delta t$.
7. Якщо $t \geq \Delta T$, то STOP і в цьому інтервалі не відбудеться перехід ризику до критичного (катастрофічного).
8. Go to Step 4.

За визначеним на попередньому кроці t можна обчислити: $\lambda(t|x)$, $P(t|x)$ та очікувані втрати EL для: допустимого, критичного та катастрофічного рівня ризику в моменти часу (t_1, t_2, t_3) .

Розроблені алгоритми дозволяють визначити не лише ступінь та рівень ризику, як передбачалось у статичному оцінюванні, описаному у третьому розділі, а і спрогнозувати момент часу, коли рівень або ступінь ризику різко змінюється.

4.5. Метод динамічного оцінювання ризиків

Для моделювання ризиків в динаміці було запропоновано метод динамічного оцінювання та прогнозування ризиків, який передбачає побудову динамічних моделей різних типів для окремих страт, визначення кращої з них, та використання такої моделі для прогнозування часу, рівню та ступеню

ризиків. Метод може бути представлений як послідовність описаних нижче кроків.

Крок 1. Визначення значущості характеристик фінансових ризиків-параметрів моделей на основі критеріїв:

– оцінки кореляції $R^2, \chi^2, ACF, PACF$ [26].

– ваги категорії змінної: $WOE_i = \ln\left(\frac{g_i}{b_i}\right)$,

де $\sum_{i=1}^k g_i = 1$ – розподіл одиничних значень цільової змінної та $\sum_{i=1}^k b_i = 1$ – розподіл нульових значень цільової змінної [342].

– інформаційного значення (IV): $IV = \sum_{i=1}^k (g_i - b_i) \ln\left(\frac{g_i}{b_i}\right) = \sum_{i=1}^k (g_i - b_i) WOE_i$.

Крок 2. Розробка моделей виживання різного виду

За можливістю може бути висунуте припущення щодо розподілу ризиків (пропорційні ризики, з часовими коваріантами, непараметричні моделі тощо). Тоді буде будуватись тільки один вид моделей виживання. Якщо явно визначити вид моделі неможливо, тоді будується множина обраних видів моделей і визначається краща з них.

2.1. Модель пропорційних ризиків

2.1.1. Оцінка функції умовного рівня небезпеки визначається як:

$$\hat{\lambda}(t | x) = \hat{\lambda}_0(t) \exp(x^T \hat{\beta}),$$

де $\hat{\lambda}_0(t)$ – оцінка базової функції рівня небезпеки $\lambda_0(t)$ [129], а $\hat{\beta}$ – оцінка вектора параметрів β , а ймовірність настання ризику:

$$P\hat{R}^{PHM}(t | x) = \frac{\hat{F}_{\hat{\beta}}(t + b | x) - \hat{F}_{\hat{\beta}}(t | x)}{1 - \hat{F}_{\hat{\beta}}(t | x)} = 1 - \frac{\hat{S}_{\hat{\beta}}(t + b | x)}{\hat{S}_{\hat{\beta}}(t | x)},$$

де $1 - \hat{F}_{\hat{\beta}}(t | x) = \hat{S}_{\hat{\beta}}(t | x) = \exp(-\hat{\Lambda}(t | x))$.

2.1.2. Оцінюється інтегральна функція базового ризику $\Lambda_0(t)$, так:

$$\hat{\Lambda}_0(t) = \frac{\sum_{i=1}^n I\{Y_i \leq t, \delta_i = 1\}}{\sum_{j=1}^n I\{Y_j \geq Y_i\}},$$

2.1.3. Параметр β оцінюється як $\hat{\beta}^{PHM} = \arg \max_{\beta} L(\beta)$, де часткова функція правдоподібності задається виразом [129]:

$$L(\beta) = \prod_{i=1}^n \frac{\exp(x_i^T \beta)}{\left(\sum_{j=1}^n I_{\{Y_j > Y_i\}} \exp(x_j^T \beta) \right)}.$$

2.1.4. Оцінка умовної інтегральної функції ризику визначається за формулою [239]:

$$\hat{\Lambda}(t | x) = \int_0^t \hat{\lambda}(s | t) ds = \exp(x^T \hat{\beta}^{PHM}) \hat{\Lambda}_0(t).$$

2.2. Побудова узагальненої лінійної моделі для оцінювання ризиків

Припущення щодо пропорційності ризиків не завжди є допустимим, а тому можливе представлення функції ризику та функцій виживання через узагальнену лінійну регресію (4.8) – (4.10). Оцінка параметрів визначається в результаті максимізації логарифмічної функції правдоподібності:

$$\hat{\theta}^{GML} = \arg \max_{\theta} l(\theta).$$

2.3. Побудова непараметричної моделі оцінювання фінансових ризиків

$$\hat{S}_h(t | x) = \prod_{i=1}^n \left(I - \frac{I_{\{Y_i \leq t, \delta_i = 1\}} B_{ni}(x)}{I - \sum_{j=1}^n I_{\{Y_j < Y_i\}} B_{nj}(x)} \right),$$

$$\text{де } B_{ni}(x) = \frac{K((x - X_i)/h)}{\sum_{j=1}^n K((x - X_j)/h)}, 1 \leq i \leq n,$$

$$P(\hat{R}^{NPM}(t | x)) = \frac{\hat{F}_h(t + b | x) - \hat{F}_h(t | x)}{I - \hat{F}_h(t | x)} = I - \frac{\hat{S}_h(t + b | x)}{\hat{S}_h(t | x)}.$$

Крок 3. Порівняння моделей та обрання кращої за множиною статистичних критеріїв

3.1. Обрання кращої моделі на основі критеріїв, описаних у розділі 3: GINI, AUC, неправильної класифікації (MR), Байєса-Шварца (BSC), Колмогорова-Смірнова (KS): $AUC = \int_{\tilde{y} \in \tilde{Y}} Se(\tilde{y}) dFPR(\tilde{y})$, $GINI = 2 \cdot AUC - 1$,

$$MR = \frac{FP + FN}{N}, BSC = N \ln \left(\sum_{k=1}^N e^2(k) \right) + n \ln(N), KS = \max_{x \in X} |F_B(x) - F_G(x)|, \text{ тощо.}$$

3.2. Перевірка гіпотези про однаковий розподіл функцій ризику

Застосовуються критерії, описані у розділі 4.3 оцінки Каплан-Майєра, LogRank, Wilcoxon (4.17–4.19) для порівняння функцій виживання, що були побудовані для різних страт.

Крок 4. Визначення моменту настання ризику

В залежності від постановки задачі, тобто необхідності визначення моменту переходу ризику на більш високий ступінь або рівень, використовується один з алгоритмів, описаних у підрозділі 4.4.

Крок 5. Визначення ймовірності настання фінансового ризику та можливих втрат

За визначеним на попередньому кроці t можна обчислити: $\lambda(t|x)$, $P(t|x)$ та очікувані втрати EL для: допустимого, критичного та катастрофічного рівня ризику в моменти часу (t_1, t_2, t_3) та відповідних функцій виживання.

4.6. Приклади динамічного моделювання ризиків фінансових систем

4.6.1. Динамічне моделювання кредитних ризиків

Задача моделювання кредитних ризиків вирішувалась у декілька етапів. На першому етапі будувалась модель, що визначала поведінку кредитної карти клієнта різними методами інтелектуального аналізу даних. Обиралась краща з них та використовувалась для подальшої оцінки. Це задача так званого класичного скорингу (оцінювання) для визначення: а) подальших дій клієнтів (дефолту/недефолту); б) визначення оптимальних лімітів та розрахунок можливих втрат внаслідок реалізації ризику невиконання взятих фінансових

зобов'язань клієнтом за заданою кредитною картою. Від задачі персоналізованого оцінювання кредитних карт також був здійснений перехід до оцінювання кредитних ризиків в цілому по портфелю і надання рекомендацій щодо подальшої кредитної політики. Друга задача, що моделювалась, це безпосередньо задача динамічного оцінювання ризиків, тобто визначення моментів, коли з'являються проблеми з обслуговуванням кредитних карт. Тут також можливе вирішення задачі виявлення шахрайства як відхилення від класичної поведінки клієнта і відповідно запобігання шахрайству з кредитними картами шляхом блокування таких операцій у реальному часі.

Збирались статистичні дані, що описували самого клієнта та його кредитні карти. Детально особливості вирішення даної задачі були описані у статтях автора [129, 185, 296, 299]. Модель кредитної карти містила наступні дані, що проілюстровано у таблиці 4.1.

Таблиця 4.1

Модель кредитної карти

Слоти	Опис
beginDate	Дата початку
finshDate	Дата номінального закінчення
closeDate	Дата фактичного закінчення
gracePeriod	Грейс-період
setLimit()	Встановлення ліміту
getLimit()	Перегляд поточного ліміту
Balance	Баланс
postDuePayments	Кількість прострочених платежів
setLoan()	Надання кредиту
getOverdue()	Отримання величини прострочки
getOutstanding()	Отримання величини термінової заборгованості
getPayment()	Отримання щомісячного мінімального платежу

В основу моніторингу кредитної карти та відслідковування шахрайських дій може бути покладена середньостатистична модель поведінки клієнта [347], що характеризується рядом параметрів, відхилення від яких система онлайн-моніторингу може сприймати як шахрайство і відмовляти у проведенні операції. Для забезпечення клієнта можливостями обслуговування кредитної

карти банк дозволяє клієнту коригувати власноруч модель кредитної карти, встановлюючи: кредитні ліміти, кількість операцій в день, можливість виконання операцій в Інтернеті, платежі за кордоном тощо.

Процедура прогнозування поведінки клієнта-власника кредитної карти є статистичним відображенням зміни кредитних лімітів, обсягу операцій та витрачених грошей у часі. Тоді шахрайством банк буде вважати дії з кредитною картою, що «випадають» з основної моделі поведінки клієнта, і це стане приводом для заблокування таких підозрілих операцій.

Проведення обчислюваних експериментів

Для експерименту було відібрано 55286 спостережень за 4037 кредитними картками зі щомісячною деталізацією. За весь період було зафіксовано 714 випадків дефолту. Для побудови моделей було сформовано список параметрів (табл. 4.2) [185].

Цільова змінна, тобто змінна, яка описувала факт дефолт/не дефолт, формувалася за критерієм [299]:

- прострочка більше 100 грн.;
- більше трьох послідовних прострочених платежів.

Таблиця 4.2

Список відібраних параметрів

Ідентифікатори	
dealb2id	Ідентифікатор угоди
Monnum	Порядковий номер місяця життя
Дані про поведінку	
Outbody	Залишок за тілом кредиту
Outextra	Залишок за комісією та/або відсотком
Ovdbody	Прострочена заборгованість за тілом
Ovdextra	Прострочена заборгованість за комісією + відсотками
pddbbody_new	Кількість днів прострочки за тілом
Pddextra	Кількість днів прострочки за комісією + відсотками
Ovd	Величина простроченої заборгованості
pdd_new	Кількість днів прострочки за кредитною картою
Limit	Розмір ліміту
Montake	Кількість знятих коштів
Monpay	Кількість повернутих коштів
Minpd	Номер місяця першого простроченого платежу понад 100 грн.

Аплікаційні дані	
clienttype	Тип клієнта
Age	Вік
beginyear	Рік видачі
term_month	Тривалість угоди
dealamount	Ліміт на початок угоди
requestamount	Сума, на яку розраховував клієнт
is_satisfied	Ознака того, що очікування клієнта справдилися
liv_is_reg	Ознака того, що адреса прописки співпадає з адресою місця постійного проживання
Childcnt	Кількість дітей
dependantcnt	Кількість утриманців
Income	Дохід
u_score	Загальний скоринговий бал
has_u_score	Признак того, що загальний скоринговий бал розраховувався
with_bank_mon	Вік клієнта як контрагента
reg_mon	Тривалість останньої прописки
Агреговані дані	
max_outbody	Максимальна заборгованість за тілом
max_ovdbody	Максимальна прострочена заборгованість за тілом
max_ovd	Максимальна прострочена заборгованість
max_pdd_new	Максимальна кількість прострочених місяців
max_limit	Максимальна величина ліміту
avg_montake	Середнє значення щомісячного зняття коштів
avg_monpay	Середнє значення щомісячного платежу
max_ovd_over_limit	Максимальне значення відношення простроченої заборгованості до встановленого ліміту
Цільова змінна	
Defaulted	Цільове поле (дефолт/не дефолт)

Будувались моделі лінійної та логістичної регресії, оцінювались їх параметри, вводились лаги для регресорів. Розроблялись моделі на основі інших методів інтелектуального аналізу даних для прогнозування факту дефолту за кредитною картою.

Реалізація здійснювалась на мові програмування R. Початкова модель M0 обирала змінні поведінки клієнтів без лагів [185]. Далі модель доповнювалась змінними з лагами (зміщеними у часі) на 1, 2, 3 лаги і, відповідно, були побудовані моделі M1, M2, M3, у яких замість поточних значень outbody, ovd, montake, monpay використовуються значення з певним лагом (табл. 4.3).

Таблиця 4.3

Значення коефіцієнтів моделей M1, M2, M3

Скорочена назва змінної	Оцінені коефіцієнти моделі M _i		
	M1	M2	M3
log(1 + outbody(t-i))	-0.1984	-0.1679	-0.0500
log(1 + ovd(t-i))	1.0041	0.9217	0.6563
log(1 + limit)	-0.1422	-0.1458	-0.2052
log(1 + montake(t-i))	-0.8695	-0.5845	-0.0513
log(1 + abs(monpay(t-i)))	-0.2420	-0.1002	-0.0302
clienttypeClient	0.2738	0.0777	0.3777
clienttypeEmployee	0.3499	0.0680	-0.4239
clienttypeInsider	0.9472	0.8103	0.6047
clienttypeNormal	0.2748	0.1149	0.3139
clienttypeOldSalary	0.6650	0.4982	0.5961
clienttypeSalary	0.6318	0.1192	-0.0524
Age	-0.0021	0.0018	0.0053
is_satisfiedt	-0.2529	-0.3165	-0.2657
liv_is_regt	0.1521	0.1887	0.1976
Childcnt	0.0275	-0.0015	0.0083
Dependantcnt	-0.0610	-0.0747	-0.0593
log(1 + income)	-0.0862	-0.0363	0.0012
has_u_scoret	0.1002	0.1747	0.4204
with_bank_mon	-0.0142	-0.0146	-0.0087
reg_mon	-0.0040	-0.0077	-0.0128

Для порівняння моделей використовувались критерії оцінки адекватності моделей, наведені у розділі 3. Результати моделювання показані у таблиці 4.4 та на рис. 4.2. Кращою моделлю за значенням площі під ROC-кривою виявилась модель з лагом 1, проте за інформаційним критерієм Акайке кращою моделлю виявилась логістична регресія.

Таблиця 4.4

Порівняння моделей

Моделі	M1	M2	M3	Logit
AUC	0.919	0.888	0.829	0,804
AIC	8946.005	7819.972	7495.905	2334.007

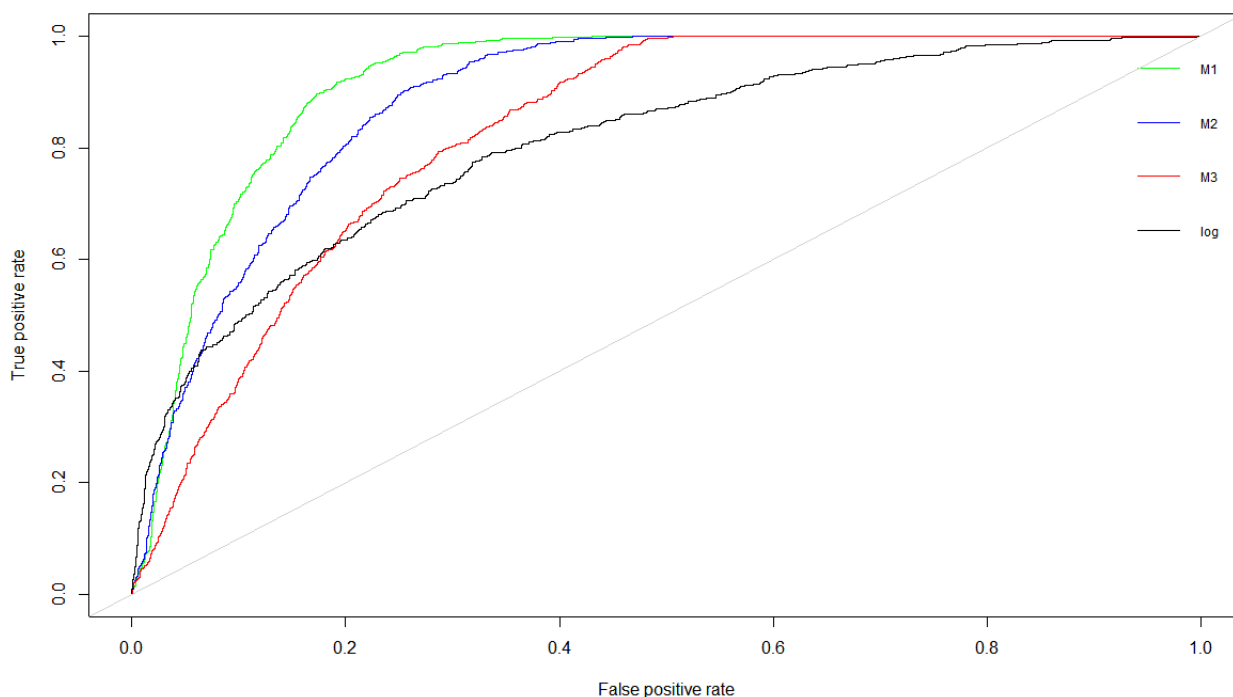


Рис. 4.2 Графіки ROC-кривих для моделей M1, M2, M3

Для подальшого порівняння з іншими методами інтелектуального аналізу даних були обрані апарати нейронних мереж, лінійної регресії та логістичної регресії. Для нейронних мереж був застосований алгоритм побудови багатошарового перцептрону з трьома прихованими шарами [370]. За різними статистичними критеріями було виконано порівняння якості моделей на навчальній та перевіірчій вибірці (таблиці 4.5 та 4.6).

Кращою моделлю за сукупністю статистичних критеріїв є нейронна мережа [129], однак результати навіть на навчальній вибірці показали необхідність покращення моделей.

Таблиця 4.5

Результати моделювання на навчальній вибірці

Модель	Misclassification rate	IKA	SBC	GINI	KS
Нейронна мережа	0.26	5772.87	5997.73	0.634	0.481
Логістична регресія	0.272	5922.28	5988.42	0.609	0.465
Лінійна регресія	0.273	9357.37	9291.24	0.609	0.462

Таблиця 4.6

Результати моделювання на перевіірчній вибірці

Модель	Misclassification rate	GINI	KS
Нейронна мережа	0.271	0.615	0.461
Логістична регресія	0.278	0.6	0.449
Лінійна регресія	0.279	0.6	0.448

Побудовані моделі [129] дозволяють визначити ймовірність переходу у стан дефолту на основі деякої інформації (параметрів) кредитної карти та, відповідно, історії платежів та обслуговування кредиту за попередні періоди [185]. Такий підхід є класичним аплікаційним скорингом, що використовується для прийняття рішень щодо видачі кредиту (кредитної карти). Для вирішення другої задачі моделювання, а саме – аналізу поведінки клієнтів з метою їх динамічного оцінювання, було застосовано динамічний метод оцінювання ризиків.

Побудова динамічних моделей оцінювання ризиків

Для прогнозування періоду проблемності кредиту були побудовані два типи моделей виживання: параметричні та напівпараметричні.

З початкової вибірки у 55286 спостережень, де було зафіксовано 714 випадків дефолту, у 3837 спостерігались прострочки. Було виконано цензурування 760 кредитних карток, що склало 16,53% із загального обсягу кредитних карт.

Далі будувались два типи динамічних моделей на основі теорії виживання: параметричні (з використанням процедури Lifereg [197]: експоненційним, логістичним та розподілом Вейбулла) та напівпараметрична модель (пропорційних ризиків Кокса). Серед параметричних моделей кращою виявилась модель з використанням розподілу Вейбулла.

Наступним етапом було визначення періоду, коли виникають проблеми з обслуговуванням кредиту. На рис. 4.3 показано моделі виживання для всієї сукупності клієнтів, і спостерігається ситуація, коли перші 3 місяці кредитні

картки добре обслуговуються [129, 299], проте з ростом періоду «життя» кредитної карти все частіше наступають випадки прострочки за кредитом.

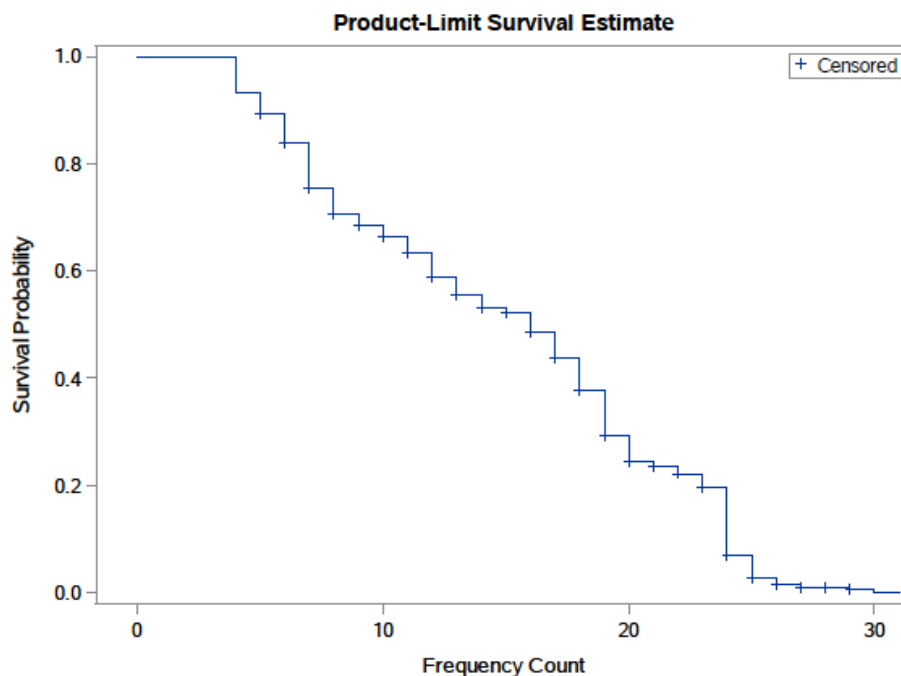


Рис. 4.3 Моделювання ризиків функціями виживання для кредитних карт

При досягненні періоду обслуговування кредитної карти у 12 місяців, ймовірність «виживання» кредитної карти становить більше 0,6, проте далі ймовірність виживання значно падає. Також спостерігається досить різке падіння проценту «ідеального обслуговування» кредиту через 18 місяців.

Для динамічного оцінювання ризиків застосовувалась функція ризику, представлена на рис. 4.4. Тут спостерігається аналогічна ситуація, коли до 18 місяців обслуговування кредиту рівень ризику є прийнятним (до 0,1), адже функції ризику і виживання пов'язані між собою. Невеликий сплеск значень функції ризику спостерігається через 8 місяців, який потім вирівнюється, проте з часом ризик значно зростає. Визначивши допустимий або прийнятний рівень ризику, можна встановити оптимальний період безпроблемного обслуговування кредиту. Для цього застосовують визначений банком ступінь ризику та

описаний у цьому розділі алгоритм розрахунку критичного часу настання ризику.

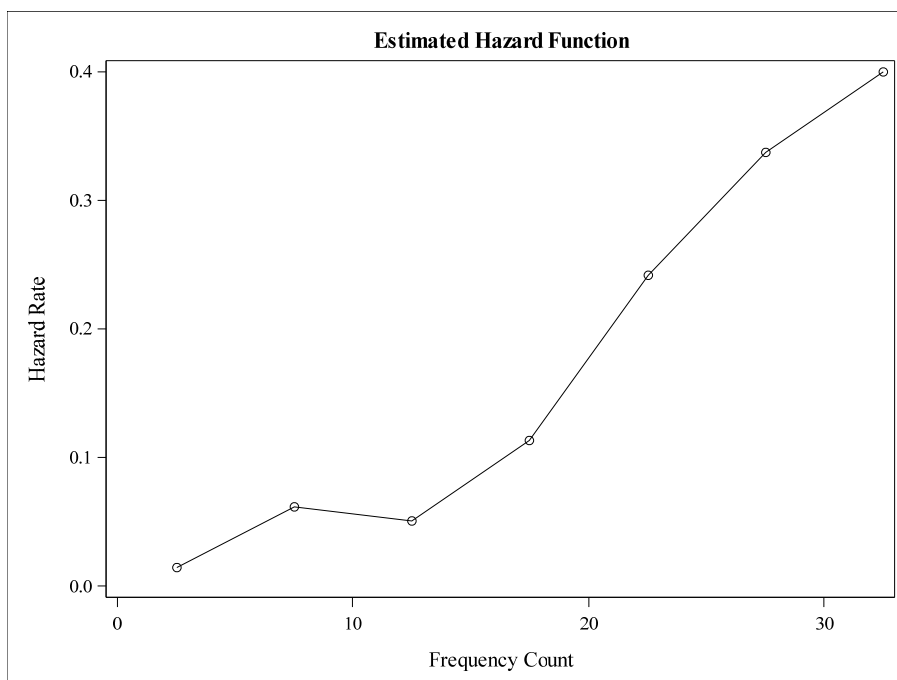


Рис. 4.4 Оцінена функція ризику в залежності від часу

Побудовані динамічні моделі (параметричні та напівпараметричні) дозволили визначити період, в який настає проблемність обслуговування кредитних заборгованостей за картками, а також визначити ймовірність настання такої події на певному етапі кредитування. Це дозволяє говорити про обсяги можливих втрат банків в результаті негативної динаміки по кредитним карткам, а також застосовувати реальні засоби боротьби з прострочками на більш ранніх етапах (ще до закінчення терміну кредиту). Моделі пропорційних ризиків Кокса, які показали найкращі результати для дослідженої задачі, можуть бути використані банками для розробки нових кредитних продуктів (в залежності від потреб споживачів на різні періоди обслуговування та з різними процентними ставками) та планування фіскальних дій в разі несвоєчасного обслуговування кредитів.

4.6.2. Приклад динамічного моделювання ризиків телекомунікаційної компанії

Для моделювання ризиків було використано реальні дані української телекомунікаційної компанії. Оцінювались потенційні втрати компанії через втрату клієнтів і прогнозувалась ймовірність відтоку клієнтів.

Опис вхідних даних. Статистична інформація про активність 150 тисяч абонентів в мережі протягом 15 місяців зібрана за 84 характеристиками. Було побудовано статистичні моделі за відібраними 10 основними показниками та 5-ма додатковими, що описували їх поведінку під час користування мобільними послугами. Детально постановка задачі, опис вхідних даних були викладені у роботі [121].

Спочатку здійснювалось статичне оцінювання фінансових ризиків з використанням методів ІАД для прогнозування ймовірності відтоку кожного абоненту та всіх абонентів компанії в цілому. Найкращі результати за характеристикою AUC показав градієнтний бустинг, а вищу точність моделі показав класифікатор випадковий ліс (табл. 4.7) [121].

Таблиця 4.7

Порівняння моделей на тестовій вибірці

Модель	precision	Recall	F1	F0.5	AUC
Градієнтний бустинг	0,70	0,64	0,66	0,68	0,842 ± 0,005
Випадковий ліс	0,72	0,60	0,65	0,69	0,832 ± 0,008
Нейронні мережі	0,69	0,59	0,63	0,66	0,825 ± 0,007
Логістична регресія	0,63	0,37	0,46	0,55	0,842 ± 0,002

На наступному етапі будувались різні типи моделей непараметричного умовного розподілу з використанням оцінок LogRank та Wilcoxon, отримані статистики, наведені на рис. 4.5. Подібне моделювання є доволі суб'єктивним і прогнозує ризик для окремих осіб, але не дає відповіді, як зменшити його, тобто який момент часу є найбільш критичним, які характеристики можуть впливати. Тому було продовжено дослідження з використанням запропонованого динамічного методу.

Rank Statistics		
SUBSCRIPTION_TYPE_CODE	Log-Rank	Wilcoxon
CRP	-128.7	-36172
f	-1829.5	-3.327E8
m	1958.2	3.3269E8

Covariance Matrix for the Log-Rank Statistics			
SUBSCRIPTION_TYPE_CODE	CRP	f	m
CRP	979.7	-622.9	-356.8
f	-622.9	14248.8	-13625.9
m	-356.8	-13625.9	13982.7

Covariance Matrix for the Wilcoxon Statistics			
SUBSCRIPTION_TYPE_CODE	CRP	f	m
CRP	1.116E13	-7.09E12	-4.07E12
f	-7.09E12	1.736E14	-1.66E14
m	-4.07E12	-1.66E14	1.706E14

Test of Equality over Strata			
Test	Chi-Square	DF	Pr >
Log-Rank	280.6203	2	<.0001
Wilcoxon	654.6026	2	<.0001
-2Log(LR)	242.8871	2	<.0001

Рис. 4.5 Результати моделювання ризиків з використанням оцінок LogRank та Wilcoxon

Оскільки дані у вибірці представлені за кожний місяць як статистика по абонентах, які обслуговуються, то необхідно було виконати групування даних, щоб отримати інформацію про конкретного абонента за весь період його обслуговування.

Побудовано моделі виживання для прогнозування ймовірності продовження обслуговування клієнтів в залежності від типу (корпоративний чи приватний клієнт). На рис. 4.6 можна відстежити потенційний відток клієнтів через 1-2 місяці [100, 121, 292], коли закінчуються наперед проплачені суми коштів, акційні пропозиції, або просто були тимчасові абоненти, які купували номери для користування послугами зв'язку в Україні, перебуваючи на відпочинку або у відрядженні.

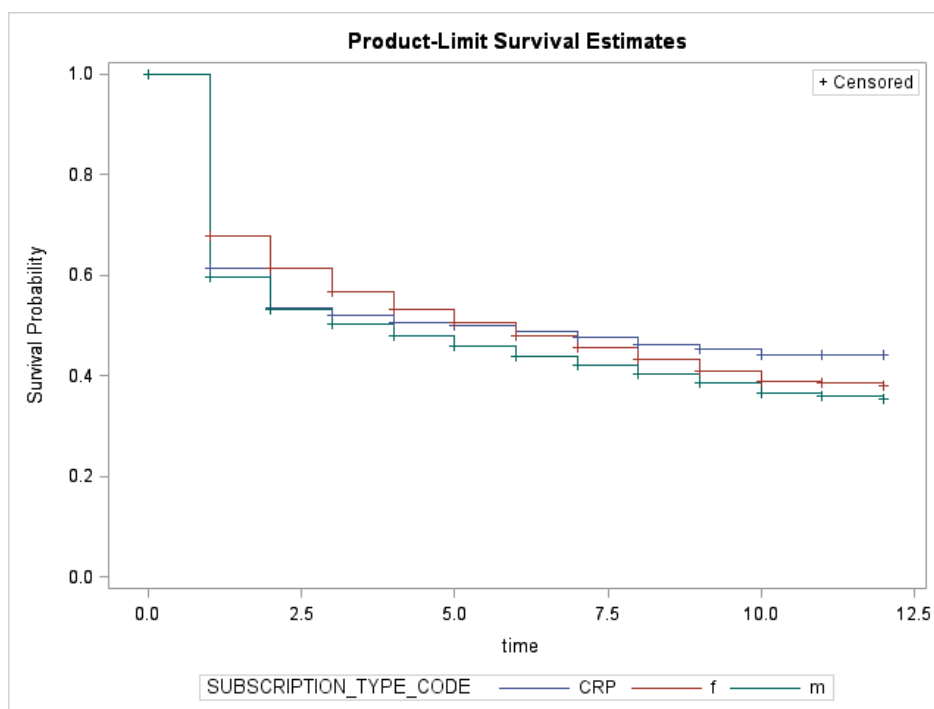


Рис. 4.6 Графіки функцій виживання для згрупованих даних

Важливим також було прогнозування моменту, коли ймовірність відтоку клієнтів перевищує 50%. Використовувався запропонований алгоритм розрахунку критичного моменту часу через заданий 50% ступінь ризику. Такий критичний момент для чоловіків настає через 3 місяці користування послугами, для корпоративних клієнтів – через 4 місяці і через 5 місяців – для жінок [121].

Отримані результати прогнозування часу є досить правдоподібними, оскільки відображають особливість людського мислення та дій: більш швидкого прийняття рішення чоловіками і більш постійних у своїх уподобаннях жінок. Для корпоративних клієнтів період відтоку більше пов'язаний з політикою обслуговування, встановленій у телекомунікаційній компанії [210], тобто списання абонентської плати за місяць наперед, і, відповідно, затримання відтоку клієнтів на місяць, пов'язане з необхідністю подачі та обробки заявки на відмову від користування послугами.

Для кожної групи клієнтів оцінювався рівень ризику протягом певного часу (рис. 4.7) та обсяг можливих втрат. Залежно від фінансової ситуації компанії може бути переоцінений сукупний рівень ризику через можливі

втрати. Слід оцінювати сукупний рівень втрат без деталізації по типах клієнтів і, відповідно, сукупних втратах за моделлю Кокса і сукупною функцією ризику, або окремо побудувати моделі збитковості по типах клієнтів [121].

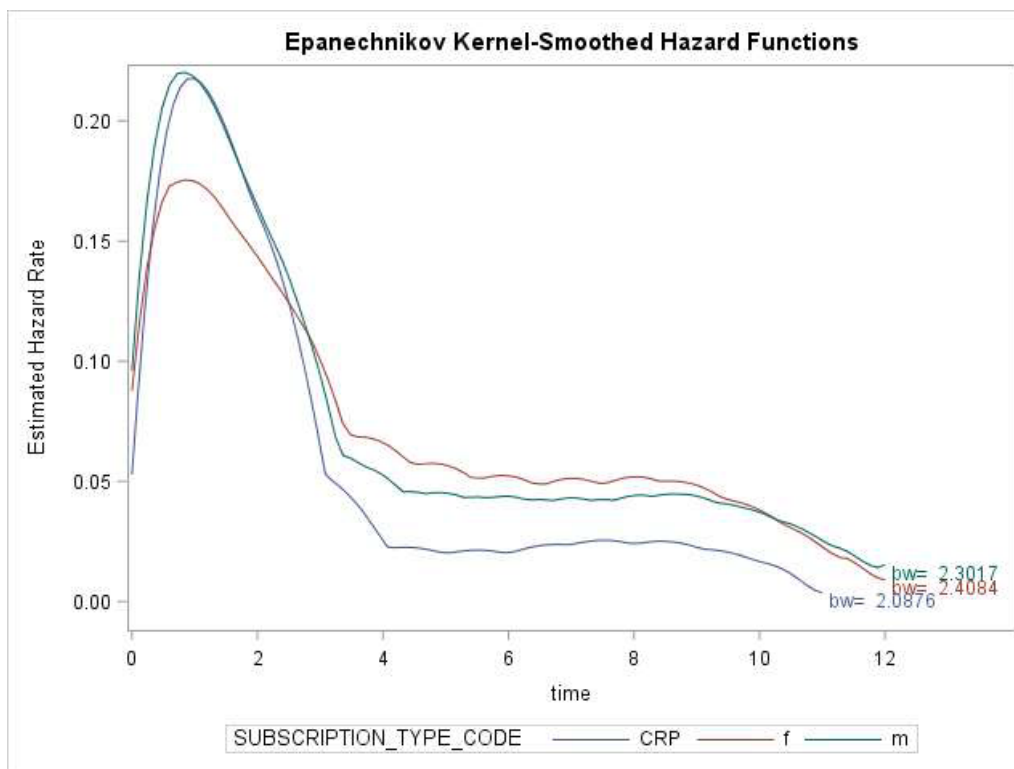


Рис. 4.7 Графік функції збитковості для згрупованих даних

Якщо на початковому етапі визначити, що телекомунікаційна компанія вважає допустимим рівнем ризику такий, рівень якого за функцією ризику не більше $\lambda(t_1) = 0,15$, критичний не більший $\lambda(t_2) = 0,3$ [121], а катастрофічний більший $\lambda(t_3) = 0,4$, то за описаним алгоритмом розрахунку часу можна визначити для кожної страти (групи) клієнтів, чи відбувався перехід в зону критичного чи катастрофічного ризику. Видно, що найбільший рівень втрат спостерігався серед груп чоловіків та корпоративних клієнтів через 1 місяць та 1,5 місяці.

Висновки до четвертого розділу

У розділі проаналізовано теоретичні і прикладні аспекти теорії виживання та запропоновано створення на її основі моделей для оцінювання

фінансових ризиків, зокрема прогнозування часу настання ризику та його переходу на більш високий рівень. Наведено формалізацію динамічних моделей на основі пропорційних ризиків Кокса та їх модифікацію, лінійних та непараметричних моделей. Запропоновано і розроблено метод динамічного оцінювання ризиків, який передбачає побудову різних типів моделей виживання і прогнозування часу переходу на вищий ступінь або рівень ризику. Запропоновано два алгоритми прогнозування часу: алгоритм, який дозволяє визначити момент переходу ризику від допустимого до критичного або катастрофічного (за визначеним обсягом критичних або катастрофічних втрат) та алгоритм розрахунку моменту переходу ризику на вищий ступінь за заданою для фінансової системи критичною ймовірністю ризику. Проведене експериментальне дослідження показало доцільність застосування динамічного методу оцінювання ризиків для практичних задач моделювання кредитних ризиків та ризиків телекомунікаційної компанії.

Динамічне оцінювання доцільно використовувати для банків та фінансових компаній, які працюють на ринку в умовах перехідного періоду і змушені швидко реагувати на зовнішні впливи та дії конкурентів. Також такий підхід може бути рекомендовано для використання у Національному банку України при проведенні стрес-тестування українських банків, для оцінювання їх діяльності і своєчасного виявлення та реагування на економічні проблеми всередині таких банків (відгук про застосування наведено у додатку Д).

РОЗДІЛ 5

РОЗВИТОК АДАПТИВНОГО ПІДХОДУ ДО ПОБУДОВИ МОДЕЛЕЙ РИЗИКІВ ФІНАНСОВИХ СИСТЕМ

Прийняття рішень в умовах невизначеностей різної природи є доволі складною задачею, адже частина інформації може бути не просто пропущеною, а навіть повністю протилежною до тої, на яку розраховувала особа, що приймає рішення. Основний ризик виявляється у прийнятті неправильного рішення щодо ситуації, яка може спричинити появу нового ризику. Фактично весь час відбувається пошук компромісу у прийнятті кращого рішення при умові неточної або неповної інформації, що обумовлює появу декількох видів ризиків одночасно, а отже може призвести до певних втрат. Фінансові системи за своєю суттю пов'язані з функціонуванням у взаємозв'язку з іншими системами, отже й прийняття рішень здійснюється в умовах, залежних від інших осіб, систем, країн, що діють відповідно до своїх задач та зовнішніх умов, приймаючи відповідні управлінські рішення так само в умовах невизначеності.

У ситуації динамічного розвитку і взаємозв'язку з іншими системами важливо не лише приймати рішення, а й мати можливість його уточнення та доопрацювання в разі різкої зміни умов, зовнішніх впливів, екстрених ситуацій або як реакції на дії конкурентів. У розділі буде запропонований метод, який базується на принципі адаптивного менеджменту, і дозволяє враховувати нові дані, критерії та навіть появу нових ризиків, уточнюючи існуючі моделі за рахунок адаптації структури і параметрів [131, 135, 152, 154], або обирати альтернативний метод і створювати нову модель, яка враховує як історичні дані, так і нові дані, пов'язані з критичними змінами зовнішніх умов. Це дозволить знизити ризики прийняття помилкових управлінських рішень щодо подальших керуючих дій та надасть ефективний інструмент для прийняття управлінських рішень в реальних системах при зміні зовнішніх умов.

5.1. Принцип адаптивного менеджменту ризиків фінансових систем

Ризики фінансових систем є по суті ризиками несистемними та в явному вигляді не визначеними, можуть бути як наслідком дії інших ризиків, так і їх комбінацією. Тому принцип їх оцінювання має дозволяти не лише враховувати зміну рівня та ступеню ризику системи, а й прогнозувати, моделювати та знижувати нові ризики, які можуть виявитись у процесі функціонування системи. Характерні особливості ризиків фінансових систем полягають у тому, що вони вимірюються у грошовому еквіваленті, пов'язані з фінансовими процесами, які можуть змінюватись від стаціонарних до нестаціонарних, від гомоскедастичних до гетероскедастичних, суттєво залежать від фінансових ринків, економічних складових, політичних рішень, та навіть не можуть бути чітко визначені і класифіковані по групах. З'являються нові валюти, нові грошові еквіваленти, платіжні системи та способи оплати [178], цифрові підписи, безготівкові та пейпал-розрахунки, що надають нові можливості для розвитку фінансових систем, а тому і нових можливостей для контрдій, махінацій, шахрайства, відповідно і ризиків нової природи та суті.

У дисертаційній роботі пропонується *принцип адаптивного менеджменту ризиків*, який полягає у можливості оцінювання та адаптації ризиків в процесі функціонування системи за рахунок використання адаптивних методів та моделей, уточнення кращої моделі, її параметрів та структури, застосування адаптивних оцінок прогнозів, використання множини комплексних критеріїв для оцінювання якості структури і параметрів моделей, введення нових критеріїв оцінки якості рішення та з передбаченою можливістю введення нових статистичних даних. Для реалізації цього принципу необхідно залучити сучасні методи інтелектуального аналізу даних, інформаційні технології з широкими обчислювальними можливостями та розробити новий метод структурно-параметричної адаптації.

5.2. Адаптивний підхід до менеджменту фінансових ризиків

Серед відомих засобів для адаптації прогнозної моделі використовують:

- рекурсивне оцінювання параметрів моделей після надходження нових вимірів для уточнення моделі та підвищення якості прогнозів практично одночасно з надходженням нової інформації;
- використання автоматизованих процедур розрахунку АКФ та ЧАКФ [17] для визначення кореляції між змінними чи виявлення необхідності урахування лагів (і якого порядку) для основних змінних;
- різні методи побудови регресійних моделей (зворотнього, почергового, методу провідних індикаторів) для виявлення найважливіших змінних для надання високоякісного прогнозу;
- оцінювання параметрів збурюючих впливів в реальному часі з надходженням нових вимірів [74];
- автоматизоване коригування структури динамічних, гібридних та статичних мереж Байеса, різних модифікацій нейронних мереж до нових вимірів, змінних та оцінок експертів;
- використання похибок прогнозів для обчислення вагових коефіцієнтів при формуванні комбінованої оцінки прогнозу з декількох індивідуальних оцінок;
- адаптивне налаштування вагових коефіцієнтів у процедурах експоненційного згладжування, тощо [217].

Для отримання додаткової інформації про процес застосовують множину критеріїв для перевірки його лінійності/нелінійності, стаціонарності, гетероскедастичності, здійснюють аналіз залишків, тощо. Це дозволяє скоригувати структуру моделі і адаптувати її до фінансового процесу, застосував контур адаптації за похибкою моделі, і уточнити її до необхідного рівня. Зворотний зв'язок за похибкою прогнозу також спрямовується на підвищення точності оцінок прогнозів, яка досягається за рахунок покращення якості (інформативності) даних, уточнення структури моделі, а також завдяки

застосуванню схем усереднення оцінок прогнозів. Цей контур також дає можливість уникнути «перенавчання», яке приводить до підвищення точності апроксимації даних побудованою моделлю, але знижує якість оцінок прогнозів.

Важливою можливістю для адаптації є застосування комбінованих та інтегрованих методів і моделей, які дозволяють подолати існуючі обмеження для використання того чи іншого методу і отримати вищі оцінки якості прогнозів при моделюванні фінансових ризиків. Наприклад, такі методи як логістична регресія або методи класифікації вимагають існування обмеженої кількості класів, які на початковому етапі можуть бути не задані для вхідних даних. Проте відомо, що ці методи надають найкращі можливості для вирішення задачі класифікації, тож є сенс використати комбінування декількох методів. Ще одним серйозним обмеженням для статистичних методів є необхідність збалансованої вибірки (в ідеалі 50/50), проте для задач аналізу ризиків це обмеження є критичним, оскільки обсяг навчальної вибірки для різних видів ризиків може бути суттєво меншим.

Коректне застосування сучасних моделей і адаптивних методів оцінки, імовірного і статистичного аналізу даних дозволяє отримати більш високу якість прогнозних оцінок в умовах структурних, параметричних і статистичних невизначеностей. Можливості для адаптації забезпечують методи на основі фільтру Калмана [74], які генерують оптимальні оцінки стану разом з короткостроковими прогнозами в умовах впливу зовнішніх стохастичних збурень і шумів вимірювань. Однак такі методи вимагають оцінок статистичних параметрів випадкових збурень і шумів вимірювань в реальному масштабі часу, що створює додаткове навантаження і помилки для процедур прогнозування.

Концепція адаптації моделі для динамічних процесів прогнозування у рамках запропонованої системної методології менеджменту ризиків передбачає ієрархічний підхід до процедур моделювання і прогнозування з урахуванням можливої структурної, параметричної та статистичної невизначеності, адаптації

математичних моделей до можливих змін у процесі навчання і при використанні альтернативних методів оцінки параметрів з метою моделювання і покращення прогнозних оцінок. Функціональна схема адаптивної системи прогнозування наведена на рис. 5.1.

Тут кожний етап обробки даних контролюється відповідним набором статистичних параметрів, які характеризують специфічні особливості даних, модель в цілому, її параметри, і, нарешті, якість отриманих прогнозних оцінок.

Адаптивна схема відрізняється кількома можливостями адаптації з використанням комплексного критерію якості. Висувається вимога попередньої підготовки зібраних даних для побудови структури моделі та оцінювання її параметрів. Визначення оптимальної структури моделі є ключовим елементом для досягнення необхідної якості прогнозів. Пропонується визначити структуру моделі наступним чином:

$$S = \{r, p, m, n, d, z, l\},$$

де r – кількість рівнянь (розмірність) моделі; p – порядок моделі (максимальний порядок диференціального або різницевого рівняння в авторегресійній моделі); m – число незалежних змінних у правій частині; n – наявність нелінійності і її тип; d – лаг (відставання) або затримка реакції в часі для виходу; z – зовнішнє збурення і його тип; l – можливі обмеження для змінних. Для автоматичного пошуку «кращої» моделі запропоновано такий комбінований критерій [217]:

$$V_N(\theta, D_N) = e^{|1-R^2|} + \ln\left(1 + \frac{MSE}{N}\right) + e^{|2-DW|} + \ln(1 + MSE) + \ln(MAPE) + e^U, \quad (5.1)$$

де θ – вектор параметрів моделі; N – розмірність часового ряду, що використовується; R^2 – коефіцієнт детермінації; DW – статистика Дарбіна-Уотсона; MSE – середньоквадратична помилка; $MAPE$ – середня абсолютна помилка в процентах; U – коефіцієнт Тейла. Ефективність запропонованого критерію була перевірена експериментально і доведена широким набором моделей і статистичних даних.

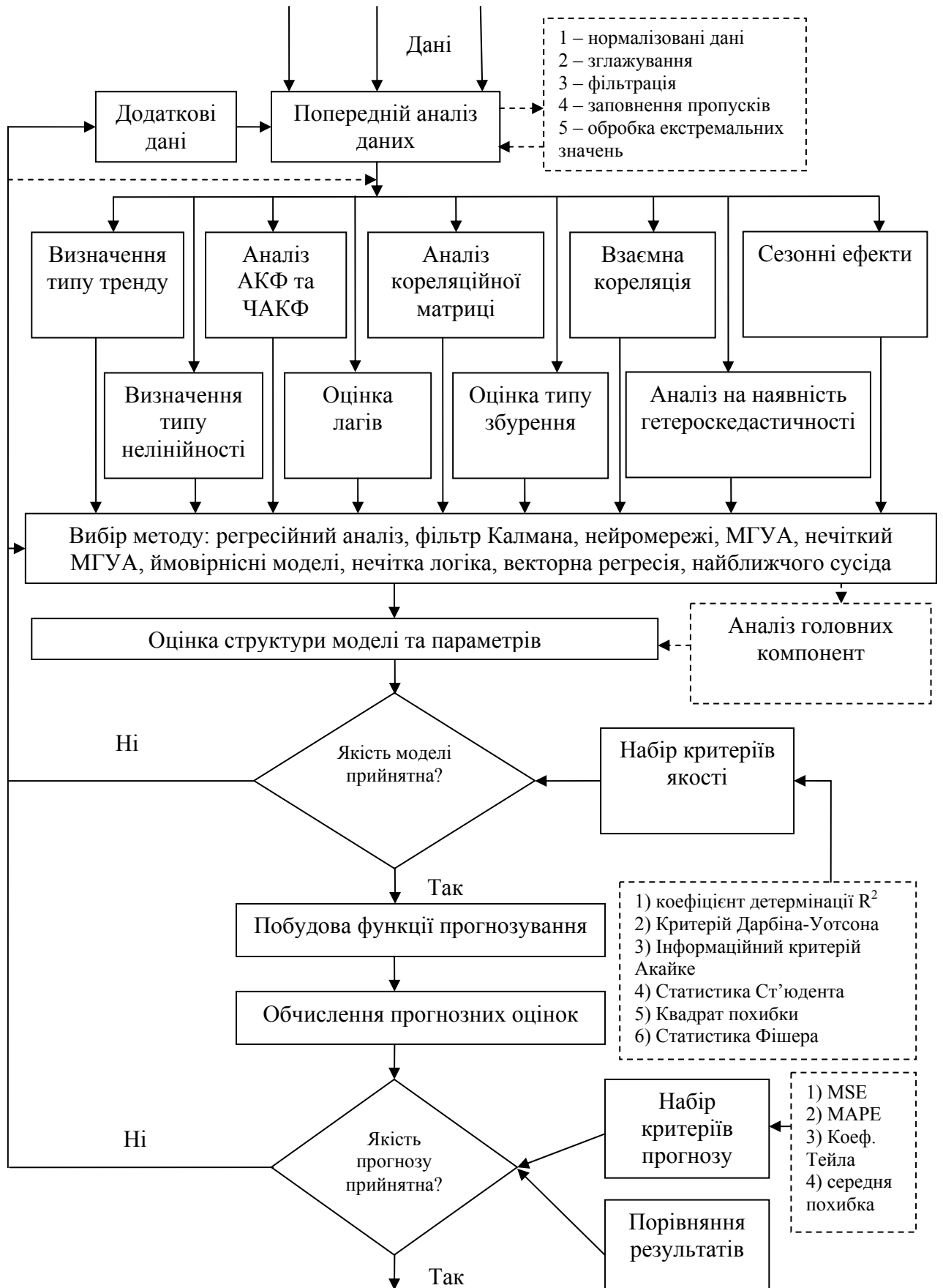


Рис. 5.1 Адаптивна схема процесу оцінювання та прогнозування моделі

Для оцінки адаптивної структури моделі використовувалась сукупність автоматизованих засобів для визначення АКФ, ЧАКФ, аналізу залишків і т.п. Використання конкретної схеми адаптації залежить від обсягу і якості даних, конкретної постановки задачі, вимог до прогнозних оцінок, а також вибором конкретного набору моделей для оцінювання і прогнозування. Логістична регресія з лінійною регресією були використані для прогнозування та опису даних, пов'язаних з фінансовими втратами, а дерева рішень і байєсівські мережі були успішно застосовані для прогнозування коливання цін на акції та деяких макроекономічних процесів.

Під час аналізу реальних статистичних даних різних фінансових систем був сформований набір методів і моделей, які доцільно використовувати для вирішення задач аналізу фінансових ризиків. Зокрема, запропоновано використання узагальнених лінійних моделей (УЛМ), регресійних моделей, методів фільтрації, мереж Байєса, дерев рішень, нейронних мереж, і т. ін. Використання кожного методу або моделі буде суттєво залежати від постановки задачі.

УЛМ як клас моделей доцільно застосовувати у випадках, коли дані демонструють статистичний розподіл, що відрізняється від нормального. Побудову УЛМ можна розглядати з класичною статистикою або з байєсівської точки зору [256]. Для побудови таких моделей необхідно висунути припущення щодо типу розподілу параметрів моделі, обрати метод оцінки параметрів на основі методів імітаційного, ієрархічного та апостеріорного моделювання. Ці моделі доцільно використовувати для задач динамічного оцінювання фінансових ризиків, оцінювання ймовірності повернення кредитів, а також для нелінійних фінансових процесів.

Ще одним способом використання адаптивного підходу є комбінація прогнозів у тих випадках, коли стандартні методи не дозволяють отримати високу якість прогнозування. Для таких випадків обирають ідеологічно різні методи прогнозування, а далі обчислюють комбіновану оцінку прогнозу з

використанням обраних ваг (наприклад, на основі досвіду експертів або практичного досвіду попереднього використання методів).

Застосування такого підходу надає такі переваги: автоматичний пошук «кращої» моделі скорочується в часі у багато разів; можна аналізувати набагато ширший набір моделей-кандидаток; оптимізується пошук завдяки використанню комплексного критерію якості; можна інтегрувати ідеологічно різні методи моделювання і прогнозування та обчислювати комбіновані оцінки прогнозів, які відрізняються кращою якістю [217]. Застосування такого адаптивного підходу в ІТ для аналізу ризиків на фінансових ринках та прогнозування макроекономічних даних дозволило отримати значення абсолютної похибки у відсотках на рівні 3 – 7% для короткострокового прогнозування.

Для застосування методів оптимальної фільтрації типу фільтру Калмана необхідно представити моделі у просторі станів. ФК дозволяє регулювати вплив випадкових зовнішніх збурень (шуму, який завжди існує) і враховувати можливі похибки вимірювання (шум вимірів). У більшості випадків практичного застосування такий підхід забезпечує високу якість короткострокових прогнозів завдяки доступності оптимальних оцінок стану, обчислених за допомогою алгоритму фільтрації [74].

Використання мереж Байеса для адаптивного менеджменту дозволяє враховувати інформацію від експертів, перераховувати нові ймовірності для кожного вузла-змінної в разі надходження нових даних, оптимізувати структуру моделі у вигляді ациклічного графу [13, 18, 20, 21, 90, 95, 99, 104, 122, 125].

Метод групового урахування аргументів (МГУА) генерує модель прогнозування у вигляді полінома Колмогорова-Габора і може бути використаний як для лінійних, так і нелінійних систем. МГУА автоматично обирає найкращі структури моделі в класі попередньо вибраної лінійної або нелінійної структури і дозволяють отримувати вищі оцінки прогнозів [71, 371].

Запропонований підхід на основі описаної вище схеми адаптації був апробований для прогнозування процесів фінансових систем, зокрема для макроекономічних процесів. Особливості запропонованого підходу полягають у наступному: тестування якості даних з набором статистичних параметрів; неперервний аналіз даних, спрямований на виявлення структури прогнозної моделі і її параметрів, генерацію моделей-кандидаток і вибір кращої з іншим набором параметрів якості моделі; формування набору оцінок прогнозів на основі моделей-кандидаток [16, 127, 217]. Найкраща оцінка прогнозу обирається за набором параметрів вищої якості прогнозів. Весь обчислювальний процес контролюється трьома наборами параметрів якості, що гарантує належну якість кінцевого результату.

Лінійний фільтр Калмана успішно застосовано для обчислення оптимальних оцінок залишків і короткострокових прогнозів, заснованих на обраних моделях. Коваріації стану шуму були оцінені рекурсивно з новими даними, які надходять, що відповідає загальній ідеології адаптації. Найкращі результати прогнозування, на прикладі прогнозування цін на золото, були отримані із застосуванням фільтра Калмана (одно- і двокрокового прогнозування) для моделей з AR (1) + трендом 4-го порядку і наведені у таблиці 5.1 [217]. Результати, отримані в цьому випадку кращі, ніж без застосування оптимального фільтра. Краща модель, побудована для прогнозування умовної дисперсії, була експоненційною УАРУГ (1,7) + з трендом 9-порядку. Досягнуте значення MAPE = 8,74% є хорошим результатом для прогнозування умовної дисперсії.

Після використання запропонованого адаптивного підходу (схема якого представлена на рис. 5.1) із застосуванням комплексного статистичного критерію якості моделі (5.1) спостерігалось поліпшення прогнозів в середньому в межах від 0,5 – 1,5%, що є важливим для вирішення задачі прогнозування фінансових ризиків, оскільки в грошовому еквіваленті надає значно кращі результати оцінок. Використання комбінації прогнозів, отриманих різними

методами прогнозування, сприяло подальшому зниженню середньої абсолютної похибки у відсотках прогнозування приблизно на 0,3 – 0,6% для описаного конкретного прикладу.

Таблиця 5.1

Результати моделювання із застосуванням адаптивного підходу для задачі оцінювання цін акцій на золото

Тип моделі	Якість моделі			Якість прогнозу			
	R^2	$\sum e^2(k)$	DW	MSE	MAE	MAPE	Theil
Результати якості моделей та прогнозування без ФК							
АР(1)	0,99	25644,67	2,15	49,82	41,356	8,37	0,046
АРКС(1,4)	0,99	25588,10	2,18	49,14	40,355	8,12	0,046
АР(1) + тренд 1го порядку	0,99	25391,39	2,13	34,39	25,109	4,55	0,032
АРКС(1,4) + тренд 1го порядку	0,99	25332,93	2,18	34,51	25,623	4,67	0,032
АР(1) + тренд 4го порядку	0,99	25173,74	2,12	25,92	17,686	3,19	0,024
Результати якості моделей та прогнозування з використанням ФК							
АР(1)	0,99	24376,32	2,11	45,21	39,73	7,58	0,037
АРКС(1,4)	0,99	24141,17	2,09	47,29	38,75	7,06	0,035
АР(1) + тренд 1го порядку	0,99	23964,73	2,08	31,15	22,11	3,27	0,029
АР(1) + тренд 4го порядку	0,99	22396,83	2,04	21,35	13,52	2,71	0,019
Результати якості моделей та прогнозування умовної дисперсії							
УАРУГ(1,7)	0,99	153639	0,113	972,5	–	517,6	0,113
УАРУГ (1,15) + тренд 2го порядку	0,99	102139	0,174	458,7	–	211,3	0,081
УАРУГ (1,15) + тренд 5го порядку	0,99	80419	0,337	418,3	–	121,6	0,058
ЕУАРУГ(1,7) + тренд 9го порядку	0,99	45184	0,429	67,8	–	8,74	0,023

Розширення адаптивної схеми прогнозування можливе за рахунок доповнення її запропонованими динамічними моделями, ймовірнісно-статистичними моделями, комбінованими методами відновлення втрачених даних, використанням нових критеріїв для оцінки якості моделей та ефективності рішень, а також застосування альтернативних методів для прогнозування фінансових ризиків.

Використання запропонованого підходу забезпечує отримання високоякісних (за точністю) коротко- та середньострокових прогнозів за умов наявності інформативних даних, а також формування на їх основі альтернативних раціональних рішень та передбачає універсальність застосування до широкого класу процесів [61].

5.3. Структурно-параметрична адаптація ймовірно-статистичних моделей

Створення адекватних моделей менеджменту ризиків є дуже важливою задачею, адже саме ці моделі дозволяють моделювати ризики, оцінювати можливі втрати, і на їх основі приймати управлінські рішення. Тому потрібно обрати коректний інструментарій, визначити обмеження, якщо вони існують при застосуванні окремих методів, а також види та типи невизначеності, які формують появу обраного модельованого класу ризику. Далі обирають найважливіші або суттєві змінні, що можуть характеризувати появу ризику, обирають критерії якості даних та адекватності моделей [127]. Ієрархічно цей етап представляється такими кроками:

1. Обрати типи фінансових ризиків, які будуть моделюватись.
2. Вибрати класи (типи) моделей для опису ризиків. Розробити різні ймовірно-статистичні моделі, які дозволяють оцінити ймовірність появи ризику та можливі втрати на основі історичних даних, що описують і характеризують появу ризику даного типу.
3. Вибрати (розробити) критерії якості:
 - ✓ критерії оцінювання якості даних;
 - ✓ критерії якості моделей;
 - ✓ критерії якості класифікації або оцінок прогнозів.
4. Виконати аналіз невизначеностей, характерних для кожного етапу аналізу ризиків [127]:
 - ✓ невизначеності даних;

- ✓ невизначеності структури і параметрів моделей;
- ✓ невизначеності оцінок прогнозів.

Цей етап є обов'язковим підготовчим етапом для реалізації методу структурно-параметричної адаптації ймовірно-статистичних моделей менеджменту фінансових ризиків.

5.3.1. Метод структурно-параметричної адаптації менеджменту ризиків

Запропонована у підрозділі 5.2 схема адаптації стане основою при розробці методу, що передбачає побудову ймовірнісних та статистичних моделей на основі статистичних даних, які характеризують появу і реалізації фінансових ризиків, а в процесі налаштування буде можливим здійснювати коригування структури самої моделі та налаштування її параметрів.

Метод структурно-параметричної адаптації відрізняється наявністю двох контурів адаптації для корегування і налаштування кращої моделі оцінювання ризиків [127]. Перший контур адаптації необхідний тоді, коли в результаті формування рішення щодо менеджменту ризиків отримуємо незадовільну якість рішення (наприклад, ймовірність ризику, яка прогнозується кращою моделлю, є нижчою порівняно з реальними даними, або втрати внаслідок реалізації фінансового ризику є занадто високими). Тоді необхідно повернутися на етап статистичного аналізу даних, виконати повторну оцінку параметрів моделі, можливо, навіть зміну її структури, і визначити нові параметри моделі. Необхідність використання у методі другого контуру адаптації пов'язана з динамічним процесом оцінювання фінансових ризиків в реальному часі [115]. Якщо розроблена модель оцінювання фінансових ризиків будується в ІСППР в реальному часі, і одразу використовується керівництвом компанії, але не задовольняє поставленим вимогам ефективності (наприклад, змінилась політика компанії щодо збору даних або оцінювання ризиків), то передбачена можливість виконання зовнішнього контуру адаптації. Знову виконується аналіз статистичних даних, з додаванням нових змінних, характеристик,

видаленням існуючих, побудовою нових оцінок, додаванням відхилених даних, зовнішніх даних, тощо [127]. Фактично відбувається навчання, тільки не самої моделі, а сукупності моделей, побудованих заново на уточнених даних, з урахуванням нових критеріїв оцінювання моделей-кандидаток. Таким чином формується цілий «портфель моделей», з якого обирається краща за визначеними критеріями, і тільки після цього відбувається передача нової моделі особі, що приймає рішення. ОПР формує нове рішення щодо менеджменту ризиків на основі запропонованих критеріїв, а це рішення щодо менеджменту ризику реалізується за обраною стратегією на практиці. Ілюстрація роботи запропонованого алгоритму адаптації при вирішенні задачі менеджменту ризиків подана на рис. 5.2.

Метод дозволяє як використовувати експертне знання щодо менеджменту ризиків, так і застосовувати попередньо зібрані статистичні дані, які об'єднуються в єдину базу знань та даних (БЗД). Початкові дані для моделювання ризиків подаються на вхід на першому кроці.

Крок 1. Попередня обробка вхідних даних.

Аналіз вхідних даних та знань виконується з метою їх перевірки, консолідації з різних джерел, обробки втрачених або пропущених значень (за можливості), вилучення некоректних та спотворених даних [127]. Перевіряється надійність та стабільність даних (відсутність маніпуляцій з даними), зручність та легкість інтерпретації даних, можливість отримання таких даних у майбутньому, урахування коефіцієнтів, важливих з точки зору бізнесу.

Крок 2. Статистичний аналіз даних та побудова множини моделей-кандидаток.

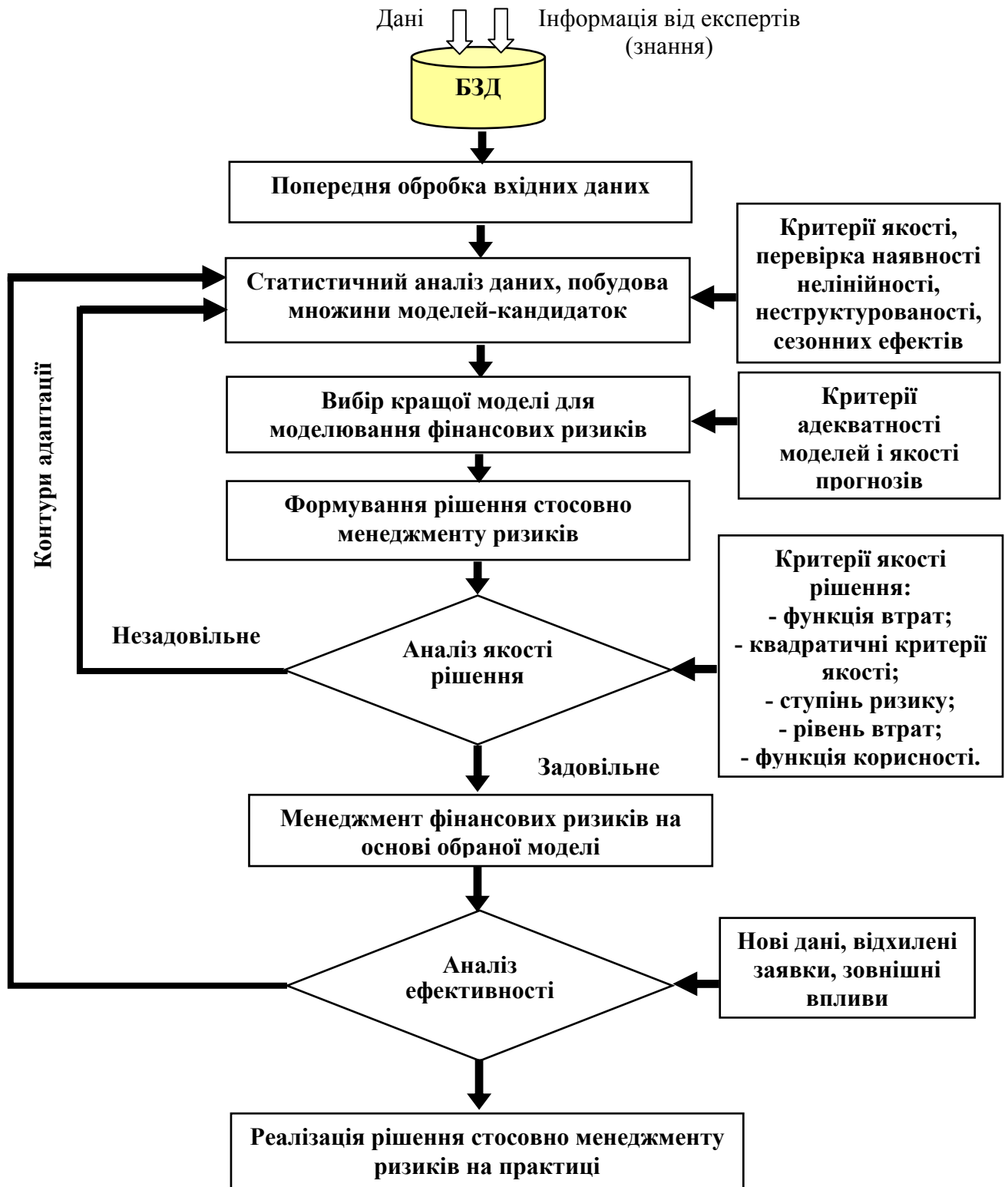


Рис. 5.2 Алгоритм реалізації процесу моделювання із структурно-параметричною адаптацією моделей для менеджменту ризиків

Обраховуються значення WOE , що визначає наскільки коректно сформовані інтервали для даної характеристики за формулою (2.4) [127]:

$$woe_j = \ln\left(\frac{Distr\ Good_i}{Distr\ Bad_i}\right). \quad (5.2)$$

Передбачуючу силу кожної характеристики оцінюють за інформаційним значенням за формулою (2.5)[108]:

$$\begin{aligned} IV &= \sum_{i=1}^L (Distr\ Good_i - Distr\ Bad_i) * woe = \\ &= \sum_{i=1}^L (Distr\ Good_i - Distr\ Bad_i) * \ln\left(\frac{Distr\ Good_i}{Distr\ Bad_i}\right), \end{aligned} \quad (5.3)$$

де L – це кількість атрибутів (рівнів або проміжків) для даної характеристики.

2.2. Формування структури моделей-кандидаток здійснюється у декілька етапів, з використанням наступних оцінок:

✓ *Дискретна функція взаємної кореляції:*

$$r_{yx}(s) = \frac{1}{N-1} \sum_{k=s+1}^N [y(k) - \bar{y}][x(k-s) - \bar{x}], \quad s = 0, 1, 2, \dots \quad (5.4)$$

Кореляція характеризує наявність (відсутність) лінійної або нелінійної залежності між змінними. Вибірковий *коефіцієнт кореляції* між двома змінними обчислюється за формулою [26]:

$$r_{yx} = \frac{1}{N-1} \frac{\sum_{k=1}^N \{[y(k) - \bar{y}][x(k) - \bar{x}]\}}{\sigma_x \sigma_y}, \quad (5.5)$$

де $-1 < r_{yx} < +1$; σ_x, σ_y – стандартні відхилення для змінних x і y

Кореляційний аналіз дозволяє встановити наявність зв'язків між значеннями однієї вибірки даних та між значеннями декількох вибірок.

✓ *Критерій для аналізу нелінійності і нестационарності.*

На практиці тестова статистика стосовно нелінійності розраховується з використанням оцінок, отриманих за МНК, таким чином [26, 127]:

– будується регресія $y(k)$ на $w(k)$, обчислюються залишки цієї моделі $\tilde{y}(k)$ і сума квадратів залишків SSR_0 ;

$\mathbf{w}(k) = [1, y(k-1), \dots, y(k-p); x_1(k), \dots, x_l(k)]^T$ – вектор вимірів основної змінної та регресорів;

- будується регресія $\tilde{y}(k)$ на $\mathbf{w}(k)$ і $\tilde{\mathbf{h}}(k)$, обчислюється сума квадратів залишків цієї моделі SSR_1 ;
- обчислюється тестова статистика

$$F(m, N - n - m) = \frac{(SSR_0 - SSR_1) / m}{SSR_1 / (N - n - m)}, \quad (5.6)$$

де $n = l + p + 1$; m – розмірність вектора параметрів θ ; обчислена величина має F – розподіл при $\theta = 0$ [127]. Використання F – статистики замість тесту χ^2 рекомендується при коротких вибірках.

При перевірці наявності нестационарності (тобто наявності одиничного кореня) використовується тест Дікі-Фуллера [26]. Для визначення наявності одиничного кореня використовують три наступні рівняння:

$$\Delta y(k) = \gamma y(k-1) + \varepsilon(k), \quad (5.7)$$

$$\Delta y(k) = a_0 + \gamma y(k-1) + \varepsilon(k), \quad (5.8)$$

$$\Delta y(k) = a_0 + \gamma y(k-1) + a_2 k + \varepsilon(k), \quad (5.9)$$

де k – дискретний час; $\gamma = a_1 - 1$ – коефіцієнт у рівнянні: $y(k) = a_0 + a_1 y(k-1) + \varepsilon(k)$. Рівняння (5.7) є моделлю випадкового кроку (або «блукання»), (5.8) включає зсув у вигляді константи a_0 , а (5.9) включає зсув та детермінований лінійний часовий тренд.

Якщо $\gamma = 0$, то послідовність $\{y(k)\}$ містить одиничний корінь. Застосування тесту Дікі-Фуллера (ДФТ) передбачає оцінювання одного або більше з наведених вище трьох рівнянь за допомогою МНК або ММП з метою отримання оцінки параметра γ та стандартної похибки цієї оцінки. На основі оцінки та її стандартної похибки обчислюється t – статистика, яка порівнюється зі значеннями, наведеними у таблицях Дікі-Фуллера [127]. На основі цього

порівняння приймається рішення щодо справедливості або відхилення нуль-гіпотези, що $\gamma = 0$.

✓ *Моделювання сезонних ефектів.*

Сезонний ефект враховується шляхом включення в модель змінної із запізненням, яке відповідає періодичності сезонного ефекту. Так, якщо сезонний ефект проявляється щоквартально, то при обробці квартальних даних ($T_s = 3$ місяці) у модель необхідно ввести змінну із запізненням $d = 4$ [26]. Ефект можна врахувати в *адитивній формі* за допомогою основної (залежної) змінної або випадкової змінної у ковзному середньому, тобто, можливі такі структури рівняння:

$$y(k) = a_4 y(k-4) + \varepsilon(k), \quad |a_4| < 1; \quad (5.10)$$

$$y(k) = \varepsilon(k) + \beta_4 \varepsilon(k-4). \quad (5.11)$$

Для рівняння (5.10) автокореляційна функція має вигляд:

$$\rho(s) = \begin{cases} (a_4)^{s/4}, & \text{якщо } s/4 \text{ ціле число,} \\ 0, & \text{якщо } s/4 \text{ дробове.} \end{cases}$$

Тобто, АКФ має ненульові значення при $s = 4, 8, 12, \dots$. Для рівняння (5.11) АКФ має одне ненульове значення при $s = 4$.

Для підвищення ступеня адекватності моделі процесу в цілому рівняння (5.10), (5.11) розширюють авторегресією з ковзним середнім, наприклад, 1-го порядку [26]:

$$y(k) = a_1 y(k-1) + a_4 y(k-4) + \varepsilon(k) + \beta_1 \varepsilon(k-1),$$

$$y(k) = a_1 y(k-1) + \varepsilon(k) + \beta_1 \varepsilon(k-1) + \beta_4 \varepsilon(k-4).$$

Для врахування сезонного ефекту у *мультиплікативній формі*:

$$y(k)(1 - a_1 L) = \varepsilon(k)(1 + \beta_1 L)(1 + \beta_4 L^4) \text{ або}$$

$$y(k)(1 - a_1 L)(1 - a_4 L^4) = \varepsilon(k)(1 + \beta_1 L).$$

✓ *Оцінювання параметрів моделей-кандидаток*

Оцінювання ризиків потребує розрахунок ймовірності їх виникнення, а тому доцільно використати ймовірнісні моделі для побудови моделі-

кандидатки. Для оцінювання рівня ризику, тобто можливих втрат, можна використати комбіновану модель. Для цього доцільно використовувати комбінації моделей на основі мереж Байеса, логістичної регресії, марківських ланцюгів [108, 127, 312].

Статистичний висновок прийнятний у випадках, коли спостереження об'єднують в групи подібної структури, кожна з яких відповідає конкретному розподілу. Припускається, що при заданих значеннях параметрів моделі спостереження для одного випадку не залежать від спостережень для інших випадків, а розподіли цих змінних є однаковими для всіх випадків. Тоді, якщо $X = \{X_{i1}, \dots, X_{in}\}$ – змінні для випадку i , а $\theta = (\theta_1, \dots, \theta_p)$ – параметри моделі, розподіл змінних для всіх випадків [263, 313, 316]:

$$P(x_1, x_2, \dots | \theta) = \prod_i P(x_i | \theta) = \prod_i P(x_{i1}, \dots, x_{in} | \theta_1, \dots, \theta_p), \quad (5.12)$$

де $P(x_{i1}, \dots, x_{in} | \theta_1, \dots, \theta_p)$ – функція параметрів моделі та значень x_{ij} . Кількість випадків вважається невизначеною, хоча практично розглядаються тільки наявні спостереження, а також ті, для яких необхідно зробити прогноз.

Байєсівський ймовірнісний висновок потребує додаткових вхідних даних, тобто апіорного розподілу ймовірностей для параметрів [313]: $P(\theta_1, \dots, \theta_p)$. При поєднанні апіорного розподілу параметрів з умовним розподілом для спостережуваних даних одержують спільний розподіл [127, 317]:

$$P(\theta_1, \dots, \theta_p, x_1, x_2, \dots) = P(\theta_1, \dots, \theta_p) P(x_1, x_2, \dots | \theta_1, \dots, \theta_p) = P(\theta) \prod_i P(x_i | \theta) \quad (5.13)$$

Звідси можна отримати *байєсівське правило* для *апостеріорного* розподілу параметрів при заданих спостереженнях для X_1, \dots, X_C [313]:

$$P(\theta | x_1, \dots, x_C) = \frac{P(\theta, x_1, \dots, x_C)}{P(x_1, \dots, x_C)} = \frac{P(\theta) \prod_{i=1}^C P(x_i | \theta)}{\int P(\tilde{\theta}) \prod_{i=1}^C P(x_i | \tilde{\theta}) d\tilde{\theta}}. \quad (5.14)$$

Пошук оцінок за методом Монте-Карло може привести до необхідності використання методів дискретизації ланцюгів Маркова [263].

Метод Монте-Карло використовують також для перевірки, чи відповідає обраний апіорний розподіл фактичним апіорним уявленням. Для спостереження будь-яких даних можна знайти прогнозуючий апіорний розподіл для X_1, \dots, X_C :

$$P(x_1, \dots, x_C) = \int P(\tilde{\theta}) \cdot \prod_{i=1}^C P(x_i | \tilde{\theta}) d\tilde{\theta}. \quad (5.15)$$

Якщо існує вибірка даних, згенерована за цим розподілом, то її можна перевірити по відношенню до очікувань, які ґрунтуються на апіорних уявленнях. У випадку, коли вона не буде репрезентативною (стосовно апіорних очікувань), необхідно переглянути апіорний розподіл, або, можливо, і всю структуру моделі.

Генерування значення з апіорного прогнозуючого розподілу може здійснюватись спочатку через генерування значення θ з апіорного розподілу параметрів, а потім через генерування значень X_1, \dots, X_C з їх розподілу, що є умовним для заданого θ . Навіть для складних моделей ці операції досить часто виявляються простими. В інших випадках для розв'язання цієї задачі використовують метод дискретизації (генерування) ланцюгів Маркова [263].

Виразення апіорних оцінок через гіперпараметри. Для зручнішого вираження апіорного розподілу параметрів моделі іноді використовують додаткові параметри – *гіперпараметри*. Наприклад, апіорна оцінка для множини параметрів $\theta_1, \dots, \theta_p$ може бути представлена як маргінальний розподіл з використанням гіперпараметра α таким чином:

$$P(\theta_1, \dots, \theta_p) = \int P(\tilde{\alpha}, \theta_1, \dots, \theta_p) d\tilde{\alpha} = \int P(\theta_1, \dots, \theta_p | \tilde{\alpha}) P(\tilde{\alpha}) d\tilde{\alpha}. \quad (5.16)$$

Ця технологія може бути поширена на будь-яке число рівнів, а отриманий результат іноді називають *ієрархічною моделлю*. Відношення залежності між гіперпараметрами, параметрами та спостережуваними змінними можна виразити за допомогою формалізму довірчих мереж (мереж Байєса) [26].

Крок 3. Вибір кращої моделі для моделювання фінансових ризиків.

Сукупність моделей-кандидаток дозволяє оцінювати ризики різними методами, і в разі накладання додаткових обмежень, які можуть не дозволити функціонувати одній з моделей, є можливість обрання іншої моделі. Саме тому на попередньому етапі будувалось декілька моделей, формувалась їх структура, оцінювались параметри [127]. Вибір кращої моделі можливий експертним шляхом з урахуванням досвіду, практичного застосування, можливостей і обмежень ІСППР. Проте більш об'єктивним є вибір кращої моделі за сукупністю критеріїв адекватності моделей та вищих прогнозних оцінок.

✓ *Критерії адекватності моделей.*

Для оцінювання адекватності моделювання використовують сукупність критеріїв якості для моделей, описаних у третьому розділі: R^2 , GINI, AIC, DW, BS, MS, тощо.

✓ *Критерії якості прогнозів.*

Для перевірки прогнозних оцінок використовуються описані у 3.1 оцінки MPE , $MAPE$, ME , MAE , SE , MSE , тощо [26].

Крок 4. Формування рішення стосовно менеджменту ризиків.

Менеджмент ризиків передбачає оцінювання ступеню (ймовірності) та рівня ризику (можливих втрат) [127]. Прийняття рішення щодо менеджменту ризиків означає, що фінансова установа має визначити, чи є заданий рівень ризику прийнятним або допустимим [160], критичним чи катастрофічним для установи. Розробляються моделі для:

✓ *оцінювання ймовірності ризику,*

$$P = F(x_i, w_i),$$

де x_i – параметри, що характеризують появу ризику, w_i – їх ваги, F – функція певного виду;

✓ *розрахунку втрат від реалізації ризику.*

Під час статичного оцінювання ризиків (п. 3.2) обчислення ймовірності настання ризику та обсяг втрат виконується за IBR-підходом [108, 109, 115]:

$$EL = \sum_{i=1}^N P(R_i) \cdot CE_i \cdot LGD_i.$$

При динамічному оцінюванні ризиків (описаному у розділі 4) будується множина динамічних моделей, що оцінюють ризики, використовуються встановлені ОПР допустимі рівні або ступені ризику та визначається момент часу настання ризику [115].

Крок 5. Аналіз якості рішення.

Для аналізу та оцінювання ризиків можуть використовуватись різноманітні критерії. Найпоширеніші критерії, що будуються на основі функцій корисності для факторів, що дозволяють оцінити ефективність менеджменту ризиків через втрати і ймовірність появи ризику.

Наприклад, *функція корисності* виду:

$$y(x) = \sum_{i=1}^m \alpha_i x_i,$$

де α_i – вагові коефіцієнти, $\sum_{i=1}^m \alpha_i = 1$,

x_i – фактори, які враховують корисність методології менеджменту ризиків.

$$\begin{cases} x_1 = EL(\text{очікувані втрати}), \\ x_2 = UL(\text{неочікувані втрати}), \\ x_3 = 1 - P(\text{втрат}). \end{cases}$$

Можливе використання стандартних критеріїв оцінювання ризиків через реальні капіталовкладення та втрати. Автором запропонований власний *оригінальний квадратичний критерій якості* опрацювання ризиків (3.8), описаний у третьому розділі:

$$I_1 = \sum_{k=1}^N [x^T(k) \times R \times x(k) + (y(k) - y_\phi(k))^T Q (y(k) - y_\phi(k)) + z^T(k) P z(k)] \rightarrow \min_{y,x,z},$$

де $x^T(k) \times R \times x(k)$ – витрати на інвестиції,

$y(k) - y_\phi(k) = \tilde{y}(k)$ – реальні втрати,

$z^T(k) \times R \times z(k)$ – компенсаційні витрати.

Цей критерій дозволяє перевірити ефективність менеджменту ризиків, виходячи з того, які сумарно витрати були здійснені для мінімізації ризиків. Перший доданок оцінює витрати на розробку та впровадження всієї методології менеджменту ризиків, її удосконалення та задіяні інвестиції [127]. Другий доданок оцінює відхилення прогнозованої суми втрати за моделлю від реального значення втрат від ризику (фактичного), третій – це витрати, які були задіяні для реалізації ефективної стратегії боротьби з ризиком (внаслідок дій конкурентів, зовнішніх невизначеностей), пов'язані із залученням додаткових засобів захисту, механізмів страхування, хеджування, маркетингу, тощо.

Крок. 6 Менеджмент фінансових ризиків на основі обраної моделі в реальному часі.

Обрана краща модель впроваджується в реальну систему підтримки прийняття рішень для оцінювання ризиків реального фінансового підприємства. Здійснюється збір статистичних даних щодо прогнозних оцінок для ризиків за кращою моделлю порівняно з реальними результатами моделювання [127], формуються так звані вибірки відхилених заявок та напрацьовуються рішення та стратегії боротьби з ризиком у випадку його переходу із зони допустимого до критичного.

Крок 7. Аналіз ефективності рішень щодо менеджменту ризиків в реальному часі.

Для перевірки ефективності запропонованої моделі та прийнятих на її основі рішень передбачена апробація моделі на практиці з урахуванням нових даних, збурень, зовнішніх впливів.

Ефективність менеджменту ризиків оцінюється за сукупністю наступних характеристик [127]:

$$E = \langle A, F, C, D, I \rangle,$$

де A – адекватність побудованої моделі,

F – прогноз прийнятної якості,

C – прийнятні обчислювальні витрати,

D – рішення, що сформоване на основі розробленої моделі та прогнозу,

I – узгодженість з даними, $I = \begin{cases} 1 - \text{дані погоджуються з моделлю,} \\ 0 - \text{дані не погоджуються з моделлю.} \end{cases}$

Для перевірки ефективності залучається експерт. Якщо модель залишається ефективною, а прийняті рішення коректними, то дана модель продовжує використовуватись на практиці. Якщо рішення некоректне, а модель неефективна, то виконується зовнішній контур адаптації [127]. При цьому до історичних даних, на яких будувалась початкова модель, додаються нові дані з урахуванням відхилених заявок, оцінок, отриманих нашою моделлю, зовнішніх даних, і здійснюється повторно побудова моделей-кандидаток та оцінювання їх параметрів.

Крок 8. Реалізація рішення для менеджменту ризиків на практиці.

Заключний крок методу передбачає впровадження моделі на практиці. У випадку, якщо модель не змінилась, це означатиме використання тої самої скорингової карти ризику, яка дозволяє оцінювати ризики, автоматично прораховуючи обрані характеристики, і визначаючи, чи ризик є допустимим [127]. У випадку будь-якої корекції структури або параметрів моделі, або обрання іншої моделі, виникає необхідність налаштування нової моделі в системі підтримки прийняття рішень, використання нової моделі для побудови скорингової карти та оцінювання ризиків.

5.3.2. Використання структурно-параметричної адаптації під час моделювання кредитних ризиків

При моделюванні кредитних ризиків найчастіше використовують ймовірнісно-статистичні методи, які дозволяють оцінити ймовірність виникнення кредитного ризику та наслідки, тобто втрати, від його прояву. Для цього розробляють так звані скорингові моделі, при цьому виконується перший контур адаптації, тобто декілька разів переналаштовуються параметри моделі для модифікації її структури і отримання високих скорингових оцінок. Далі за

сукупністю критеріїв якості обирається краща модель, на основі якої будується скорингова карта, яка і впроваджується в реальній фінансовій установі. Після того, як розроблена скорингова карта відпрацювала певний період, були отримані статистичні дані щодо прогнозних та реальних оцінок кредитного ризику, виконується переоцінка і перевірка придатності розробленої раніше моделі. На рис. 5.3 показано, що вибірка, яка використовувалась для побудови скорингової карти, оцінюється компонентою Score. Далі розраховується ймовірність настання дефолту і порівнюється з порогом відсікання та реальним станом, чи був кредит повернутий. Розраховується процент неправильної класифікації (помилки першого та другого роду). Для нашої тестової вибірки він становив 11%. Для нової вибірки, яка була отримана після впровадження скорингової карти, також розраховуємо ймовірність і перевіряємо процент неспівпадіння при відомому порозі відсікання. Він становить 15%, проте для банку це занадто високий результат. Це означає, що потрібно виконати адаптацію моделі до нових статистичних даних. Виконуємо другий контур адаптації, який передбачає підвантаження статистичних даних щодо відхилених заявок за заданою моделлю, прийнятих кредитних рішень, тощо. На зовнішньому контурі адаптації може бути побудована нова скорингова карта. Для нашої задачі при структурно-параметричній адаптації моделі була отримана доля неправильної класифікації на рівні 12,5% для нового набору даних, що є нижчим від початкової моделі і прийнятним для банку, тож ця модель була використана для побудови нової скорингової карти. Послідовність виконання двох контурів адаптації для кредитних ризиків на прикладі налаштування і структурно-параметричної адаптації скорингових моделей показана на рис. 5.3.

Детальніше етапи та особливості розробки скорингових карт були описані у [108]. Структурна адаптація передбачає додавання в модель нових параметрів, які раніше не були включені, можлива зміна порядку моделі, та навіть вибір нової моделі з сукупності відібраних на початковому етапі

ймовірно-статистичних моделей. Параметрична адаптація передбачає як налаштування параметрів самої моделі, так і адаптацію і коригування параметрів і ваг для скорингової карти, що будується на основі обраної моделі.

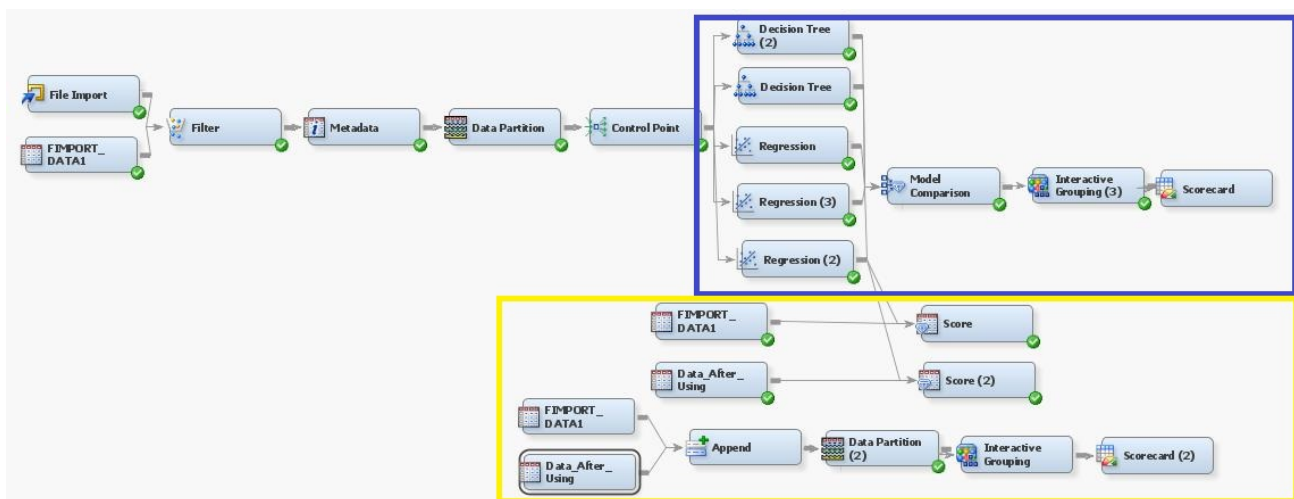


Рис. 5.3 Структурно-параметрична адаптація при розробці скорингової карти

Структурно-параметрична адаптація менеджменту кредитних ризиків продовжується і після повторного впровадження оновленої скорингової карти [106], здійснюється процес моніторингу та аналізу ефективності скорингової карти, уточнення і коригування її параметрів [108], оскільки змінюється зовнішній світ, фінансове становище, законодавство, фактори та характеристики, що були включені в скорингову модель, та ключові фактори оцінки. За рахунок такої адаптації є можливість реального уточнення і подальшого застосування скорингової карти в реальному часі, що є важливим для фінансової установи, яка продовжує здійснювати свою фінансову діяльність і має опрацьовувати кредитні рішення і приймати рішення, не зупиняючи свою діяльність в очікуванні розробки нового методу та відповідних моделей для оцінювання ризиків.

Висновки до п'ятого розділу

Запропонований принцип адаптивного менеджменту ризиків, розвинутий у динамічному підході та методі структурно-параметричної адаптації, передбачає реалізацію двох контурів адаптації: внутрішнього і зовнішнього, що забезпечує динамічне оцінювання фінансових ризиків в реальному часі через ймовірнісну та вартісну складову. Використання двох контурів адаптації дозволяє одразу частково подолати ризик, оскільки зменшується невизначеність, яка спричиняє появу ризиків, через урахування нових статистичних даних та збурень, а також і скоригувати рішення в разі, якщо воно виявилось неефективним.

Схема структурно-параметричної адаптації є універсальною і може бути застосована для різних типів ризиків. Апробація запропонованого підходу здійснювалась на фінансових ризиках, зокрема кредитних, та для прогнозування цін на акції на фінансових ринках, проте у подальших дослідженнях можливе розширення її застосування для задач моделювання виробничої та технологічної діяльності, вирішення задач керування тощо. Єдине обмеження, яке накладалося, пов'язане скоріше з класом задач, які вирішувались (оцінювання ризиків). Необхідно було забезпечити використання таких методів, які дозволяють отримати коректні оцінки за факторами, що характеризують ризик: ймовірність і втрати. Використання методів, які дозволяють лише обрахувати втрати (наприклад, лінійна регресія) і не дають можливість оцінити ймовірність, означають неповне оцінювання ризиків і відповідно формування лише часткового рішення щодо ризику.

Запропоновані етапи попередньої підготовки даних і моделей, метод структурно-параметричної адаптації були реалізовані у вигляді окремих застосувань в інформаційній технології. Аналіз ефективності і якості рішення здійснюються в окремому блоці ІСППР, яка описана детально у шостому розділі.

РОЗДІЛ 6

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ОЦІНЮВАННЯ, МОНІТОРИНГУ ТА МЕНЕДЖМЕНТУ РИЗИКІВ ФІНАНСОВИХ СИСТЕМ

Розроблена у третьому розділі системна методологія на основі запропонованих методів, моделей та принципів для використання на практиці у реальних задачах ризик-менеджменту має бути представлена у вигляді послідовності кроків виконання та обов'язкових модулів. Оскільки фінансові системи суттєво відрізняються, концептуально можуть складатися з різного набору модулів і включати як основні методи, так і додаткові, то зрозуміло, що саме інформаційні технології є засобом представлення закладеної методології для розробки ІСППР для конкретної фінансової системи. Інформаційні технології пов'язують між собою основні потоки взаємодії, ролі, дані та знання, визначають структуру та набір модулів для вирішення набору задач. Ідеальною для фінансових систем є інформаційна технологія для інтеграції в існуючу ERP-систему, а тому доцільно розглянути технологію як реалізацію окремого веб-додатку виконання менеджменту ризиків для подальшого його інтегрування в існуючу систему управління підприємством.

У розділі описано структуру загальної (розширеної) архітектури для ІСППР на основі уніфікованої інформаційної технології, що містить можливі модулі та застосування. Ця ІСППР розроблялась і створювалась для фінансової системи, у якій не було власної системи прийняття рішень, а тому не було необхідності інтегруватись з існуючими додатками, а виникала потреба забезпечити зв'язок з існуючими базами знань і даних для аналізу, оцінювання та мінімізації постійно виникаючих у фінансовій системі ризиків.

Оскільки вирішувались також і реальні прикладні задачі ризик-менеджменту для функціонуючих фінансових систем, де застосовувались розроблені у дисертації методи і моделі, у розділі також будуть наведені

основні елементи інформаційних технологій, які інтегрувалися в інформаційні системи цих підприємств.

6.1. Формування вимог до моделюючого інструментарію

У дисертаційній роботі було показано, що фінансові ризики є несистемними, а фінансові процеси – нелінійними, тому слід включати в інформаційну технологію (ІТ) і ІСППР на її основі ті методи моделювання процесів, які дають можливість прогнозувати нелінійні та нестационарні фінансові процеси. ІТ та реалізована на її основі ІСППР дозволяють визначати найкращі моделі з мінімальними на це затратами ресурсів. Також передбачено інтуїтивно зрозумілий інтерфейс для роботи з програмним продуктом.

При розробці інформаційної системної підтримки прийняття рішень для аналізу, оцінювання та мінімізації ризиків фінансових систем важливо дотримуватись основних принципів та вимог, які висуваються загалом до СППР [238, 243, 274, 360], а також враховувати специфіку застосування розробленого комплексу у фінансових системах [300]. Інформаційна технологія та ІСППР на її основі має відповідати усім основним принципам розробленої системної методології та забезпечувати комплексний, ефективний та адекватний ризик-менеджмент з урахуванням особливостей фінансової діяльності [291], необхідності динамічного прогнозування ризиків, ступеню, профілю та рівня ризику та базуватися на таких принципах:

- ✓ ефективність менеджменту ризиків фінансових систем, що дозволяє оцінити ступінь та рівень ризику, усю повноту заходів для мінімізації втрат з урахуванням критеріїв якості опрацювання ризиків та ефективності застосованого комплексу заходів щодо менеджменту ризиків;
- ✓ своєчасність – ІСППР має працювати в реальному часі для забезпечення виявлення, оцінювання ризику на початковому етапі та забезпечити його моніторинг і своєчасне застосування запобіжних заходів;

- ✓ структурованість – розроблена ІСППР має складатися з окремих модулів, прикладних програм, додатків, що виконують недублюючі функції, має бути забезпечена послідовність їх виконання з розподілом даних, результатів, оцінок та звітів;
- ✓ розподіл обов'язків (відокремлення функції контролю і оцінювання ефективності рішень від безпосередньо менеджменту ризиків) – забезпечення розподілу функцій між окремими дійовими особами та надання обмежених повноважень щодо доступу та прийняття рішень щодо даних, критеріїв, моделей та рішень;
- ✓ всебічність та комплексність – розроблена система має здійснювати аналіз, оцінювання та моніторинг всіх обумовлених видів ризику (кредитного, фінансового, інформаційного, операційного, валютного, тощо) від початкового етапу виявлення ризику до застосування антиризикових дій і оцінювання ефективності системи ризик-менеджменту;
- ✓ відповідність – створення ефективної системи для менеджменту ризиків, що задовольняє всі етапи бізнес-процесів у фінансовій системі, відповідає вказаним видам ризику та дозволяє опрацьовувати їх згідно описаних показників (ступеню, рівню та часу);
- ✓ незалежність – від ОПР та зовнішніх факторів і впливів, що можуть спотворити результати відпрацювання системи ризик-менеджменту;
- ✓ конфіденційність – збереження особистих даних, дотримання вимог законодавства щодо захисту персональної або секретної інформації;
- ✓ однозначність та чіткість – надання результатів, що можуть бути інтерпретовані лише в певному вигляді, є зрозумілими і недвоєкими для ОПР та керівництва компанії, установи або банку;
- ✓ прозорість – можливість надання результатів для внутрішнього та зовнішнього моніторингу і аудиту щодо доцільності та ефективності роботи системи.

6.2. Розширена інформаційна технологія менеджменту ризиків фінансових систем

Особливість природи фінансових ризиків, невизначеностей, що їх обумовлюють, враховувалась в комплексному інструментарії на основі запропонованої у дисертаційній роботі системної методології менеджменту фінансових ризиків.

6.2.1. Клієнт-серверна архітектура

Інформаційна технологія і інформаційна система підтримки прийняття рішень для фінансового ризик-менеджменту реалізуються у трирівневій клієнт-серверній архітектурі і передбачають використання комплексу застосувань (застосунків) сервера застосувань, сервера бази знань і даних фінансової системи (БЗД ФС) та клієнтських частин. Розширена ІТ включає у своїх застосуваннях всі запропоновані методи та моделі, реалізації методики аналізу нестационарних процесів довільної природи на основі методів системного аналізу, передбачає і реалізує можливість ієрархічного аналізу процесів моделювання та прогнозування фінансових ризиків, враховує невизначеності структурного, параметричного і статистичного характеру. Окремі застосування реалізують динамічне оцінювання ризиків, прогнозування часу переходу ризику з прийняттого або допустимого до критичного або катастрофічного, адаптування моделей оцінювання та прогнозування ризиків до змін у процесах та застосування альтернативних методів оцінювання з метою пошуку кращих структур моделей за допомогою множини статистичних критеріїв якості. Запропоновані нові обчислювальні методи із можливістю адаптації і реагування на нові дані та зовнішні впливи, обмеження, законодавчі акти та критерії оцінювання зі зворотним зв'язком на основі використання множини статистичних параметрів якості оцінок прогнозів і моделей були інтегровані в єдину інформаційну технологію.

На рис. 6.1 наведена трирівнева архітектура розширеної інформаційної технології (ІТ) фінансового підприємства. Особливістю розробленої клієнт-

серверної технології є те, що вона дає можливість успішної реалізації необхідної функціональності розроблених алгоритмів реалізації методів у вигляді окремих програмних застосунків, які викликаються за необхідності і забезпечують ефективну підтримку бізнес-діяльності фінансової компанії.



Рис. 6.1 Клієнт-серверна архітектура розширеної інформаційної технології фінансового ризик-менеджменту

Розширена інформаційна технологія доступна користувачам фінансового підприємства через автоматизовані робочі місця. Щоб надати окремим групам користувачів доступ лише до необхідних застосунків (реалізації методів, моделей або формування звітності) адміністратори встановлюють сервер застосунків, сервер бази даних і знань фінансової системи (БЗД ФС) і застосування клієнта на фізичному сервері, формуючи обчислювальне середовище для користувачів цього клієнта. У випадку групи серверів таке

середовище формується на кластері, можуть застосовуватися технології віртуалізації [118, 158, 277].

База даних і знань об'єднує між собою базу даних (БД) у вигляді: зовнішніх та внутрішніх даних, інформації з архівів, статистичної інформації про клієнтів, товари і послуги, форми фінансової звітності, тощо. База знань (БЗ) об'єднує нормативну документацію від національних та міжнародних фінансових організацій щодо норм капіталу, рейтингів ризиків тощо; законодавчі нормативні та регулятивні акти для фінансової діяльності в країні; експертні оцінки, отримані на попередньому етапі для початкового аналізу ризиків та виявлення необхідної для збору статистичної інформації.

У запропонованій інформаційній технології на базі сервера застосувать реалізовані такі застосування:

- ✓ попередня обробка даних та знань;
- ✓ побудова моделей-кандидаток для оцінювання фінансових ризиків;
- ✓ побудова ймовірно-статистичних моделей оцінювання ризиків;
- ✓ перевірка адекватності моделей та оцінювання якості прогнозів;
- ✓ розробка скорингових карт оцінювання фінансових ризиків;
- ✓ динамічне моделювання фінансових ризиків;
- ✓ менеджмент фінансових ризиків на основі кращої моделі та скорингової карти;
- ✓ оцінювання ефективності менеджменту фінансових ризиків;
- ✓ структурно-параметрична адаптація моделей фінансових ризиків;
- ✓ формування звітності;
- ✓ інші застосування.

Інформаційна технологія може включати як усі зображені на рис. 6.1 можливі застосування, так і лише деякі з них. Сукупність сервісів і бібліотек, які реалізують найбільш важливі і критичні функції (базову функціональність) формують ядро сервера застосувать. Особливість реалізації полягає в тому, що один клієнт може отримати доступ за робочою необхідністю до багатьох різних

застосувань, а різні типи клієнтів (відповідно і різні автоматизовані робочі місця (АРМ)) отримують доступ до одного і того самого застосування, наприклад, з різним рівнем доступу. Клієнти взаємодіють через сервер застосувань, надсилаючи запити і отримуючи відповіді та необхідну інформацію з БЗД (для запису, читання, редагування). Доступ клієнтів до окремих застосувань та бази даних і знань здійснюється через сервер застосувань.

Для фізичного проектування отриманої моделі даних обрана система управління базами даних Microsoft SQL Server 2016, яка забезпечує побудову централізованих баз даних з використанням технології клієнт-сервер, дозволяє ефективну організацію роботи з одними і тими самими даними декільком користувачам [118,158]. Можливе також підключення інших застосувань в інформаційну технологію та ІСППР для виконання додаткових функцій.

6.2.2. Реалізація розробленої інформаційної технології у вигляді мікросервісів

Виходячи з бізнес-вимог, які пов'язані з необхідністю практичної розробки і інтеграції запропонованої інформаційної технології менеджменту ризиків в існуючу і вже функціонуючу систему підтримки прийняття рішень, автором було запропоновано використати архітектуру мікросервісів, яка передбачає, що кожне застосування розробляється як сукупність невеликих сервісів, кожен з яких працює у власному процесі та зв'язується з іншими сервісами через протоколи передачі даних у комп'ютерних мережах, наприклад найпоширеніший Http. Реалізація мікросервісів дозволяє розподілити окремі бізнес-задачі між різними командами розробників, кожен буде вирішувати поставлену задачу, виходячи з потреб компанії щодо часу їх впровадження та використання (термінові, нетермінові), використовуючи автоматизоване середовище для їх реалізації, процеси будуть розпаралелені, значно спрощені, а тому «легші». Також значно спрощується комунікація як усередині

застосування, так і з іншими застосуваннями, майже відсутнє централізоване керування [3].

Архітектура інформаційної системи підтримки прийняття рішень у вигляді сукупності мікросервісів, реалізована автором у дисертації, виявилась актуальною з практичної точки зору. Для розробки власних методів і підходів використовувались мікросервіси, які інтегрувались в існуючу систему. Фінансова система продовжувала працювати в нормальному режимі, інформаційна система виконувала свої звичні функції, а мікросервіси інтегрувались в існуючу систему по мірі закінчення їх розробки, надаючи додаткові функції і можливості користувачам так швидко, як це можливо (ASAP), а не по завершенню розробки всіх функцій і застосувань одночасно. Ще одним ключовим моментом виявилось те, що фінансові підприємства працюють у галузі, що дуже швидко розвивається, а тому мають орієнтуватися і швидко реагувати на зміни, щоб вирішувати нові задачі. Тому потреба «передбачення» нових функцій, можливостей та застосувань для системи є досить актуальною, проте доволі складною. Розгортання мікросервісів і їх інтеграція з основною системою – просте і оригінальне рішення, що не потребує зупинки існуючих фінансових процесів та не виникає необхідності розробки нової архітектури ІСППР. Так само, в разі необхідності видалити певну функціональність з системи з мікросервісами, це реалізується значно швидше і простіше за рахунок видалення необхідного мікросервісу без потреби переписувати всю систему цілком. Також мікросервіси дозволяють досягти масштабованості архітектури завдяки багаторазовому використанню сервісів у багатьох екземплярах. Ще однією перевагою виявилась можливість використання у різних мікросервісах різних технологій та бібліотек. Наприклад, для запропонованих на R сервісів, що реалізували розроблений автором динамічний метод і динамічні моделі, є можливість інтеграції з уже функціонуючими сервісами інформаційної системи на Python, де було реалізовано збір даних та їх обробка. Звичайно, з точки зору стандартизації

доцільно застосовувати одні й ті самі сервіси і методи розробки, тож для проектування нових методів бажано обирати ті самі методи і стандарти. Проте, якщо вже існують реалізовані алгоритми в інших середовищах, то, можливо, доцільно використати їх (хоча б на перших етапах), щоб запустити додаткові можливості одразу.

При реалізації систем у вигляді мікросервісів слід враховувати і їх слабкі місця: необхідність додаткових текстів для перевірки успішності інтеграційних тестів, реалізації для кожного мікросервісу алгоритму його розгортання, перевірки цілісності контрактів і даних [3].

6.2.3. Реалізація хмарних сервісів для ІТ фінансового ризик-менеджменту

Реалізація інформаційних технологій може бути у різних формах: Storage-as-a-Service для зберігання даних на зовнішньому сховищі, у «хмарі» [357], Information-as-a-Service («інформація як сервіс») для можливості віддаленого використання інформації, яка може змінюватися щохвилини або навіть щомиті, та Infrastructure-as-a-Service («інфраструктура як послуга»), яка надає користувачам комп'ютерну інфраструктуру у вигляді віртуальної машини або платформи, пов'язані у єдину мережу, які він самостійно налаштовує для вирішення бізнес-задач. Існують і інші хмарні рішення, проте для наших задач найцікавішими і перспективнішими є використання цих трьох технологій [143].

Для реалізації останніх тенденцій до хмарних обчислень компанія SAS розробила інструментарій SAS[®] Viya™ [338], який дозволяє об'єднати реалізацію окремих задач у вигляді мікросервісів, отримати доступ через спеціальні платформи, здійснювати аналітику на хмарі (рис. 6.2).

Така архітектура дозволяє виконати швидкі обчислення, не висуває значних вимог до персональних комп'ютерів користувачів, оскільки всі обчислення здійснюються на хмарі, а самі користувачі лише потребують доступу, наприклад через VMWare [365]. ІТ дозволяє обирати найбільш ефективне використання ресурсів хмар від публічних, приватних до гібридних,

тим самим знявши повністю або частково (за потреби) залежність від внутрішніх операцій з інформацією. Технологія дозволяє змінювати за потреби і постачальників хмар, відповідно до фінансів, потужностей та можливостей, тобто забезпечує гнучкість і лояльність системи [338].

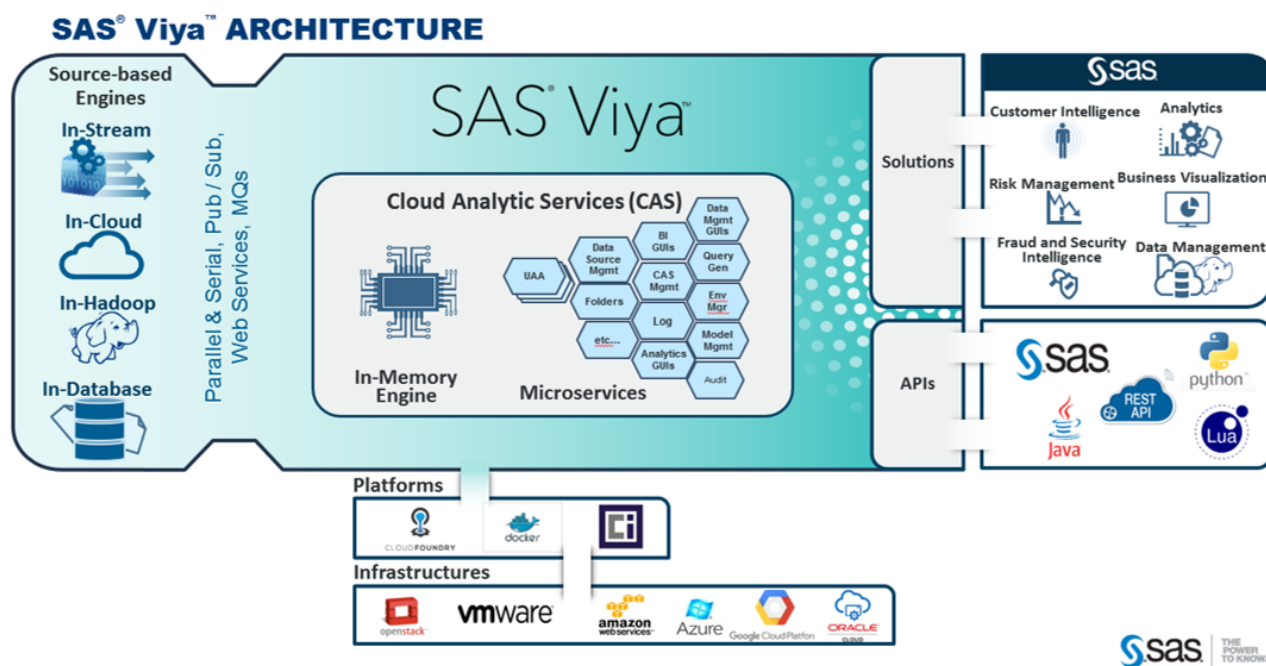


Рис. 6.2 Архітектура IaaS у SAS Viya [338]

Використання аналітики на хмарі дозволяє покращити час виконання моделі за замовчуванням, подвоїти продуктивність аналітичної групи. З'являється більше часу саме на розробку і вирішення більшої частини бізнес-задач, а це дозволяє підвищити якість обслуговування клієнтів. Використання мікросервісів [143] дозволяє залучити спеціалістів, які будуть вирішувати окремі бізнес-задачі, запускати їх на відпрацювання, не очікуючи завершення всіх розробок, паралельно здійснювати процедури.

В цій системі також додана додаткова можливість взаємодії з програмним середовищем SAS через інтерфейси прикладного програмування – це невеликий інтерфейс додатку (англ. application programming interface, API) до можливостей продукту SAS Viya [338]. Додатковою перевагою є використання різних мов програмування для реалізації власних моделей, методів та застосувань. Запускаючи різні моделі на Python / R/ SAS [2, 113, 174, 197, 205,

257, 337, 351], оцінювати їх якість та обирати ту модель реалізації, яка є швидшою по розгортанню. Написані власні коди можуть розгортатись усюди велику кількість разів. Це дозволяє легше тестувати, захищати та керувати експериментальними аналітичними застосуваннями [338].

6.3. Розробка архітектури ІСППР

Розроблена ІСППР має надати ОПР повний спектр механізмів для моделювання, оцінювання та прогнозування ФР та критеріїв для оцінювання якості даних та моделей, з можливістю адаптивного реагування для запобігання і зниження фінансових ризиків з урахуванням нових свідчень, даних та викликів в процесі моделювання.

Для розробки ефективної інформаційної системи підтримки прийняття рішень необхідно виконання всіх вимог, описаних у розділі 6.1, та забезпечення ефективного використання розробленої ІСППР для практичних задач.

Система підтримки прийняття рішень за означенням генерує множину альтернатив для надання рекомендацій особі, що приймає рішення. ІСППР має містити множину функцій та процедур для реалізації достатньої повноти рішень. Сформулюємо наступне твердження про функціональну повноту розробленої ІСППР.

Твердження про функціональну повноту ІСППР для менеджменту ризиків. Функціональна повнота ІСППР для менеджменту ризиків

$$IDSS = \{ KDB, DP, IR(MS, MMP, FMP), RE(SRE, DRE, SC), MR(DQ, MAQ, FQ, DE, SPA) \}$$

передбачає реалізацію всієї послідовності процедури менеджменту ризиків у вигляді відповідних модулів: ідентифікації, моделювання, оцінювання та мінімізації ризиків, що містять функції для обробки даних та урахування невизначеностей і інших видів ризиків, статичного та динамічного оцінювання ризиків, прогнозування ймовірності та наслідків, побудови скорингової карти як міри ризику, адаптації та перевірки якості та ефективності запропонованих

рішень. Тут *KDB* – база даних і знань; *DP* – множина процедур і функцій попередньої підготовки, консолідації і обробки даних з використанням запропонованих методів заповнення неповних/втрачених даних; *IR* – процедури ідентифікації ризику, що передбачають розробку моделі, що описує зміну ризику, оцінювання її структури і параметрів, прогнозування наслідків та ймовірності ризику. Процедури *IR* містять такі елементи: *MS* – процедури підготовки даних та оцінювання структури математичної моделі (наприклад, на основі мереж Байєса); *MMP* – блок оцінювання параметрів математичної моделі; *FMP* – модуль оцінювання прогнозів втрат (наслідків) та ймовірності настання ризику на основі обраної математичної моделі. *RE (SPE, DRE, SC)* – процедури оцінювання ризиків: статичного *SRE*, динамічного *DRE*, побудови скорингових карт ризиків *SC*. *MR (MAQ, FQ, DE, SPA)* – модуль менеджменту ризиків: на основі критеріїв адекватності математичних моделей *MAQ*, точності прогнозів *FQ*, ефективності рішень *DE* та функції структурно-параметричної адаптації моделей оцінювання ризиків *SPA*.

Кожен модуль ІСППР є її окремим функціональним елементом, який реалізує весь набір процедур, функцій і класів для інтеграції з БЗД, іншими модулями і користувачами.

Система, що розробляється, має задовольняти загальні вимоги до систем підтримки прийняття рішень [14, 360, 361]: містити сучасні бази даних, моделей, критеріїв, та необхідні обчислювальні процедури; мати зручний і простий інтерфейс; послідовність функцій виконання має відповідати людському сприйняттю та представленню; накопичувати знання та адаптуватись в процесі свого функціонування; мати необхідну швидкість виконання процедур та обчислень та необхідну точність; генерувати необхідні для ОПР форми та звіти; забезпечувати інтерактивну взаємодію з іншими користувачами; обмінюватись даними і знаннями з іншими системами обробки

інформації за допомогою комп'ютерних мереж; мати можливість для додавання нових процедур, функцій та модулів.

6.3.1. Основні рівні ІСППР

Архітектура ІСППР, що розроблялась на основі запропонованої системної методології менеджменту ризиків фінансових систем, містить наступні рівні:

- 1) завантаження і обробки даних, зокрема опрацювання невизначеностей та заповнення пропущених даних;
- 2) аналізу даних про фінансові процеси на основі методів ІАД;
- 3) побудови та обрання кращої моделі оцінювання ФР;
- 4) моделювання і прогнозування фінансових ризиків;
- 5) динамічного оцінювання ризиків на основі динамічних моделей з можливістю прогнозування фактору часу на основі заданого критичного рівня втрат або ймовірності настання ризику;
- 6) адаптації на основі структурно-параметричного методу.

Послідовне оцінювання фінансових ризиків здійснюється знизу вверх на різних рівнях розробленої архітектури ІСППР і передбачає наступні етапи (рис. 6.3):

✓ перший рівень надає можливість завантаження даних шляхом імпорту із файлу/ів (підтримка різних форматів та можливість імпорту з інших джерел), консолідація даних, можливість вибору лише обраних характеристик для аналізу та введення додаткової інформації (наприклад, маркування даних) вручну. Передбачене визначення статистично значущих змінних, застосування методів обробки даних, заповнення втрачених або пропущених даних на основі запропонованого комплексного методу, можливість опрацювання і усунення надлишковості даних, підготовка даних до аналізу;

✓ на другому рівні здійснюється аналіз даних про фінансові процеси. Передбачена можливість візуальної оцінки даних для виявлення трендів,

перевірки на стаціонарність та нелінійність, проведення статистичного та кореляційного аналізу, визначення найважливіших змінних, що характеризують фінансовий процес і мають бути включені для побудови моделей на основі методів ІАД, інтегрованих моделей та мереж Байеса;



Рис. 6.3 Рівні архітектури створеної ІСППР

✓ третій рівень передбачає множину засобів для побудови моделей-кандидаток різного типу, зокрема запропонованих ймовірно-статистичних моделей; перевірка адекватності моделей та якості прогнозів;

✓ четвертий рівень реалізує моделювання фінансових ризиків, тобто прогнозування можливих втрат, або ймовірності настання ризиків, або обох параметрів одночасно (за допомогою запропонованого автором ймовірно-статистичного методу). Прогнозування відбувається на основі обраної на третьому рівні кращої моделі;

✓ п'ятий рівень надає можливість здійснити динамічне оцінювання ризиків на основі розробленого динамічного методу та визначити момент переходу ризиків на більш високий ступінь або рівень на основі кращої динамічної моделі;

✓ шостий рівень реалізує адаптацію моделей або побудову нових адаптивних моделей з урахуванням надходження нових даних і знань, критеріїв, оцінок.

6.3.2. Структура ІСППР для менеджменту фінансових ризиків

Архітектура ІСППР доволі складна, проте може бути представлена у вигляді сукупності ідеологічно послідовно поєднаних модулів та додатків на основі розробленої системної методології, зокрема:

- ✓ *Модуль опрацювання невизначеностей та неповноти даних та знань, реалізує:*
 - комбінований метод обробки неповних (втрачених та пропущених) даних на основі МБ та регресійних моделей;
 - урахування інформаційного ризику, пов'язаного з недостовірністю, несвоєчасністю, суперечливістю та неправдивістю інформації та неточності, неповноти, нечіткості вхідних даних та знань.
- ✓ *Модуль розробки скорингової карти як міри ризику виконує:*
 - побудову скорингових моделей на основі методів ІАД;
 - обрання кращої моделі на основі одного з визначених статистичних критеріїв або за сукупністю критеріїв;
 - побудову скорингової карти і встановлення скорингових балів;
 - нейро-нечіткий метод доповнення відхилених заявок при розробці скорингової карти.
- ✓ *Модуль динамічного оцінювання фінансових ризиків здійснює:*
 - побудову динамічних моделей ризиків для прогнозування рівня та ступеня ризику;
 - розрахунок часу настання ризику (критичного, катастрофічного) в динамічному методі на основі ступеню ризику (ймовірності настання ризику);

- визначення моменту часу настання ризику на основі рівня ризику (допустимого, критичного, катастрофічного);
 - ймовірно-статистичне оцінювання ризику фінансових втрат (розрахунок обсягу втрат та ймовірності);
- ✓ *Модуль оцінювання якості даних та рішень* включає:
- критерії перевірки якості даних та знань;
 - аналіз ефективності рішень щодо менеджменту ризиків;
 - критерії аналізу якості рішень (квадратичні критерії на основі функції корисності та власний квадратичний критерій).
- ✓ *Модуль менеджменту ризиків* включає:
- розрахунок реальних втрат від можливої реалізації ризику;
 - управлінські модулі: розрахунку капіталовкладень та інвестицій; компенсаційних втрат на боротьбу з ризиками; встановлення обмежень та лімітів для рівня втрат та ймовірності реалізації ризику (допустимий, критичний, катастрофічний);
 - рекомендації щодо вибору стратегії зниження ризиків.
- ✓ *Модуль адаптації* дозволяє:
- введення нових даних та знань, підвантаження нових критеріїв, обмежень;
 - біхевіоріальне оцінювання ризику (оцінювання ризику в процесі та перевірка його ступеню та рівня порівняно з початковим);
 - виконання структурно-параметричної адаптації з двома контурами адаптації та використанням відхилених заявок.

Розширена архітектура багатofункціональної ІСППР представлена на рис. 6.4. Система може бути представлена у різних модифікаціях та різній наповненості в залежності від необхідності фінансової компанії, проте саме така архітектура нового типу уособлює нову концепцію запропонованої системної методології для аналізу, оцінювання та менеджменту фінансових ризиків.



Рис. 6.4 Розширена структура інформаційної системи підтримки прийняття рішень менеджменту фінансових ризиків

БЗД ІСППР складається з бази даних, знань та бази моделей (БМ), де зберігаються кращі моделі, побудовані для вирішення задач статичного та динамічного оцінювання ризиків.

Модуль оцінювання якості даних та рішень є модулем, пов'язаним з основними модулями, виконується як на етапі передачі даних для моделювання і початкової перевірки їх якості, так і на етапі перевірки якості отриманих прогнозів та побудованих моделей, перевірки ефективності розроблених рішень щодо зниження фінансових ризиків.

Кожен модуль ІСППР є її окремим функціональним елементом, який реалізує весь набір процедур, функцій, класів для інтеграції з БЗД, іншими модулями і користувачами. Функціональна схема розширеної ІСППР є доволі складною і громіздкою для представлення, тому подання її модулів використовувалось на UML для представлення кожного блоку у вигляді моделі класу. Приклад такої моделі для блоку моделювання на основі регресійних моделей у вигляді діаграми класів наведено на рис. 6.5.

Модуль опрацювання невизначеностей включає операції: попередньої обробки та консолідації даних, заповнення пропущених/втрачених даних, оцінювання інформаційних ризиків, опрацювання невизначеностей; модуль розробки скорингової карти реалізує статистичний аналіз даних, формування структури моделі та оцінювання параметрів скорингової моделі, побудови скорингової карти оцінювання ризиків, обчислення короткострокових і середньострокових прогнозів з використанням побудованих моделей та спеціального модулю перевірки якості вхідних даних та ефективності запропонованих рішень.

Використовується модуль динамічного оцінювання на основі запропонованого динамічного та ймовірно-статистичного методів із застосуванням 3-4 типів різних динамічних моделей оцінювання ризиків, обранням кращої з них, та використанням двох алгоритмів розрахунку моменту настання ризику на основі ступеню та рівню ризику.

6.4. Приклад інформаційної технології для оцінювання ризиків і інтеграції в існуючу систему на основі мікросервісів

У додатку В детально описано задачу аналізу фінансових ризиків, яка вирішувалась для німецької компанії Shop24Direct. У цьому підрозділі зупинимось лише на описі інтеграції розроблених мікросервісів з існуючою системою та базою даних.

База даних компанії складається з таких компонентів:

- користувацька інформація (стать клієнта, дата його народження (ці характеристики мають пропуски), місце проживання (оскільки товари доставляються поштою));
- інформація щодо товару: назва, тип і характеристики товару, кількість придбаного товару;
- дата покупки.

У базі даних наявні пропущені та втрачені дані. Актуальними для компанії є задачі моделювання:

- прогнозування: оцінка продажів, прогнозування навантаження або часу простою сервера для забезпечення швидким доступом користувачів до каталогу замовлень;
- аналіз і оцінювання ризиків та їх мінімізація: вибір найбільш перспективних клієнтів для цільової розсилки (ризик обрання клієнтів, що не будуть купувати товари. Тоді втрати розраховуються як сума коштів, витрачених на такі недієві розсилки), оцінювання ймовірностей та втрат від недобросовісних клієнтів [215];
- надання рекомендацій: визначення продуктів, які з високою ймовірністю можуть бути придбані разом та розробка рекомендацій таких товарів за уподобаннями клієнтів [269];
- пошук послідовностей: аналіз вибору клієнтів під час здійснення покупок, прогнозування наступної можливої події [334];

- групування: розподіл клієнтів або товарів на певну скінченну кількість кластерів, що об'єднують схожі або пов'язані елементи, для аналізу особливостей та прогнозування спільних рис [362, 367].

Вирішувалась задача: розробити інформаційну технологію для автоматизації процесу надання рекомендації щодо супутнього/наступного товару для клієнта і прогнозування часу його повернення до інтернет-магазину.

Компанія Shop24Direct використовує власну ERP-систему для управління всіма ресурсами компанії та роботи з клієнтами. Тому задача полягає у створенні такої інформаційної технології, яка дозволить реалізувати всі необхідні задачі з моделювання та аналізу фінансових ризиків у вигляді окремих мікросервісів, які власне і будуть інтегруватися з існуючою системою.

ІТ у вигляді веб-додатку була з легкістю інтегрована в систему планування ресурсів підприємства компанії, адже інтеграція мікросервісу жодним чином не зачіпає існуючі системи. Загальна схема архітектури мікросервісів підприємства представлена на рис. 6.6.



Рис. 6.6 Схема архітектури мікросервісів компанії Shop24Direct

Складові частини ІТ, яка розроблялася для компанії, представляють набір наступних компонент:

- модулі попередньої обробки вхідних даних з вбудованим комбінованим методом відновлення втрачених даних щодо клієнтів;
- модулі статистичного аналізу даних та розробки статистичних моделей-кандидаток з використанням розробленого динамічного методу для прогнозування часу наступного замовлення;
- модулі серіалізованої моделі для розгортання у веб-додатку;
- веб-додаток, який обирає кращу ймовірно-статичну модель та надає їй можливість виконувати сервіс-запити по мережі в рамках протоколу JSON.

Всі компоненти виконані на мові Python [257] із застосуванням спеціальних пакетів для наукових та інших задач [205]. Аналіз, моделювання та тестування виконані за допомогою програмного середовища JupyterLab. Доцільність вибору мови та середовища програмування зрозуміла: необхідно було інтегруватись в існуючу систему, а тому доцільно, щоб і моделювання, і імплементація здійснювались в одному програмному середовищі. Були використані наступні бібліотеки як допоміжні: `implicit`, `pandas`, `numpy`, `iciry`, `implicit.evaluation`, `matplotlib`, `seaborn`, `sklearn`, `random`, `pymsql`, `pickle`.

Модель серіалізується за допомогою пакету `pickle`, далі може бути десеріалізована безпосередньо у веб-додатку.

Веб-додаток використовує фреймворк `flask`, за допомогою якого додаток розгортається та інтегрується в систему сервісів компанії. Процес розгортання інформаційної технології представлений на рис. 6.7.

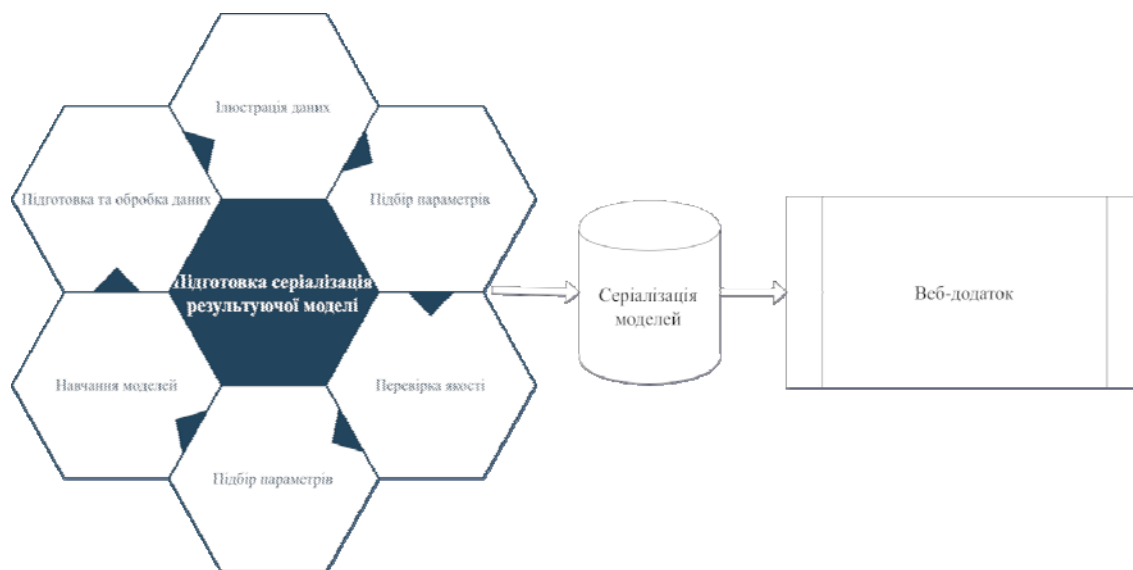


Рис. 6.7 Процес розгортання інформаційної технології

6.5. Інформаційна технологія опрацювання інформаційних ризиків в системі фінансового ризик-менеджменту підприємства

З урахуванням описаних вище особливостей фінансових даних і процесів у різних фінансових системах та існуючих інформаційних ризиків було запропоновано створення системи фінансового ризик-менеджменту [112]. Зокрема, окрім моделей для аналізу різноманітних фінансових ризиків, створено модулі обробки зовнішніх інформаційних ризиків та інформаційних впливів, які супроводжують появу фінансових ризиків (і часто є причиною їх появи), а тому мають бути розраховані за критерієм I_f , описаним у третьому розділі. Також система враховує вплив внутрішніх фінансових ризиків [112], пов'язаних з особливостями і складнощами роботи компанії (завантаженістю складських приміщень, проблемами на виробництві, малими потужностями, роботою персоналу), помилковістю раніше прийнятих управлінських рішень, короткостроковим плануванням тощо, і зовнішніх фінансових ризиків, пов'язаних з коливанням фінансового становища в країні, зміною попиту та пропозиції, діями конкурентів, коливаннями курсів валют у світі тощо. Система менеджменту фінансового ризику оцінює виявлені фінансові та інформаційні ризики і використовується для прогнозування і мінімізації втрат (рис. 6.8).

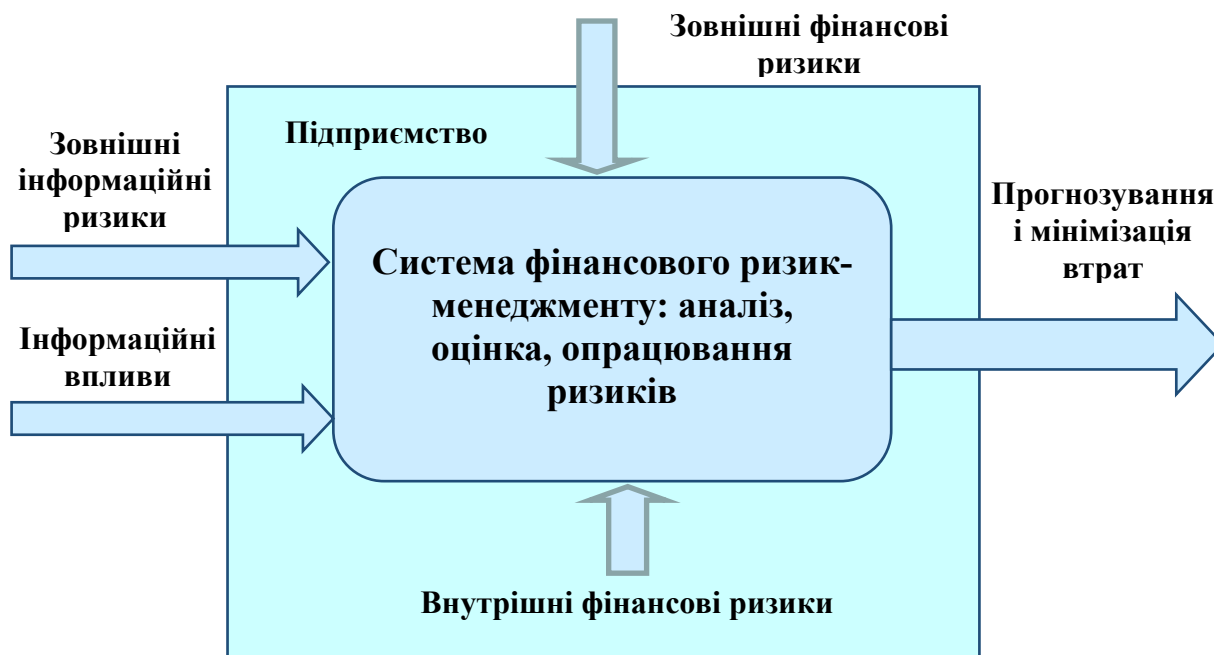


Рис. 6.8 Інформаційна технологія системи фінансового ризик-менеджменту

Функціональність системи фінансового ризик-менеджменту включає наступні обчислювальні операції: попередню обробку даних; статистичний аналіз, спрямований на підготовку даних для визначення структури моделі та оцінки параметрів, обчислення короткострокових і середньострокових прогнозів з використанням сконструйованих функцій прогнозування і підсистеми аналізу якості, яка будується на трьох наборах статистичних критеріїв якості. Критерії допомагають оцінити якість даних, адекватність моделі і якість прогнозів. Кожен раз виділяються окремі набори статистичних параметрів, спрямовані на вирішення конкретних завдань моніторингу для обчислювальних процесів. Такий системний підхід спрямований на поліпшення якості для всіх обчислень в системі. Враховується необхідність застосування різних видів критеріїв при аналізі лінійних або нелінійних моделей [112]. Наприклад, нелінійні моделі класифікації вимагають застосування індексу GINI, загальної точності, площі під кривою (AUC) і т. ін [270].

Постійний моніторинг та удосконалення процесів ризик-менеджменту виконується в єдиній інформаційній системі, що дозволяє прискорити виявлення ризикових процесів і знизити ймовірність можливих помилок.

Запропонована структура та методологія аналізу і оцінювання інформаційних ризиків у складі фінансових ризиків впроваджена автором в єдиній інформаційній технології як окремий модуль, реалізована на платформі SAS-технологій [174, 337, 338], підтвердила доцільність урахування інформаційних ризиків.

Урахування інформаційного ризику при розробці системи фінансового ризик-менеджменту шляхом оцінки якості вхідних даних, коректності моделювання та правильності прийнятих рішень з використанням різних критеріїв дозволяє здійснити процедуру аналізу та оцінювання як інформаційних, так і фінансових ризиків багаторазово, і покращити якість прийнятих рішень, якщо значення критерію буде недостатньо високим. Таким чином забезпечується процес адаптації системи фінансового ризик-менеджменту до реальної ситуації, урахування нової інформації та зовнішніх збурень, зміни політики фінансової установи щодо прийнятих рішень, наприклад, від консервативної до агресивної, і навпаки, а також урахування нових видів фінансових ризиків, які не моделювалися до цього.

6.6. Інформаційна система підтримки прийняття рішень для аналізу інвестиційних ризиків на фінансовому ринку

Для оцінювання обсягу можливих втрат (як очікуваних, так і неочікуваних) та резервного капіталу хедж-фонду [208] для покриття ризиків клієнтів було розроблено систему підтримки прийняття рішень, детально описану в авторській роботі [124] та додатку Б.1. Основною відмінністю та особливістю запропонованої СППР для аналізу фінансових ризиків хедж-фонду була можливість прогнозувати час настання можливих втрат за портфелем інвестицій для інвесторів, що вклали гроші у певний пакет акцій.

Передбачалось, що інвестиційний портфель складається з акцій n_1, \dots, n_k . Для кожної складової n_1, \dots, n_k є історія зміни цін за деякий період і потрібно побудувати модель, яка описуватиме портфель та оцінюватиме вартість

окремих складових на майбутнє за деякими описовими параметрами. Для розв'язання цієї задачі використовувалась множина моделей на основі методів ІАД та розроблених у дисертації. Задача розв'язувалась як з точки зору прогнозування ціни на акції на 1, 2, 3 і т.д. кроків уперед, так і прогнозування самої тенденції зміни ціни (зростання або падіння) [124].

Даними для моделювання зміни тенденції на ринку акцій були обрані щоденні звіти по абсолютних та процентних змінах цін закриття акцій відносно цін їх закриття попереднього дня (взято всі робочі дні з 3 січня 2006 року по 27 жовтня 2017 року). Дані обрано для 100 великих за капіталізацією компаній, які входять в S&P100 [124]. Оскільки ціни акцій суттєво відрізняються для різних компаній, то для розробки моделей використано нормовані значення.

На основі створених математичних моделей і з використанням принципів системного аналізу [76] розроблено програмний продукт на С# для реалізації ІСППР, яка дозволяє обчислювати необхідну величину резервного капіталу для покриття втрат з вказаною ймовірністю впродовж розрахованого періоду до моменту часу настання ризику. Передбачена можливість варіації інвестиційного портфеля шляхом варіації його складу (є можливість вибрати бажані акції) та розрахунку суми інвестицій (у доларах або в процентах). Для певного рівня інвестиційного капіталу ІСППР дозволяє прорахувати для користувача прогнозовані втрати (очікувані та неочікувані) за портфелем акцій, який також можна варіювати на певний обраний період. Результатом роботи ІСППР [124] є розрахунок втрат на основі методології VaR при визначенні суми інвестиції та кількості днів, можливо розрахувати час настання ризику, якщо задано значення обсягів втрат, які покриваються резервами.

На рис. 6.9. показано результати роботи СППР для обчислення часу настання ризику за вказаним обсягом інвестицій та ймовірністю ризику (на рис. 6.9, а) для оцінювання очікуваних втрат та на рис. 6.9, б) для неочікуваних втрат).

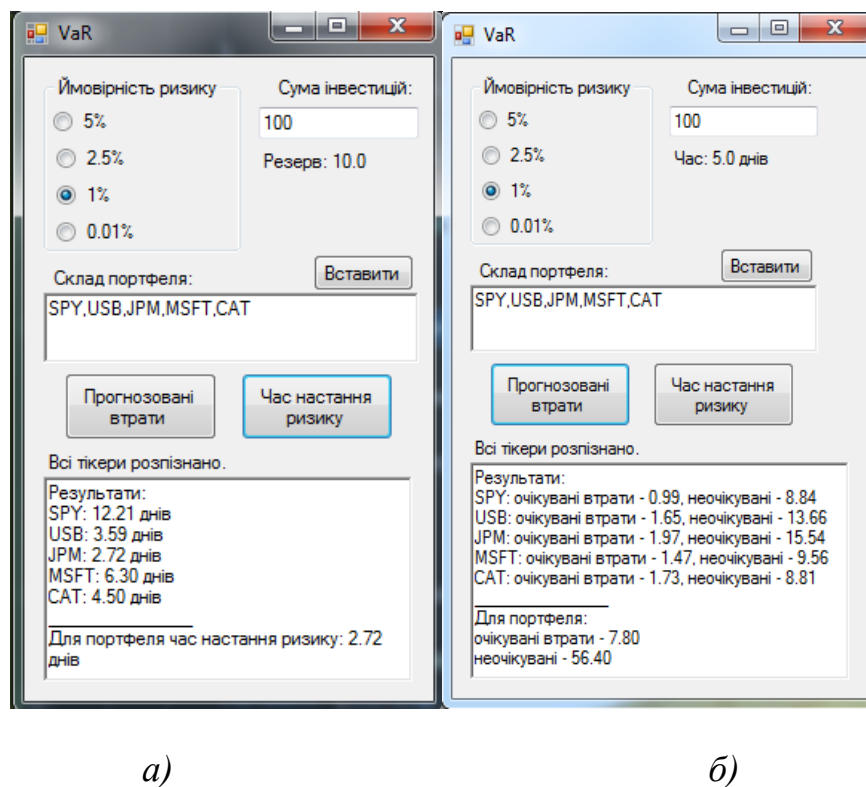


Рис. 6.9 Елементи ІСППР: а) для розрахунку кількості днів до настання ризику; б) для оцінювання очікуваних та неочікуваних втрат для портфеля

Було отримано, що JPM є акцією, що зазнає найбільших коливань, тому час настання ризику для неї є найменшим, і вона є найбільш ризиковою. Для визначення критичного часу настання ризику обчислюється мінімальне з усіх періодів акцій для компаній у портфелі [124]. Якщо сумарні інвестиції у портфель акцій становлять 100 доларів, то з ймовірністю 99% можна стверджувати, що втрати за портфелем не перевищать 10 доларів впродовж наступних 2,72 днів. За припущенням про рівномірний розподіл акцій у портфелі втрати по кожній компанії не перевищать 2 доларів впродовж знайденого періоду.

Також можна через розраховану кількість днів поступово виходити з позиції, при цьому розрахункові втрати будуть знаходитись у визначеному ліміті. Для того щоб довше знаходитися в позиції, можна змінювати також баланс лімітів по всіх компаніях: для більш волатильної надавати більший ліміт, для менш волатильної – менший [124].

Ще одна функція ІСППР передбачає оцінювання очікуваних і неочікуваних втрат. Для кожної акції можна обрати дані за певний часовий інтервал з вибірки (можна взяти за весь період впродовж 10 років або впродовж останніх 30 днів) і здійснити оцінювання волатильності та VaR для кожної акції. Для того, щоб обчислити очікувані і неочікувані втрати для портфеля, необхідно додати отримані втрати для кожного тикера у портфелі. Для сформованого портфеля SPY, USB, JPM, MSFT, CAT отримано значення, показані на рис. 6.9, б) [124].

Розроблена за описаними у третьому розділі принципами система підтримки прийняття рішень для оцінювання інвестиційних ризиків є універсальним інструментом для моделювання інвестиційних ризиків за методологією VaR. Така ІСППР за рахунок гнучкого використання множини існуючих та розроблених методів та моделей дозволяє прогнозувати очікувані та неочікувані втрати за довільним портфелем акцій [124]. Особливістю є також можливість розрахувати ймовірність та час, впродовж якого інвестиції за кожною акцією та портфелем в цілому є забезпеченими. У подальших дослідженнях можливе введення вагових коефіцієнтів для надання більших лімітів з метою перерозподілу акцій у портфелі і забезпечення найбільш волатильних компаній. Для розробки даної ІСППР було обрано мову програмування C# та середовище розробки програмного забезпечення Microsoft VisualStudio 2015.

6.7. Розробка і реалізація ІТ для динамічного оцінювання ризиків відтоку клієнтів

Постановка задачі динамічного оцінювання ризиків здійснювалась у четвертому розділі при розробці власного динамічного методу та експериментальному дослідженні його на прикладах. Задача є актуальною для усіх сфер застосування, оскільки вирішується саме задача утримання клієнта та визначення моменту його потенційного відтоку [336, 353, 364, 368]. Для

телекомунікаційної сфери та онлайн гемблінгу розроблені моделі, описані у додатку В. Розглянемо відповідну архітектуру ІТ та елементи реалізації описаних методів, моделей у вигляді окремих додатків.

Для компанії, яка здійснює забезпечення онлайн-гри та підтримку користувачів, необхідно оцінити ризик втрати клієнта та прогнозування моменту, коли він може бути втрачений, та коли він може повернутися до гри [298, 310, 336, 344]. Статистичні дані містять 150 тис. спостережень за діями гравців за 42 характеристиками, що характеризують саму гру (Казино/Покер), дані клієнтів, стратегії гри, та їх фінансову стабільність (суми ставок, виграші та програші, кількість ставок протягом доби)[130].

Ця інформація зберігається в базі знань та даних, на основі PostgreSQL. За запитами, написаними на SQL, отримують доступ до вкладених функцій та процедур обробки та коригування даних. Для безпосереднього аналізу даних використовуються модулі побудови моделей (динамічних та статичних) та модуль оцінки параметрів, реалізовані на мові R [130, 372]. Всі розроблені модулі зберігаються на сервері, поєднуються з БЗД, коригуючи дані та зберігаючи кращі моделі для оцінювання користувачів, результати представляються користувачу у вигляді таблиць, графіків, дерев кластеризації тощо (рис. 6.10).

Структура вітрин для представлення активності користувачів, оновлення його даних та статистичних даних гри протягом дня наведена на рис. 6.11.

Для щоденного оновлення даних клієнтів, які використовуються при створенні моделі аналізу клієнтів, використовується створена вітрина `scoring_model_view` (у додатку Г). Для наочного представлення та оцінювання ймовірності відтоку клієнтів використовується вітрина `vip_score_view`. Це фактично таблиця, що містить характеристики: *user_id* – ідентифікатор користувача, *type* – тип гравця: казино, покер або обидва, *score* – вектор розрахованої ймовірності відтоку, *is_bad* – ознака чи клієнт перестане грати в одну з ігор.

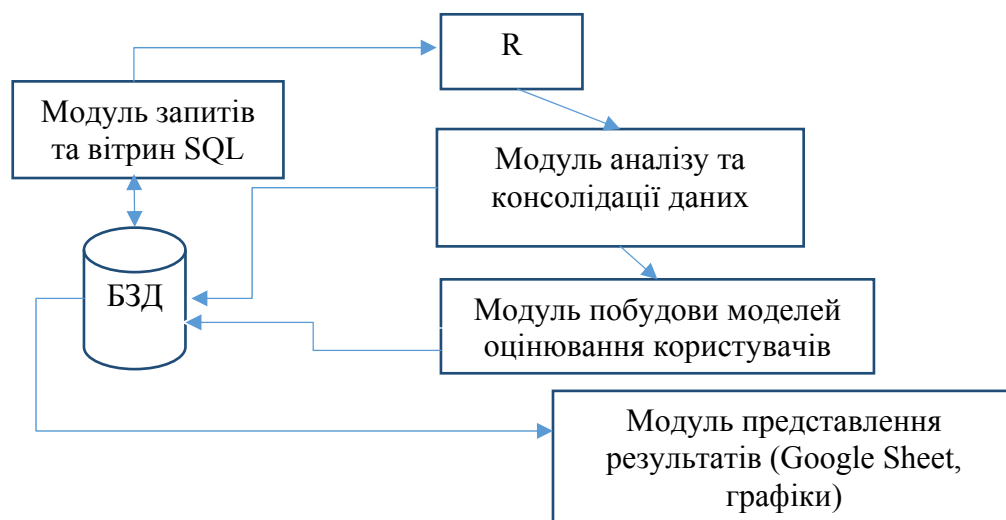


Рис. 6.10 Функціональна схема взаємодії елементів СПДР

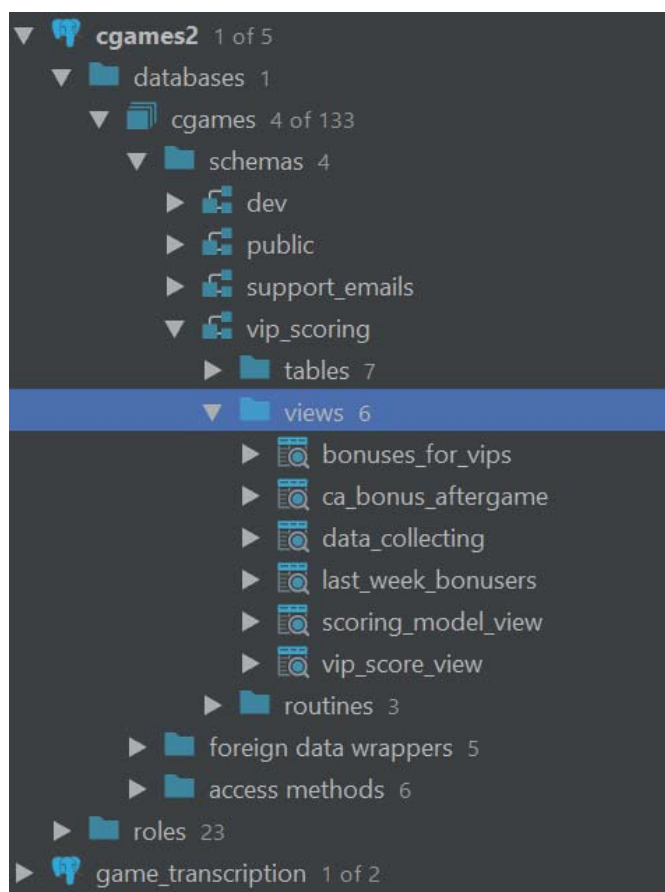


Рис. 6.11 Структура вітрин на сервері SQL

Для аналізу та консолідації даних клієнтів та представлення їх у зручному форматі у відповідному модулі використовується запропонований автором метод заповнення втрачених або пропущених даних і бібліотеки для попередньої обробки даних та маніпуляцій із ними: *dplyr*, *reshape2*, *zoo*.

Модуль побудови моделей оцінювання параметрів користувачів складається з двох блоків: для побудови різних моделей `fit_models.R` та безпосередньо оцінювання ймовірності їх відтоку `eval_score.R` (у додатку Г). Для побудови моделей використовуються стандартні бібліотеки R [130, 372], а також бібліотеки: *RODBC* – для підключення до БЗД; *ggplot2* – основна бібліотека, яка використовується для візуалізації; *survival*, *survminer* – бібліотеки для розробки динамічних моделей виживання та їх візуалізації на основі динамічного методу; *xgboost* [246], *ROCR* – бібліотеки, які використовуються для побудови ансамблів дерев та оцінки їх якості на основі критеріїв ROC та GINI.

Модуль забезпечує побудову множини динамічних моделей [115] та моделей ансамблів дерев, обрання кращої з них на основі статистичних критеріїв, збереження моделей у БЗД та за необхідності їх адаптації на основі структурно-параметричного методу, запропонованого у п'ятому розділі дисертаційної роботи. Модуль дозволяє оцінювати ймовірність настання ризику відтоку кожного окремого клієнту, а також групи клієнтів з певними ознаками, прогнозувати очікувані втрати від такого відтоку на основі побудованих моделей.

Для представлення результатів користувачам у зручному та зрозумілому вигляді був створений модуль, який дозволяє звести результати аналізу у вигляді веб-інтерфейсу таблиці у Google Sheet (рис. 6.12), яка оновлюється в реальному часі та доступна лише обмеженому колу користувачів з різним рівнем доступу (для читання/редагування/збереження тощо).

Модульна архітектура цієї системи є простою, гнучкою і зручною для використання, в разі необхідності є можливість доопрацювання і додавання нових модулів.

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
Player ID	Prize CA Bonus + Poker	Prize CA Bonus Casino	Prize Wagered Bonus	Prize Free Spins	Casino Game - #1	Casino Game - #2	Casino Game - #3	Poker Rating	Casino Rating	Total Rating	Poker Rating score	Casino Rating score	Days from Last Poker Game	Days from Last Casino Game	Last CA Bonus Date	Last CA Bonus Sum	Revenue after CA Bonus	Revenue % Bonus Sum	Casino NG Last 7 days	Rate Last 7 days		
1	0.00	0.00	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	233	2018-02-08	103.28	4727.80	1	0.00	100.0%			
2	0.00	0.00	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	48	2018-02-08	296.80	1218.48	1	0.00	100.0%			
3	0.00	0.00	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	48	2018-02-08	4.00	15.00	1	0.00	100.0%			
4	0.00	0.00	0.00	0.00				Yes	Yes	Yes	87.5%	100.0%	48	2018-02-08	81.93	2054.71	1	799.33	754.34			
5	0.00	0.00	0.00	0.00	Fireworks Blast			Yes	Yes	Yes	100.0%	100.0%	48	2018-02-13	77.13	206.13	0	1448.57	742.20			
6	0.00	0.00	0.00	0.00	Big Bad Wolf	Salute Fortune	Crash vs Pines	Yes	Yes	Yes	100.0%	87.5%	83	2018-02-08	16.36	165.78	1	33.80	402.45			
7	0.00	0.00	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	100	2017-02-09	80.07	108.18	0	0.00	638.43			
8	0.00	0.00	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	48	2018-02-08	0.07	0.01	0	0.00	384.25			
9	0.00	0.00	0.00	0.00				Yes	Yes	Yes	76.7%	100.0%	48	2018-02-08	208.36	641.80	1	0.00	341.45			
10	0.00	0.00	0.00	0.00	Dragon's Family	Rainbow Ryan	Virgo Gun Thor and Loki	Yes	Yes	Yes	100.0%	100.0%	104	2018-02-10	174.24	1084.10	1	27.87	331.47			
11	0.00	0.00	0.00	0.00				Yes	Yes	Yes	88.7%	100.0%	103	2018-02-13	33.07	288.34	1	0.00	309.36			
12	0.00	0.00	0.00	0.00	Fluffy's	Double Dragons		Yes	Yes	Yes	100.0%	100.0%	39	2018-02-10	101.40	101.40	1	28.30	488.03			
13	0.00	0.00	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	277	2018-02-28	59.86	760.80	1	0.00	487.03			
14	0.00	0.00	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	90	2018-02-13	1.04	0.00	0	0.00	141.62			
15	0.00	0.00	0.00	0.00	Lucky Lady's Charm	The Money Game	Jack and The Beanstalk	Yes	Yes	Yes	91.7%	80.0%	53	2018-02-13	15.80	0.00	0	0.00	247.54			
16	0.00	0.00	0.00	0.00				Yes	Yes	Yes	85.0%	100.0%	47	2018-02-21	60.28	446.88	1	0.00	384.87			
17	0.00	0.00	0.00	0.00				Yes	Yes	Yes	95.0%	100.0%	11	2018-02-08	60.33	414.84	1	0.00	307.86			
18	0.00	0.00	0.00	0.00				Yes	Yes	Yes	91.0%	100.0%	48	2018-02-08	2.58	64.84	1	4.78	344.23			
19	0.00	0.00	0.00	0.00	Virgo Gun Bonus	Rainbow Ryan	Sixty Sixties	Yes	Yes	Yes	72.7%	100.0%	48	2018-02-08	30.16	212.24	0	181.84	212.24			
20	0.00	0.00	0.00	0.00	Dragon			Yes	Yes	Yes	71.3%	100.0%	48	2017-02-02	50.48	806.48	1	1.20	202.40			
21	0.00	0.00	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	388	2018-02-08	46.88	665.67	1	23.43	188.20			
22	0.00	0.00	0.00	0.00				Yes	Yes	Yes	74.7%	100.0%	48	2017-02-09	16.07	889.03	1	0.22	152.29			
23	0.00	0.00	0.00	0.00				Yes	Yes	Yes	84.0%	100.0%	20	2017-02-09	17.00	364.47	0	2.78	74.84			
24	0.00	0.00	0.00	0.00				Yes	Yes	Yes	87.8%	100.0%	50	2017-02-13	31.22	29.16	0	0.00	56.34			
25	0.00	0.00	0.00	0.00	Lucky Prizes			Yes	Yes	Yes	100.0%	100.0%	308	2017-02-02	109.40	121.84	0	291.79	402.2			
26	0.00	0.00	0.00	0.00	Fluffy Wild	Fruity Wild	Backpack Circus	Yes	Yes	Yes	47.1%	100.0%	49	2018-02-08	68.78	1126.07	1	26.70	42.14			
27	0.00	0.00	0.00	42.81				Yes	Yes	Yes	100.0%	100.0%	299	2017-02-09	83.24	0.00	0	530.09	30.80			
28	0.00	0.00	0.00	0.00	Wheel Fortune	Invention Roulette		Yes	Yes	Yes	92.0%	100.0%	48	2018-02-13	82.13	81.38	1	3489.30	8.20			
29	0.00	40.41	173.30	0.00	Seven High	European Roulette Gold	Fruity N'Stars: Holiday Edition	Yes	Yes	Yes	84.8%	100.0%	53	2018-02-08	221.78	2037.45	0	1130.07	0.00			
30	0.00	19.80	138.64	0.00	The Three Musketeers	Secret of the Stones	Hidden Valley	Yes	Yes	Yes	80.0%	100.0%	48	2018-02-20	130.38	0	386.90	128.40				
31	0.00	3.84	10.80	0.00	Personal Jolly	Phantom Zeus		Yes	Yes	Yes	68.8%	100.0%	48	2018-02-08	1.62	136.80	1	77.40	48.10			
32	0.00	8000.00	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	138	2017-11-13	48.00	1385.38	1	0.00	0.00			
33	0.00	1634.90	0.00	0.00	Fluffy's	Dragon		Yes	Yes	Yes	100.0%	100.0%	156	2018-02-20	82.25	308.08	0	1959.81	751.20			
34	0.00	1943.11	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	111	2017-02-09	30.04	0.00	0	0.00	41.80			
35	0.00	1317.62	0.00	0.00	Baccarat			Yes	Yes	Yes	100.0%	100.0%	108	2017-02-08	1133.22	2088.28	1	1788.49	64.20			
36	0.00	1000.00	0.00	0.00				Yes	Yes	Yes	99.0%	100.0%	48	2017-02-08	30.49	2097.70	1	0.00	41.80			
37	0.00	869.93	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	162	2017-02-08	0.00	0.00	0	409.00	206.86			
38	0.00	869.93	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	286	2017-02-10	175.36	58.31	0	0.00	51.80			
39	0.00	869.93	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	147	2018-02-11	188.74	0.00	0	0.00	7.80			
40	0.00	782.20	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	63	2018-02-10	86.67	4.77	0	602.25	0.00			
41	0.00	688.00	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	264	2017-02-08	227.31	111.88	0	2077.33	0.00			
42	0.00	418.20	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	498	2018-02-02	378.46	347.78	0	34.80	0.00			
43	0.00	330.00	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	53	2018-02-11	205.83	282.58	0	0.00	0.00			
44	0.00	321.74	0.00	0.00				Yes	Yes	Yes	88.0%	100.0%	49	2017-02-08	10.01	4149.01	1	0.00	290.40			
45	0.00	308.81	0.00	0.00				Yes	Yes	Yes	98.9%	99.9%	51	2017-02-02	524.18	1085.41	1	21.42	0.00			
46	0.00	437.34	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	249	2018-02-08	0.00	0.00	0	2188.89	0.00			
47	0.00	399.33	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	58	2018-02-08	69.20	1612.70	1	17.21	8.80			
48	0.00	388.700000	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	87	2018-02-08	779.29	971.41	0	34.88	655.83			
49	0.00	379.20	0.00	0.00				Yes	Yes	Yes	100.0%	100.0%	81	2018-02-10	849.42	888.26	1	0.00	747.20			

Рис. 6.12 Представлення результатів аналізу у Google Sheets

Для отримання результатів з бази даних і знань в інтерфейс користувача розроблений невеликий скрипт на Python для оновлення даних по всім клієнтам з періодичністю 1 раз на добу. Для цього використовується процес-демон на сервері cron, що кожного дня запускає скрипти оцінювання параметрів клієнтів. Саме цей планувальник відповідальний за послідовність періодичного оновлення даних та виклик модулів побудови моделей.

Прототип архітектури у вигляді моделі «зірка» із буфером – базою даних на PostgreSQL наведений у цьому підрозділі і реалізований на відкритих безкоштовних технологіях (типу Ubuntu, PostgreSQL, R) є зручним та практичним рішенням для невеликих компаній, яким не має потреби здійснювати масштабні розрахунки та прогнозування на розподілених або хмарних ресурсах.

6.8. IT та ІСПП менеджменту ризиків на базі SAS

У другому розділі була наведена схема прийняття кредитних рішень та розробки скорингових карт на базі SAS-технологій, яка була реалізована з використанням віртуальної машини SAS 2014_DEC_93_EDU_Credit Scoring, що надає можливість користувачам з будь-яких комп'ютерів та програмних

середовищ отримати доступ через VMware [365] до віддаленої машини з необхідним SAS-функціоналом (рис. 6.13).

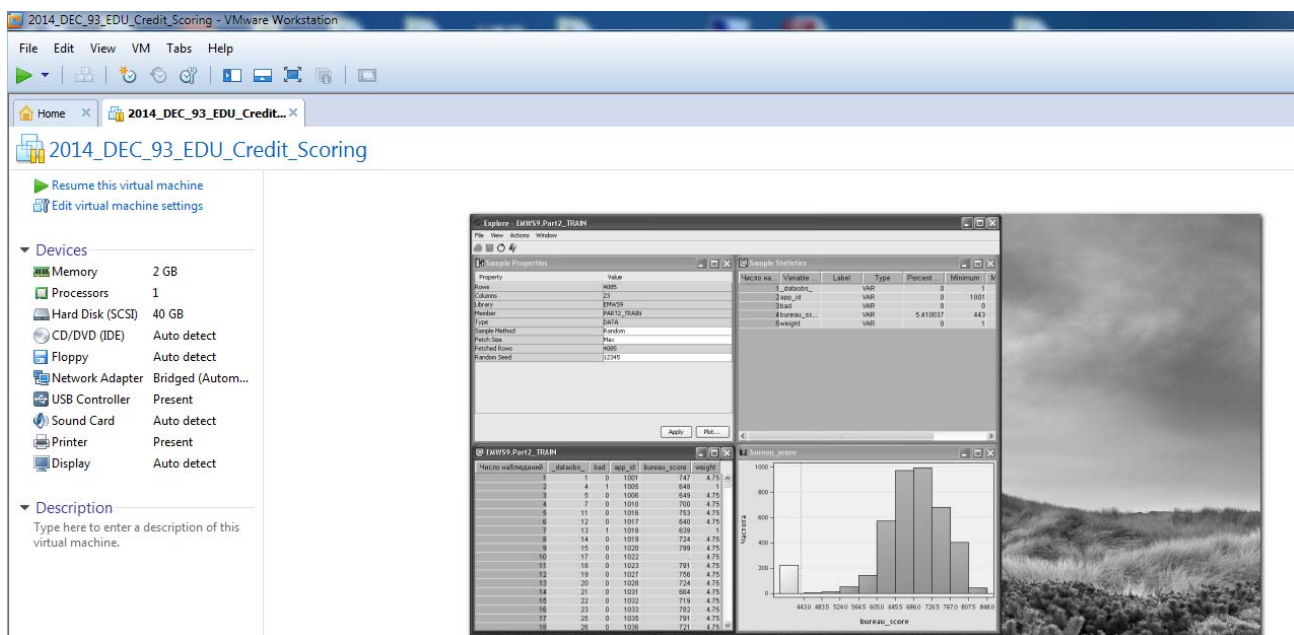


Рис. 6.13 Доступ до SAS-сервісів через віртуальну машину

Архітектура ІТ для прийняття рішень щодо кредитних заявок, аналізу і оцінювання кредитних рішень наведена на рис. 6.14 [338].

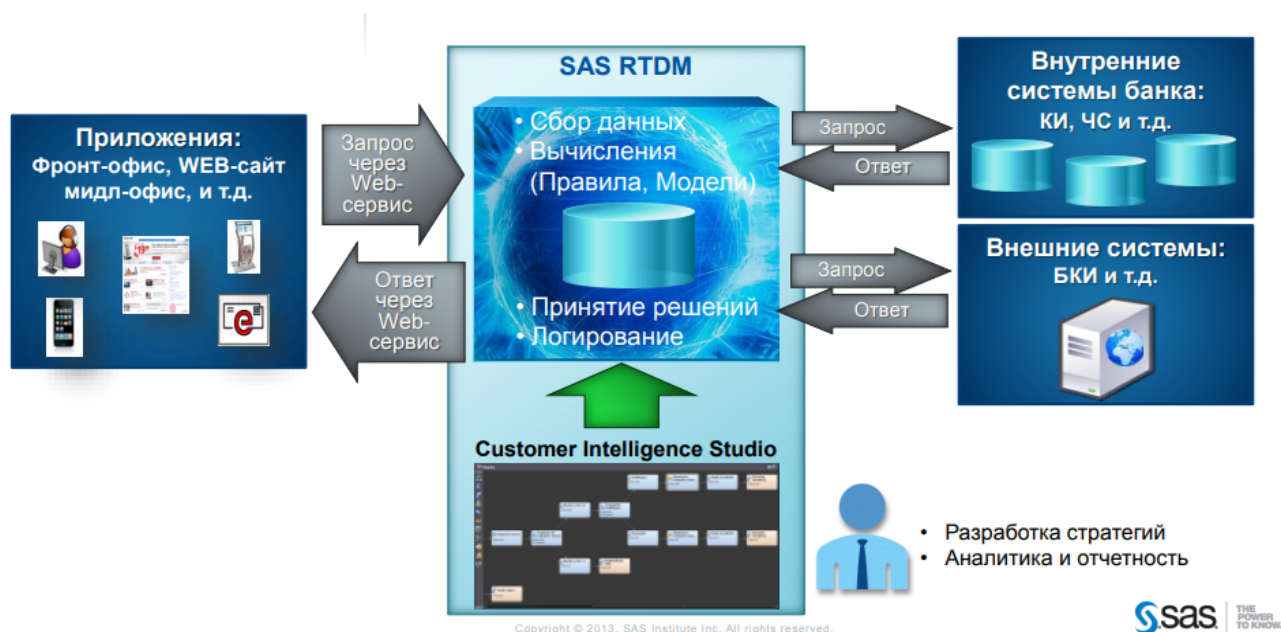


Рис. 6.14 Ілюстрація архітектура ІТ на базі SAS RTDM з сайту SAS [113]

Загальна реалізація процесу представляється у послідовності виконання бізнес-процесу: фронт-офіс (кредитні спеціалісти на фронт-енді) [342] вводять дані з анкети заявника, реєструють заявника у БД (наприклад, реалізованій на Oracle [320]), здійснюється аналіз кредитної історії (підвантажуються інформація з бюро кредитних історій та власної бази кредитних історій), здійснюється розрахунок ризик-змінних, розрахунок кредитних лімітів, побудова скорингової оцінки на основі скорингової карти, запис скорингового балу в БД та видача результатів, чи слід надавати кредит позичальнику (порівняно з порогом відсікання, встановленим на основі політики банку).

Особливості RTDM дозволяють використати всі можливості інтеграції: можливість на рівні бізнес-користувачів сформувати запит даних, різні методи інтеграції через звернення напряму до БД, веб-сервіси, черги, файли та використання множини моделей інтелектуального аналізу даних та розроблених власних моделей на основі запропонованих методів, написаних на SAS Base (детальніше у додатку Г), в реальному часі.

Висновки до шостого розділу

У розділі описано розширену інформаційну технологію, створену на основі запропонованої у дисертаційній роботі системної методології, що об'єднує методи і моделі динамічного, ймовірно-статистичного, регресійного оцінювання фінансових ризиків, комбіновані методи і моделі відновлення пропущених і втрачених даних, методи структурно-параметричної адаптації. Показано, що запропонована інформаційна технологія нового типу, яка може бути реалізована у вигляді як клієнт-серверної архітектури, так і мікросервісів та хмар, є гнучкою і адаптивною з практичної точки зору як до задач (дозволяє включати лише необхідні ресурси і блоки з оптимальним використанням серверу), так і до інтеграції в існуючу інформаційну систему функціонуючого підприємства, за рахунок розроблених власних мікросервісів,

які підключаються до системи за необхідності, реалізуючи окремі функції та підзадачі.

Існує можливість інтеграції і виконання хмарних обчислень з використанням сучасних технологій розподілених сервісів та збереження даних у «хмарах», що робить запропоновану інформаційну технологію дійсно сучасною і перспективною для подальшого використання. У додатку Г наведено акти впровадження розроблених у дисертаційній роботі інформаційних технологій на реальних підприємствах. Застосування запропонованих ІТ дозволило покращити менеджмент ризиків і отримувати точніші прогнози при оцінюванні фінансових ризиків. Інформаційні технології реалізовані у вигляді інформаційних систем. Розширена інформаційна система підтримки прийняття рішень рекомендована для реалізації і впровадження у новий банк, підприємство або фінансову систему, з можливістю виконання всіх запропонованих у розробленій системній методології етапів менеджменту фінансових ризиків. Доцільною є реалізація такої системи на базі сукупності застосувань або мікросервісів, щоб забезпечити додавання нових можливостей реалізації нових функцій та алгоритмів за потреби, у процесі функціонування системи.

Отже, результати виконання розділу підтверджують практичну доцільність і ефективність застосування запропонованих у дисертаційній роботі моделей, методів і методологічних підходів до оцінювання та менеджменту фінансових ризиків.

ВИСНОВКИ

На основі виконаних теоретичних та експериментальних досліджень у дисертаційній роботі вирішено важливу науково-прикладну проблему підвищення ефективності менеджменту ризиків фінансових систем шляхом розробки системної методології дослідження фінансових ризиків та єдиної інформаційної технології менеджменту ризиків.

Основні отримані наукові та практичні результати роботи полягають у наступному:

1. Проведено аналіз сучасного стану проблеми оцінювання ризиків фінансових систем, сутності ризиків, існуючих методів та інструментальних засобів моделювання фінансових ризиків. Складність менеджменту ризиків пов'язана з відсутністю узагальненої методології побудови інформаційних технологій для аналізу ризиків, недосконалістю моделей і методів для оцінювання та прогнозування ризиків з урахуванням неточності та неповноти як вхідних даних, так і критеріїв оцінювання. Показано доцільність розробки системної методології та інструментарію для менеджменту ризиків фінансових систем.

2. Запропоновано і розроблено системну методологію менеджменту фінансових ризиків, яка містить зокрема складові, що враховують багатофакторність ризику, засоби подолання невизначеностей, які притаманні фінансовим процесам, нові запропоновані комбіновані методи і моделі статичного і динамічного оцінювання ризиків, і дозволяє підвищити ефективність опрацювання ризиків фінансових систем. Методологію застосовано під час розробки системи моделювання, оцінювання та менеджменту фінансових ризиків.

3. Показано, що джерелом ризиків у фінансових системах можуть бути неповнота, неточність, нечіткість та недостатня якість інформації. Встановлено комплексний характер фінансового ризику і показано вплив інформаційних

ризиків на появу та якість оцінювання ФР та запропоновано критерій їх урахування як складової фінансового ризику.

4. Удосконалено методiku оцінювання фінансових ризиків на основі мереж Байєса (динамічних та гібридних), що дозволило здійснювати поглиблений аналіз причин появи невизначеностей даних, враховувати фактор часу при оцінюванні фінансових ризиків та здійснювати динамічне прогнозування ступеню та рівня фінансового ризику.

5. Розроблено комбінований метод обробки неповних даних при моделюванні фінансових ризиків, який дає можливість здійснити глибинний аналіз повноти, інформативності та якості даних, проаналізувати причини появи пропусків за допомогою мереж Байєса та відновити втрачені дані на основі застосування існуючих методів заповнення пропусків, регресійного моделювання та ЕМ-оцінювання.

6. Запропоновано нейро-нечіткий метод доповнення вибірки відхиленими заявками при моделюванні фінансових ризиків, який передбачає побудову нейронної мережі на навчальній вибірці для оцінювання відхилених заявок і доповнення вибірки при моделюванні фінансових ризиків. Розроблений метод разом зі способом урахування інформаційної складової був включений у методологію застосування скорингової карти як інструменту оцінювання і прогнозування фінансових ризиків та забезпечує підвищення якості оцінок можливих втрат.

7. У системній методології запропоновано і розроблено ймовірнісно-статистичний метод прогнозування ризику фінансових втрат, який дозволяє опрацьовувати неточність та неповноту вхідних даних за рахунок фільтрації даних, прогнозувати одночасно рівень ризику за допомогою регресійних моделей та ймовірність його виникнення в часі завдяки використанню динамічних мереж Байєса.

8. Показано доцільність динамічного аналізу ризиків та розроблено принцип і метод динамічного оцінювання та прогнозування ФР, які

передбачають не лише урахування ступеня та рівня ризику у фінансовому ризик-менеджменті, а й фактору часу як моменту переходу на вищий ступінь або рівень ризику. Метод ґрунтується на побудові різних динамічних моделей: параметричних, напівпараметричних та непараметричних, передбачена можливість формування різних моделей для різних страт та розроблено алгоритми визначення критичного часу як на основі допустимого рівня втрат для фінансової системи, так і на основі ймовірності.

9. Узагальнено існуючі критерії оцінювання якості прогнозування та прийнятих рішень і запропоновано критерій якості опрацювання ризику з урахуванням витрачених інвестицій та компенсаційних витрат на подолання невизначеностей та протидію зовнішнім впливам.

10. Запропоновано принцип адаптивного менеджменту ризиків з використанням поведінкових моделей та функцій виживання, який надає можливість здійснювати налаштування параметрів та структури моделей відповідно до зміни навколишнього середовища, надходження нових даних та параметрів. Розроблено новий метод структурно-параметричної адаптації, який відрізняється урахуванням можливих структурних і параметричних невизначеностей моделей і дозволяє здійснювати аналіз фінансових ризиків в реальному часі та коригувати скорингову модель і скорингову карту ризику, ураховувати нові параметри та змінювати структуру моделі.

11. Запропоновано і розроблено структуру інформаційної технології та функціональний склад розширеної ІСППР нового типу, що відрізняються комплексним застосуванням системного підходу до розробки архітектури, функціональної схеми та процедур аналізу даних, запропонованих ймовірнісно-статистичного, комбінованого, динамічного та структурно-параметричного методів моделювання, і забезпечують можливість генерування множини взаємодоповнюючих альтернатив та об'єктивного вибору кращої з них за допомогою кількох множин статистичних критеріїв якості даних, моделей та прогнозів. Запропонована ІСППР передбачає динамічне оцінювання і

прогнозування ризиків і забезпечує адаптацію і побудову нових адекватних математичних моделей фінансових ризиків з урахуванням нових даних, критеріїв, вимог та уточнень.

12. На основі запропонованих методів, моделей і технологій аналізу фінансових ризиків розроблено інструментальні засоби для розв'язання прикладних задач аналізу та менеджменту фінансових ризиків у вигляді інформаційних технологій та ІСППР (клієнт-серверної архітектури, з використанням мікросервісів та хмар). Використання технологій розподілених сервісів та збереження даних у «хмарах» робить такі технології гнучкими і «легкими» завдяки виконанню хмарних обчислень для великих обсягів даних. Застосування мікросервісів, які підключаються до системи за необхідності, дозволяє адаптувати такі технології для використання у вже існуючих системах.

13. Напрацьована системна методологія менеджменту фінансових ризиків та запропоновані методи, принципи і моделі оцінювання ризиків отримали впровадження у таких установах:

- Національному Банку України при розробці методології оцінювання та аудиту комерційних банків;

- компанії ТОВ «САС Інстїтют ЕЛ.ЕЛ.СІ.» при розробці нових інформаційних технологій та інструментальних рішень для банківської сфери з можливістю біхевіоріального скорингу та динамічного прогнозування фінансових ризиків;

- компанії ТОВ «ЕЛСІКО» при аналізі тендерних закупівель та поведінки учасників торгів на онлайн-платформі ProZorro;

- телекомунікаційній компанії ТОВ «Лайфселл» при прогнозуванні відтоку абонентів та можливих втрат, а також прогнозуванні витрат абонентів при користування мобільними послугами за кордоном, що дало можливість зменшити втрати компанії на 12%;

– іноземній компанії Artcom Venture GmbH при аналізі поведінки клієнтів, що дозволило на 7% підвищити якість прогнозів та збільшити обсяги продажів;

– на кафедрі математичних методів системного аналізу ІПСА НТУУ «КПІ ім. Ігоря Сікорського» при викладанні навчальних курсів «Системний аналіз та прогнозування ризиків», «Прикладні методи прогнозування», «Системний аналіз фінансових ризиків методами штучного інтелекту», «Управління ризиками на фінансових ринках», «Ризик-менеджмент методами інтелектуального аналізу даних», «Системи і методи підтримки прийняття рішень».

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Ажмухамедов И. М., Выборнова О.Н. Формализация понятий приемлемого и толерантного риска. *Инженерный вестник Дона*. 2015. № 3. URL: ivdon.ru/ru/magazine/archive/n3y2015/3240 (дата звернення 25.09.2018).
2. Александрова М. М. Страхування: навч. посіб. К.:ЦУЛ, 2002. 208с. Аналитика, бизнес-аналитика и управление данными: SAS. URL: https://www.sas.com/ru_ru/home.html (дата звернення 25.09.2018).
3. Алмазов С. .NET Core: как работают микросервисы в контейнерах. URL: <https://dou.ua/lenta/articles/microservices-net-core/> (дата звернення 25.09.2018).
4. Андерсон Т. Статистический анализ временных рядов. М.: Мир, 1976. 756с.
5. Аттетков А. В., Галкин С. В., Зарубин В. С. Методы оптимизации. Москва: МГТУ им.Баумана, 2003. 440с.
6. Балабанов И. Т. Риск – менеджмент. М.:Финансы и статистика, 1996. 188 с.
7. Балабанов И. Т. Основы финансового менеджмента. Как управлять капиталом? М.: Финансы и статистика, 1995. 384 с.
8. Бабенко В. Г. Страхування фінансових ризиків як механізм надання гарантій суб'єктам підприємницької діяльності. Суми: Вид-во Укр. акад. банк. справи Нац. банку України, 2007. 21 с.
9. Балашова Н. Управление операционным риском – анализ современных тенденций. *Рынок ценных бумаг*. 2001. № 7 (190). С. 68–70.
10. Башкіров О. В. Порівняльний аналіз VAR-методів оцінки ризику активів банку. *Проблеми і перспективи розвитку банківської системи України*: зб. наук. праць. Суми: УАБС НБУ, 2005. Вип. 14. С. 302 – 309.
11. Белевець М. С., Бідюк П. І., Кузнєцова Н. В. Розробка системного підходу до менеджменту ризиків. *Системні науки та кібернетика*. 2015. №1. С. 30

- 48. URL: http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc_1_2015.pdf (дата звернення 25.09.2018).
12. Бідюк П. І., Баклан І. В. Системний підхід к построению регрессионной модели по временным рядам. *Системні дослідження та інформаційні технології*. 2002. № 3. с. 110–135.
 13. Бідюк П. І., Баклан Я. І. Модифікований метод формування структури мережі Байєса. *Наукові вісті НТУУ «КПІ»*. 2008. № 1. С.18-27.
 14. Бідюк П. І., Гожий О. П., Коршевніук Л. О. Комп'ютерні системи підтримки прийняття рішень : навч. посіб.: Нац. техн. ун-т України «Київ. політехн. ін-т», Ін-т приклад. систем. аналізу, Чорномор. держ. ун-т ім. Петра Могили. Миколаїв : Вид-во ЧДУ ім. Петра Могили; К., 2012. 379 с.
 15. Бідюк П. І., Коновалюк М. М. Визначення величини ризику VaR на основі оцінок параметрів моделі стохастичної волатильності. *Системні дослідження та інформаційні технології*. 2012. № 3. С. 85 – 94.
 16. Бідюк П. І., Коновалюк М. М. Оцінювання параметрів моделей стохастичної волатильності з використанням алгоритму Гіббса. *Регіональний міжвузівський збірник наукових праць «Системні технології»*. Дніпропетровськ, 2011. Вип. 6(77). С. 12-27.
 17. Бідюк П. І., Коршевніук Л. О., Кузнєцова Н. В. Моделі і методи прикладної статистики: навч. посіб. з грифом МОН України. Київ: НТУУ «КПІ», 2014. 722 с.
 18. Бідюк П. І., Коршевніук Л. О. Система для оцінювання і прогнозування стану підприємства на основі мереж Байєса. *Наукові праці : Наук.- мет. журнал*. Миколаїв: Вид-во ЧДУ ім. Петра Могили.2010. Вип. 121. Т. 134 С.60–74.
 19. Бідюк П. І., Кузнєцова Н. В. Ймовірно-статистичний метод оцінювання ризику фінансових втрат. *KPI Science News*. 2018. №2. С. 7–17.

20. Бідюк П. І., Кузнецова Н. В. Наближені методи формування ймовірнісного висновку та їх модифікація для гібридних мереж Байєса *Наукові праці. Серія: Комп'ютерні технології*. Миколаїв, 2009. Вип. 104. т. 117. С. 17–30.
21. Бідюк П. І., Кузнецова Н. В. Основні етапи побудови і приклади застосування мереж Байєса. *Системні дослідження та інформаційні технології*. 2007. № 4. С. 26–39.
22. Бидюк П. И., Кузнецова Н. В. Прогнозирование волатильности финансовых процессов с помощью моделей условной дисперсии. *Проблемы управления и информатики*. 2014. № 5. С. 47–54.
23. Бідюк П. І., Кузнецова Н. В., Терентьев О. М. Система підтримки прийняття рішень для аналізу фінансових даних. *Наукові вісті НТУУ "КПІ"*. 2011. №1. С. 48–61.
24. Бідюк П. І., Матрос Є. О. Моделі оцінки ризиків кредитування фізичних осіб. *Кібернетика та обчислювальна техніка*. 2007. №153. С. 87–95.
25. Бідюк П.І., Меньяйленко О.С., Половцев О.В. Методи прогнозування. Луганськ: Альма-Матер, 2008. 608 с.
26. Бідюк П.І., Романенко В.Д., Тимошук О.Л. Аналіз часових рядів: навч. посіб. Київ: НТУУ «КПІ», 2013. 600 с.
27. Бланк И.А. Управление финансовой стабилизацией предприятия. К.: Вид-во "Ніка-Центр", 2003. 496 с.
28. Бланк І. А. Управління фінансовими ризиками. К.: Ніка-Центр, 2014. 600 с.
29. Бобиль В. В. Фінансові ризики банків: теорія та практика управління в умовах кризи: монографія. Дніпропетровськ: Дніпропетр. нац. ун-т залізн. трансп. ім. акад. В Лазаряна, 2016. 298 с.
30. Боди З., Мертон Р. Финансы. М.: «Вильямс», 2000. 592 с.
31. Бригхем Ю., Гапенски Л. Финансовый менеджмент. Полный курс: в 2-х т. / пер. с англ. под ред. В.В. Ковалева. СПб.: «Экономическая школа», 1997.
32. Буднік М. М., Мартюшева Л. С., Сабліна Н. В. Фінансовий ринок: навч. посіб. К.: Центр навчальної літератури, 2009. 334 с.

33. Бучко І. Є. Скоринг як метод зниження кредитного ризику банку. *Вісник Університету банківської справи Національного банку України*. 2013. Вип. 2 (17). С. 178–182.
34. Бююль А., Цёфелью П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей: пер. с нем. Спб.: ООО «ДиаСофтЮП», 2005. 608 с.
35. Васильев В. А., Летчиков А. В., Лялин В. Е. Математические модели оценки и управления финансовыми рисками хозяйствующих субъектов. *Аудит и финансовый анализ*. 2006. № 4. С. 200–237.
36. Великоіваненко Г. І., Трокоз Л. О. Нейро-нечітка модель оцінювання прострочених позик комерційного банку. *Нейро-нечіткі технології моделювання в економіці*. 2014. №3. С. 23-66. URL: http://nbuv.gov.ua/UJRN/Nntm_2014_3_4 (дата звернення 25.09.2018).
37. Верхуша Н. П. Сутність кредитного ризику банку та його фактори. URL:http://lib.uabs.edu.ua/library/Visnik/Numbers/2_31_2011/31_03_04.pdf (дата звернення 25.09.2018).
38. Вітлінський В. В., Великоіваненко Г. І. Ризикологія в економіці та підприємстві: монографія. Київ: КНЕУ, 2004. 480 с.
39. Вітлінський В. В., Верченко П. І. Аналіз, моделювання та управління економічним ризиком: навч. посіб. К.: КНЕУ, 2000. 292 с.
40. Вітлінський В. В. Концептуальні засади ризикології у фінансовій діяльності. *Фінанси України*. 2009. № 3. С. 3–9.
41. Вітлінський В. В., Наконечний С. І. Ризик у менеджменті. К.: Борисфен-М, 1996. 326 с.
42. Вовчак О. Д. Страхова справа: підручник. К.: Знання, 2011. 391 с.
43. Воробьев С. Н., Бадин К. В. Системный анализ и управление рисками в предпринимательстве: учеб.пособие М.: Из-во Московского психолого-социального института; Воронеж: Из-во НПО «МОДЭК», 2009. 760 с.

44. Воронцов К. Математические методы обучения по прецедентам. Курс лекций. М: МФТИ, 2006. 58с.
45. Вошило М. Основи управління ризиками у банківській справі. *Вісник Національного банку України*. 2001. №12. С. 51– 52.
46. Все продукты Microsoft. URL: <https://www.microsoft.com/ru-ru/> (дата звернення 25.09.2018).
47. Вуколов В. В. Інформаційні ризики в державному управлінні. URL:http://archive.nbuv.gov.ua/e-journals/Patp/2010_2/10vvvrdu.pdf (дата звернення 25.09.2018).
48. Гланц С. Медико-биологическая статистика. Москва: Практика, 1999. 459с.
49. Говорушко Т. А. Страхові послуги: навч. посібн. К.: Центр навчальної літератури, 2005. 400 с.
50. Горячек І. Види банківських ризиків та управління ними. *Економіка. Фінанси. Право*. 2000. №8. С. 36–38
51. Грабовецький В. Є. Методи експертних оцінок: теорія, методологія, напрямки використання. Вінниця: ВНТУ. 2010. 170с.
52. Гуськова В. Г., Кузнєцова Н. В. Комбінування оцінок прогнозів, обчислених за різними методами для обраних країн світу. *Системні науки та кібернетика*. 2015. №1. С. 49 – 57. URL: http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc_1_2015.pdf (дата звернення 25.09.2018).
53. Данилов В. Я., Зінченко А. Ю., Данилов В. Я. Системний підхід до розв'язання прямих і обернених задач у системах з хаосом. *Системні дослідження та інформаційні технології*. 2017. № 2. С. 7–18. URL: http://www.ela.kpi.ua/bitstream/123456789/20944/1/SD2_01Danylov.pdf (дата звернення 25.09.2018).
54. Довгий С. О., Бідюк П. І., Трофимчук О. М. Системи підтримки прийняття рішень на основі ймовірно-статистичних методів. Київ: Логос, 2014. 419с.

55. Доклады о развитии человека (2003 – 2012 гг.). URL: <http://hdr.undp.org/en/reports/hdr2012/chapters/ru> (дата звернення 25.09.2018).
56. Долгошея Н. О. Страхування в запитаннях та відповідях : навч. посіб. К.: Центр учбової літератури, 2010. 318 с.
57. Дрейпер Н., Смит Г. Прикладной регрессионный анализ, 3-е изд.: М. Изд.дом «Вильямс», 2007. 912 с.
58. Дрозд А. О., Капустян В. О. Загальна модель ціноутворення кредитного та депозитного продуктів комерційного банку за умови випадкового запізнення при поверненні кредитів. *Бізнес Інформ*. 2016. № 7. С. 47–51.
59. Друри К. Управленческий учет для бизнес-решений. М.: ЮНИТИ-ДАНА, 2015. 656 с.
60. Дубініна С. В. Байєсівські методи моделювання актуарних процесів та оцінювання ризиків страхових компаній: дис. ... канд. техн. наук : 05.13.23 Київ, 2017. 199 с.
61. Дудка Б.Р., Бідюк П.І. Реалізація методики побудови моделей часових рядів. *Системний аналіз та кібернетика*. 2016. С.233-255. URL:http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc_5_2016.pdf (дата звернення 25.09.2018).
62. Егоров В. А. Система управления рисками в банке. *Финансы*. 2003. №9. С.78.
63. Ермасова Н. Б. Риск-менеджмент организации. 2008. URL: http://velib.com/read_book/ermasova_natalja/risk_menedzhment_organizacii/glava_2_analiz_riska_v_deyatelnosti_firmy/21_sistemnyjj_podkhod_k_analizu_riskov_v_deyatelnosti_firmy/ (дата звернення 25.09.2018).
64. Жихор О. Б., Штехан М. О. Сутність поняття «фінансовий ризик» та його класифікація. *Науковий вісник НЛТУ України*. 2013. Вип. 23.10. С. 145–150.
65. Завгородний В. И. Парадигма информационных рисков. URL: http://www.fa-kit.ru/main_dsp.php?top_id=591 (дата звернення 25.09.2018).

66. Загірська І. О. Аналіз та прогнозування стану підприємства на основі байєсівського методу *Наукові праці Чорноморського державного університету імені Петра Могили. Сер. : Комп'ютерні технології*. 2011. Т. 160. Вип. 148. С. 27–35.
67. Загородній А. Г., Вознюк Г. Л. Фінансово-економічний словник. К.: Знання, 2007. 1072 с.
68. Загоруйко Н. Г., Ёлкина В.Н., Лбов Г.С. Алгоритмы обнаружения эмпирических закономерностей . Новосибирск: Наука. 1985. 120 с.
69. Загоруйко Н. Г. Методы распознавания и их применение. М.: Сов. радио, 1972. 216 с.
70. Загоруйко Н. Г. Прикладные методы анализа данных и знаний. Новосибирск: Изд-во института математики, 1999. 61 с.
71. Зайченко Ю. П. Основи проектування інтелектуальних систем. Київ: Слово, 2006. 352 с.
72. Закон України Про публічні закупівлі. URL:
<http://zakon2.rada.gov.ua/laws/show/922-19> (дата звернення 25.09.2018).
73. Зангиева И. К. Решение проблемы неполноты данных массовых опросов. *Российская социология завтрашнего дня: Сборник студенческих работ*. М.: Издательский дом ГУ-ВШЭ. 2008. Вып. 3. С.84–95.
74. Згуровский М. З., Подладчиков В. Н. Аналитические методы калмановской фильтрации для систем с априорной неопределенностью. К.: Наукова думка, 1995. 284 с.
75. Згуровский М. З., Бидюк П. И., Терентьев А. Н. Методы построения байесовских сетей на основе оценочных функций. *Кибернетика и системный анализ*. 2008. №2. С. 81–88.
76. Згуровский М. З., Панкратова Н. Д. Системный анализ: Проблемы. Методология. Приложения. К. : Наук. думка, 2005. 743 с.

77. Згуровський М. З., Бідюк П. І., Терентьев О. М., Просянкіна-Жарова Т. І. Байесівські мережі в системах підтримки прийняття рішень : навч. посіб. Київ: ТОВ ВП «Едельвейс», 2015. 300 с.
78. Згуровський М. З., Панкратова Н. Д. Основи системного аналізу: підруч. для студ. вищ. навч. закл. К. : Вид. група ВНУ, 2007. 544 с.
79. Зинкевич В.А., Штатов Д.Н. Информационные риски: анализ и количественная оценка (начало). *Бухгалтерия и банки*. М., 2007. №1. С. 50–55.
80. Зинкевич В. А., Штатов Д. Н. Информационные риски: анализ и количественная оценка (окончание). *Бухгалтерия и банки*. М., 2007. №3. С. 48–53.
81. Злоба Е., Яцкив И. Статистические методы восстановления пропущенных данных. *Computer Modelling & New Technologies*. 2002. Vol.6. № 1. С.51–61.
82. Зоріна О.А. Методи аналізу фінансових ризиків. *Проблеми теорії та методології бухгалтерського обліку, контролю і аналізу*. 2011. № 2(20). С. 221–229.
83. Зур'ян О.В. Економічна сутність та класифікація фінансових ризиків. *Наукові праці МАУП*. 2016. Вип. 50. С. 141–146.
84. Камінський А. Б. Моделювання фінансових ризиків. К. : Видавничо-поліграфічний центр «Київський університет», 2006. 306 с.
85. Камінський А. Б., Писанець К. К. Модель кредитного скорингу, заснована на концепції виживання. *Формування ринкової економіки в Україні*. 2012. Вип. 27. С. 136 – 142.
86. Капустян В. О., Дрозд А. О. Моделювання кредитного ризику в потоковій моделі банку *Сучасні проблеми економіки і підприємництва: зб.наук.пр.* 2010. Вип. 5 (частина II).С. 103–106.
87. Капустян В. О., Ільченко К. О. Моделювання прибутку банківської установи в умовах кризи. *Бізнес Інформ*. 2010. № 4. С. 92–95.

88. Киселев М., Соломатин Е. Средства добычи знаний в бизнесе и финансах *Открытые системы*. 1997. № 4. С. 40–44.
89. Кочетыгов А. А. Финансовая математика: Серия: Учебники, учебные пособия. Ростов на Дону: Изд-во «Феникс», 2004. 480 с.
90. Кузнєцова Н. Аналіз даних клієнтів за допомогою мереж Байєса. *Комп'ютерні науки та інженерія: мат. V міжнар. конф. молодих вчених CSE-2011*, (Львів, 24–26 листоп. 2011 р.). Вид. Львівської політехніки, 2011. С.50–53.
91. Кузнєцова Н. В. Аналіз і оцінювання інформаційних ризиків. *Системний аналіз та інформаційні технології (SAIT 2013)*: матеріали XV міжнар. наук.-техн. конф. (Київ 27-31 травня, 2013 р.). К.: ННК “ІПСА” НТУУ “КПІ”, 2013. С. 122–123.
92. Кузнєцова Н. В. Аналіз та прогнозування ризиків шахрайства з кредитними картками. *Інформатика та математичні методи в моделюванні*. 2018. Т.8, №1. С. 16–25.
93. Кузнєцова Н. В. Виявлення та оброблення невизначеностей у формі неповних даних методами інтелектуального аналізу. *Системні дослідження та інформаційні технології*. 2016. № 2. С. 104–115.
94. Кузнєцова Н. В. Деякі аспекти мінімізації інформаційних ризиків у банківській діяльності. *Системні дослідження та інформаційні технології*. 2014. № 1. С. 7–19.
95. Кузнєцова Н. В. Динамічно-рівномірна дискретизація даних у гібридних мережах Байєса. *Вісник Хмельницького Національного Університету*. 2011. №5. Т.3(180). С. 247–251.
96. Кузнєцова Н. В. Застосування мереж Байєса до оцінювання інформаційних ризиків. *Информационные технологии и безопасность. Оценка состояния: мат. междун. науч.-практ. конф. ИТБ-2013*. (Киев, 18 июня 2013г.). Вип. 13. К.: НАН України. 2013. С. 95–104.

97. Кузнєцова Н. В. Інтегровані моделі аналізу фінансових даних. *Системний аналіз та інформаційні технології (SAIT 2011)*: матеріали XIII міжнар. наук.-техн. конф. (Київ 23-28 травня, 2011 р.). К.: ННК «ІПСА» НТУУ «КПІ», 2011. С. 274.
98. Кузнєцова Н. В. Інтегрований підхід до оцінювання кредитних ризиків. *Тр. Одес. политехн. ун-та*. Одеса, 2010. Вып.1(33) – 2(34).С. 187–192.
99. Кузнєцова Н. В. Інформаційна технологія аналізу кредитоспроможності позичальника на основі мереж Байєса. *Вісник Хмельницького Національного Університету*. 2010. №5. Т.2. С. 192–196.
100. Кузнєцова Н. В. Інформаційні технології аналізу клієнтської бази абонентів та прогнозування їх поведінки. *Информационные технологии и безопасность*: мат. междунар. науч.-практ. конф. ИТБ-2017 (Київ, 30 листопада, 2017 г.). К.: ИПРИ НАН України, 2017. С. 114 – 120.
101. Кузнєцова Н. В. Інформаційні технології для аналізу фінансових зловживань на платформі ProZorro. *Наукові праці ВНТУ*: електронне наукове фахове видання. Електронні дані. Вінниця: Вінниц. Нац. тех. ун-т: 2018. №1. URL: <https://praci.vntu.edu.ua/index.php/praci/article/view/534/525> (дата звернення 25.09.2018).
102. Кузнєцова Н. В. Інформаційні технології обробки та аналізу даних у фінансовому ризик-менеджменті. *Інформаційні технології та спеціальна безпека*. 2015. №1. С. 86–98.
103. Кузнєцова Н. В. Методи оцінювання ризиків роздрібного кредитування *Системний аналіз та інформаційні технології (SAIT 2010)* : матеріали XII міжнар. наук.-техн. конф. (Київ, 25 – 29 травня 2010 р.).К.: ННК «ІПСА» НТУУ «КПІ», 2010. С. 272.
104. Кузнєцова Н. В. Методика оцінювання ризику зниження фінансової стабільності за допомогою мереж Байєса. *Вісник Хмельницького Національного Університету*. 2013. №2. Т.3 (198). С. 187–190.

105. Кузнєцова Н. В. Питання якості, конфіденційності та коректності даних в інформаційних технологіях аналізу фінансових ризиків. *Информационные технологии и безопасность: мат. междунауч.-практ. конф. ИТБ-2015* (Киев, 21 октября 2015 г.). Вып. 15. К.: ИПРИ НАН Украины, 2015. С. 131–136.
106. Кузнєцова Н. В. Практичне застосування методології розробки скорингових карт для аналізу ризиків автокредитування. *Електротехнічні та комп'ютерні системи*. 2017. № 24(100). С. 104 – 111.
107. Кузнєцова Н. В. Практичні підходи до визначення та урахування невизначеностей, що формують фінансові ризики. *Тр. Одес. политехн. ун-та*. 2014. Вып.2 (44). С. 160–170.
108. Кузнєцова Н. В. Розробка скорингових карт для аналізу ризиків банківської діяльності. *Реєстрація, зберігання і обробка даних*. 2017. Т.19. №4.С.35–44.
109. Кузнєцова Н. В. Скорингові карти для аналізу ризиків банківської діяльності. *Інформаційні технології та спеціальна безпека*. 2018. №1. С. 26–35.
110. Кузнєцова Н. В. Скорингові карти як інструмент аналізу ризиків. *Системний аналіз та інформаційні технології (SAIT 2016)* : матеріали 18-ї міжнар. наук.-техн. конф. (Київ, 25 – 29 травня 2016 р.).К.: ННК «ІПСА» НТУУ «КПІ», 2016. С. 106 –107.
111. Кузнєцова Н. В. Скорингові технології оцінювання ризиків шахрайства в банківській діяльності. *Информационные технологии и безопасность: мат. междунауч. науч.-практ. конф. ИТБ-2016*. (Киев, 1 декабря 2016 г.) К.: ИПРИ НАН Украины, 2016. С. 43 – 47.
112. Кузнєцова Н. В. Фінансовий ризик-менеджмент з урахуванням інформаційних ризиків. *Реєстрація, зберігання і обробка даних*. 2018. Т.20. №1. С. 30–39.

113. Кузнецова Н. В., Бідюк П. І. Аналіз фінансових ризиків з використанням SAS-технологій обробки даних. *Електротехнічні і комп'ютерні системи*. 2016. № 22(98). С. 267 – 271.
114. Кузнецова Н. В., Бідюк П. И. Выявление рисков мошенничества в ходе тендерных закупок методами теории выживания. *Управляющие системы и машины*. 2018. № 2 (274). С.87-96.
115. Кузнецова Н. В., Бідюк П. І. Динамічне моделювання фінансових ризиків. *Індуктивне моделювання складних систем*. 2017. Вип. 9. 2017. С. 122–137.
116. Кузнецова Н. В., Бідюк П. І. Інтегрований підхід до аналізу фінансових ризиків. *Сучасна інформатика: проблеми, досягнення та перспективи розвитку: матеріали міжнар. наук. конф. (12–13 вересня 2013 року, Київ)*. Київ: Ін-т кібернетики ім. В.М.Глушкова НАН України, 2013. С. 153–155.
117. Кузнецова Н. В., Бідюк П. І. Інформаційна система підтримки прийняття рішень на основі інтегрованого підходу для аналізу фінансового стану підприємства *Наукові праці. Серія: Комп'ютерні технології*. Миколаїв, 2010. Вип. 130. Т. 143. С. 49–56.
118. Кузнецова Н. В., Бідюк П. І. Інформаційна технологія аналізу фінансових даних на основі інтегрованого методу. *Системні дослідження та інформаційні технології*. 2011. № 1. С. 22–33.
119. Кузнецова Н. В., Бідюк П. І. Математичні моделі виживання для прогнозування фінансових ризиків. *Моделювання та прогнозування економічних процесів: матеріали XI науково-практичної конференції. (Київ, 6-8 грудня 2017р.)*. К.: КПІ ім. Ігоря Сікорського, 2017. С. 47–48.
120. Кузнецова Н. В., Бідюк П. И. Моделирование кредитных рисков на основе теории выживания. *Проблемы управления и информатики*. 2017. №6. С. 33– 46.
121. Кузнецова Н.В., Бідюк П. І. Моделювання фінансового ризику в телекомунікаційній сфері. *Наукові вісті НТУУ “КПІ”*. 2017. №5. С. 51–58.

122. Кузнєцова Н.В., Бідюк П. І. Нейронні та мережі Байєса у задачі аналізу кредитних ризиків. *Реєстрація, зберігання і обробка даних*. 2015. Т.17, №2. С. 61–71.
123. Кузнєцова Н.В., Бідюк П. І. Порівняльний аналіз характеристик моделей оцінювання ризиків кредитування. *Наукові вісті НТУУ «КПІ»*. 2010. № 1. С. 42 – 53.
124. Кузнєцова Н.В., Бідюк П. І. Система підтримки прийняття рішень для аналізу інвестиційних ризиків фінансових ринків. *Вісник Національного університету «Львівська політехніка». Серія «Інформаційні системи та мережі»*. 2018. Вип. 887. С. 115–121.
125. Кузнєцова Н. В., Бідюк П. І. Системний підхід до аналізу кредитних ризиків з використанням мереж Байєса. *Наукові вісті НТУУ «КПІ»*. 2008. №3. С. 11–24.
126. Кузнєцова Н. В., Бідюк П. І. Системний підхід до менеджменту фінансових ризиків. *Системні дослідження та інформаційні технології*. 2018. № 2. С. 124–140.
127. Кузнєцова Н. В., Бідюк П. І. Структурно-параметрична адаптація ймовірно-статистичних моделей для оцінювання фінансових ризиків *KPI Science News*. 2018. №3. С.23–34.
128. Кузнєцова Н. В., Кінда В. В. Обробка і аналіз даних обрахунку тепловтрат приміщень у реальному часі. *Системні науки та кібернетика*. 2015. №1. С. 77 – 86. URL:http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc_1_2015.pdf (дата звернення 25.09.2018).
129. Кузнєцова Н. В., Куца К.В., Штогрін С. Р. Застосування методології аналізу виживання для дослідження споживчих ризиків. *Системні науки та кібернетика*. 2017. №6. С. 126–135. URL:
http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc_6_2017.pdf (дата звернення 25.09.2018).

130. Кузнецова Н. В., Фомін О. В. Прогнозування ризику втрати користувачів онлайн-платформи. *Системний аналіз та інформаційні технології (SAIT 2018)*: матеріали 20-ї Міжн. науково-техн. конф., (Київ 21-24 травня, 2018 р.). К.: ННК «ІПСА» НТУУ «КПІ», 2018. №20. С. 139–140.
131. Кунцевич В. М., Лычак М. М. Синтез оптимальных и адаптивных систем управления. Игровой подход. Киев: Наукова думка, 1985. 248с.
132. Ладанюк А. П. Основи системного аналізу: навч. пос. Вінниця: Нова книга, 2004. 176 с.
133. Легка Я. І. Поняття та зміст інвестиційних ризиків як об'єкту управління. *Теоретичні та прикладні питання економіки: зб. наук. пр.* К.: Видавничо-поліграфічний центр "Київський університет", 2009. Вип. 20. С. 289 – 292.
134. Литтл Р. Дж. А., Рубин Д. Б. Статистический анализ данных с пропусками. М.: Финансы и статистика, 1991. 336с.
135. Лукас В. А. Теория управления техническими системами: уч. пос. Екатеринбург: Изд-во УГГУ, 2005. 677 с.
136. Масалітіна В. В. Планування руху грошових коштів в системі управління фінансовими ризиками. К.: Вид-во "Родовід", 2002. 20 с.
137. Матвійчук А., Бень В. Використання logit- та probit-регресій для оцінки кредитоспроможності позичальника. *Вісник Національного банку України*. 2015. № 5. С. 37–41.
138. Метод і модель аналізу виживання (Каплана-Мейєра). URL: https://www.ibm.com/support/knowledgecenter/en/SSLVMB_23.0.0/spss/advanced/idh_kmei.html (дата звернення 25.09.2018).
139. Методичні рекомендації щодо організації та функціонування систем ризик-менеджменту в банках України: Постанова Правління Національного банку України від 02.08.2004 р. № 361. URL: http://www.bank.gov.ua/Bank_supervision/Risks/361.pdf (дата звернення 25.09.2018).

140. Методичні вказівки з інспектування банків “Система оцінки ризиків”: вказівки, затверджені правлінням НБУ від 15.03.2004 № 104. URL: <http://zakon.nau.ua/doc/?code=v0104500-04> (дата звернення 25.09.2018).
141. Міщенко В. І., Слав’янська Н. Г., Коренева О. Г. Банківські операції: підручник. Київ: Знання, 2007. 283 с.
142. Никитин И., Цулая М. Системный подход к процессу управления рисками. URL: <http://www.bainr.ru/article25.html> (дата звернення 25.09.2018).
143. Облачные технологии для земных пользователей. URL: <http://help.starline.lg.ua/internet/oblachnie-tehnologii-dlja-zemnih-polzovatelei.html> (дата звернення 25.09.2018).
144. Павлюк О. В., Бідюк П. І., Павлюк О. О. Методика побудови динамічних мереж Байеса. *Наукові вісті НТУУ "КПІ"*. 2010. № 2. С. 55-62.
145. Паклин Н. Логистическая регрессия и ROC-анализ – математический аппарат. URL: <https://basegroup.ru/community/articles/logistic> (дата звернення 25.09.2018).
146. Панкратова Н. Д. Рациональный компромисс в системной задаче концептуальной неопределенности. *Кибернетика и системный анализ*. 2002. № 4. С.162–180.
147. Парасюк І. М., Єршов С. В., Алексеєнко О. А. Трансформаційний підхід типу «модель–модель» для реалізації Байєсівських механізмів інтелектуального аналізу даних. *Проблеми програмування*. 2006, № 2 – 3. Спеціальний випуск. С. 12–16.
148. Перепелиця В.О. Математичні моделі та методи оцінки ризиків економічних, соціальних та аграрних процесів. К.: Центр навчальної літератури. 2013. 210 с.
149. Пернарівський О. Аналіз, оцінка та способи зниження банківських ризиків. *Вісник Національного банку України*. 2004. №4. С. 44–48.
150. Петренко С., Симонов С. Методики и технологии управления информационными рисками. URL: <http://citforum.ru/security/articles/risk/>.

151. Половцев О. В. Державне управління регіональним розвитком в умовах невизначеності: аналіз підходів до прийняття рішень. *Теорія та практика державного управління і місцевого самоврядування*. 2013, №2. URL: http://nbuv.gov.ua/UJRN/Ttpdu_2013_2_11 (дата звернення 25.09.2018).
152. Попович М. Г., Ковальчук О. В. Теорія автоматичного керування. К.: Либідь, 2007. 656с. URL: <http://www.hindawi.com/journals/mpe/2013/162938/> (дата звернення 25.09.2018).
153. Про порядок формування і використання резерву для відшкодування можливих втрат за кредитними операціями банків: положення, затверджене постановою правління НБУ від 06.07.2000 № 279. URL : zakon.rada.gov.ua/cgi-bin/laws/main.cgi?nreg=v8184500-08 (дата звернення 25.09.2018).
154. Растрингін Л. А. Адаптація складних систем. Методи і приложення. Рига: Зинатне, 1981. 375 с.
155. Рзаев Р. Р., Бабаева С. Т., Бабаев Т. А. Автоматизированная информационная система комплексной оценки финансовой устойчивости коммерческих банков. *Проблемы управления и информатики*. 2017. № 3. С. 71–86.
156. Рішняк І. В. Системний аналіз категорій ризику та невизначеностей. *Вісник науц. унів. «Львівська політехніка»*. 2003. №489. С. 263–275.
157. Рогов М. А. Риск-менеджмент. М.: Финансы и статистика, 2001. 118 с.
158. Ролік О. І., Теленик С. Ф., Жаріков Є. В. Управління рівнем послуг в системі Інтернету речей з мікрохмарною архітектурою. *Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка* : зб. наук. праць. 2017. Вип. 65. С. 110–117.
159. Романенко Л., Коротеєва А. Ризики у банківській діяльності. *Фінанси України*. 2003. №5. С. 121–127.
160. Рыхтикова Н.А. Аналіз и управление рисками организации: уч. пос. Москва: ИНФРА, 2009. 240с.

161. Саати Т. Принятие решений. Метод анализа иерархий: пер. с англ. / под ред. Р. Г. Вачнадзе. М. : Радио и связь, 1993. 278 с.
162. Саати Т., Кернс К. Аналитическое планирование. Организация систем. М.: Радио и связь, 1991. 224 с.
163. Сиванич І.Б., Капустян В.О. Моделювання динаміки об'ємів банківських кредитів та депозитів за допомогою моделей перемикання. Актуальні проблеми економіки та управління : зб. наук. праць молодих вчених. 2016. Вип.10. URL: http://ela.kpi.ua/jspui/bitstream/123456789/22511/1/5_4_Syvanyc_h.pdf (дата звернення 25.09.2018).
164. Сиддики Н. Скоринговые карты для оценки кредитных рисков: пер. с англ. Е. Ильичева. М.: Манн, Иванов и Фербер, 2014. 268 с.
165. Система ризик-менеджменту в банках: теоретичні та методологічні аспекти: монографія / за ред. В. В. Коваленко. Одеса: ОНЕУ, 2017. 304 с.
166. Ситникова Н. Ю. Кредитные риски в системе финансового риск-менеджмента: автореф. дис. ... канд. экон. наук: 08.00.10. М., 2004. 25 с.
167. Слепухина Ю. Э. Риск-менеджмент на финансовых рынках: учебное пособие. URL: <http://elar.urfu.ru/bitstream/10995/34716/1/978-5-7996-1522-2.pdf> (дата звернення 25.09.2018).
168. Снитюк В. Е. Эволюционный метод восстановления пропусков в данных *Интеллектуальный анализ информации*: сб. трудов УІ междунар. конф. Киев, 2006. С. 262–271.
169. Снитюк В. Є. Прогнозування. Моделі. Методи. Алгоритми : навч. посіб. Київ : Маклаут, 2008. 364 с.
170. Спирли Э. Корпоративные хранилища данных. Планирование, разработка, реализация. М.: Изд. дом Вильямс, 2001. 396с.
171. Справочное пособие по экологической оценке: в 3 т. Washington : World Bank, 1991.
172. Сравнение методов интеллектуального анализа данных при оценивании кредитоспособности физических лиц / А. Н. Терентьев, П. И. Бидюк, А. В.

- Миронова и др. *Проблемы управления и информатики*. 2009. № 5. С. 141–149.
173. Страхування: теорія та практика : навч. посіб. / за заг. ред. проф. Н. М. Внукової. Харків: Вид-во "Бурун Книга", 2004. 377 с.
174. Терентьев А. Н., Домрачев В. Н., Костецкий Р. И. SAS BASE: Основы программирования. К.: Эдельвейс, 2014. 304 с.
175. Терентьев А. Н., Бидюк П. И. Методы построения Байесовских сетей *Адаптивные системы автоматического управления: межвед. научно-техн. сб.* 2005. №8. С. 130 – 141.
176. Тулупьев А. Л., Николенко С. И., Сироткин А. В. Байесовские сети: логико-вероятностный подход. СПб.: Наука, 2006. 607с.
177. Тычинский А. В. Управление инновационной деятельностью компаний: современные подходы, алгоритмы, опыт. Таганрог: ТРТУ, 2006. URL:<http://www.aup.ru/books/m87/pril5.htm> (дата звернення 25.09.2018).
178. Українська міжбанківська асоціація членів платіжних систем. URL: <https://ema.com.ua/> (дата звернення 25.09.2018).
179. Управління ризиками банків: монографія у 2 томах. Т. 2: Управління ринковими ризиками та ризиками системних характеристик / А. О. Єпіфанов, Т. А. Васильєва, С. М. Козьменко та ін. / за ред. А. О. Єпіфанова, Т. А. Васильєвої. Суми : ДВНЗ “УАБС НБУ”, 2012. 299 с.
180. Усачев С. Кредитный скоринг: решение класса desktop или enterprise?. *Банки и технологии*. 2008. № 4. С. 50 – 54.
181. Финансы предприятий: учебн. пособ. / под ред. Е.И. Бородиной. М.: Изд-во "Банки и биржи", ЮНИТИ, 1995. 208 с.
182. Фінансовий менеджмент: підручник / за заг. ред. Т.А. Говорушко. Львів: «Магнолія 2006», 2014. 344с.
183. Фінансові ризики. URL: <http://www.vuzlib.org/fdsg/2.htm> (дата звернення 25.09.2018).

184. Финансовый риск. URL: <http://www.risk24.ru/finriski1.htm> (дата звернення 25.09.2018).
185. Фомін О. В., Кузнецова Н. В. Скорингові моделі поведінки клієнтів-власників кредитних карток для оцінки їх платоспроможності. *Системні науки та кібернетика*. 2016. №5. С. 56 – 67. URL: http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc_5_2016.pdf (дата звернення 25.09.2018).
186. Хаб'юк О. Банківське регулювання та нагляд через призму рекомендацій Базельського комітету: монографія. Івано-Франківськ: ОППО; Снятин: ПрутПринт, 2008. 260с.
187. Харчук М. В. Аналіз масштабів та основні напрями мінімізації ризиків шахрайства членів міжнародних платіжних систем. *Ефективна економіка*. №6, 2013. URL: <http://www.economy.nauka.com.ua/?op=1&z=2120> (дата звернення 25.09.2018).
188. Центральний банк та грошово-кредитна політика: підручник / кол. авт.: А. М. Мороз, М. Ф. Пуховкіна, М. І. Савлук та ін.; за ред. д-ра екон. наук, проф. А. М. Мороза і канд. екон. наук, доц. М. Ф. Пуховкіної. К.: КНЕУ, 2005. 556 с.
189. Чубукова И. А. Data Mining. М. : Бином ЛБЗ, 2008. 384 с.
190. Шаховська Н. Б., Угрин Д. І. Аналіз задачі опрацювання відсутності та неповноти інформації у сховищі даних. *Вісник Національного університету «Львівська політехніка». Інформаційні системи та мережі*. 2008. № 519. С. 215 – 225.
191. Шеремет А. Д., Сайфулин Р. С. Методика фінансового аналізу. М.: ИНФРА. 1996. 196 с.
192. Ширяев А. Н. Основы стохастической финансовой математики. Т. 1: Факты. Модели. Москва: ФАЗИС. 1998. 512 с.
193. Энциклопедия фінансового риск-менеджмента / под ред. А. А. Лобанова, А. В. Чугунова. М.: Альпина Паблишер. 2003. 786 с.

194. Agarwal V., Taffler R. “Twenty-five years of the Taffler z-score model: does it really have predictive ability?” *Accounting and Business Research*. 2007. Vol. 37. No. 4. P.285 – 300.
195. AgenaRisk. URL: <https://www.agenarisk.com/> (last accessed 25.09.2018).
196. Allen L. N., Rose L. C. Financial survival analysis of defaulted debtors. *Journal of Operational Research Society*. 2006. Vol. 57. P. 630–636.
197. Allison P. D. *Survival Analysis Using SAS: A Practical Guide: Second Edition*. Cary, NC: SAS Institute Inc., 2010. 324 p.
198. Altman E. *Corporate Financial Distress and Bankruptcy: 3rd edition*. – John Wiley and Sons, 2005. 1029p.
199. Altman E. I., Onorato M. An Integrated Risk Model for Defaultable Loans and Bonds. *NYU Working Paper*. 2003. FIN-03-009. 21p.
200. Altman E. I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*. 1968. 189–209. doi:10.1111/j.1540-6261.1968.tb00843.x. (last accessed 25.09.2018).
201. Altunbas Y., Carbo S., Gardener E. P. M., Molyneux P. Examining the Relationships between Capital, Risk and Efficiency in European Banking *European Financial Management*. 2007. Vol. 13. No.1.P. 49–70.
202. Analytic Solver for Excel. URL: <https://www.solver.com/analytic-solver-platform> (last accessed 25.09.2018).
203. Anderson R. *The Credit scoring toolkit*. Oxford: Oxford University Press, 2007. 790 p.
204. Angoss Predictive analytics and business intelligence software and solutions. URL: <http://www.angoss.com/> (last accessed 25.09.2018).
205. Ashlee V. Data Analysts Captivated by R’s Power. URL: http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html?pagewanted=all&_r=0 (last accessed 25.09.2018)
206. Aven T. *Foundations of Risk Analysis. A Knowledge and Decision-Oriented Perspective*. John Wiley & Sons Ltd, 2003. 198p.

207. Azzalini Adelchi, Scarpa Bruno. Data Analysis and Data Mining. An Introduction. Oxford University Press. 2012. 289 p.
208. Baba N., Goko H. Survival analysis of hedge funds *Working Papers Series*. 2006. Vol. 5. No. 6. P. 1–2.
209. Baillie R. T., Bollerslev T., Mikkelsen H. O. Fractionally Integrated Generalized Conditional Heteroskedasticity . *Journal of Econometrics*. 1996. Vol. 74. P. 3–30.
210. Bali Raghav, Dipanjan Sarkar. R Machine Learning By Example. Packt Publishing Ltd, 2016. 340 p.
211. Basel Committee-Basel III. URL: https://www.bis.org/list/bcbs/tid_132/index.htm. (last accessed 25.09.2018)
212. Basel II: International Convergence of Capital Measurement and Capital Standards: a Revised Framework – Comprehensive Version. URL: <http://www.bis.org/publ/bcbs54.htm> . (last accessed 25.09.2018).
213. Beninel F. Bouaguel W., Belmufti G. Transfer Learning Using Logistic Regression in Credit Scoring. Cornell University Library, 2012. URL: <https://arxiv.org/pdf/1212.6167.pdf>. (last accessed 25.09.2018).
214. Beran J., Dja A. K. Credit risk modeling based on survival analysis with immunes *Statistical Methodology*. 2007. Vol. 4. P. 251–276.
215. Berry Michael J.A., Linoff Gordin S. Data Mining Techniques for Marketing, Sales and Customer Relationship Managers:2ed.: Wiley publishing. 2011.888 p.
216. Best's Enterprise Risk Model: A Value-at-Risk Approach . *Seabury Insurance Capital*. April. 2001. 64 p. URL: <http://www.casact.org/coneduc/specsem/erm/2001/handouts/freestone1.pdf>. (last accessed 25.09.2018).
217. Bidyuk P. I., Konovalyuk M. M., Kuznetsova N.V., Pudlo I. V. Adaptive Short-Term Forecasting of Selected Financial Processes. *Research bulletin of NTUU “KPI”*. 2014. N1. P.35–41.
218. Bidyuk P. I. Kuznietsova N.V. Decision support system for adaptive processes forecasting in various applied areas with alternative techniques.

Геоинформационные системы и компьютерные технологии эколого-экономического мониторинга: сб. докл.междун. научно-техн. конф. (г. Днепропетровск, 13–15 апреля 2016 г.) / под ред. Л.В.Сарычевой.: Электрон. данные. Днепропетровск: ГВУЗ «НГУ» МОН Украины, 2016. URL: http://gis.dp.ua/conf2016-publications/sections/iad/1_Bidyuk_Kuznetsova.pdf.

219. Blonda P., Baraldi A., D'Addabbo A., Tarantino C., De Blasi R. RBF Networks Exploiting Supervised Data in the Adaptation of Hidden Neuron Parameters: *Advances in Artificial Intelligence. Lecture Notes in Computer Science AI*IA*. 2001. Vol. 2175. P.51 – 56.
220. Bodyanskiy Ye., Zaychenko Yu., Pavlikovskaya E., Samarina M., Viktorov Ye. The neo-fuzzy neural network structure optimization using the GMDH for the solving forecasting and classification problems. *Proc. Int. Workshop on Inductive Modeling*. Krynica, Poland, 2009. P. 77 – 89.
221. Bollerslev T. A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*. 1987. Vol. 69. P. 542–547.
222. Bollerslev T., Chow R., Kroner K. ARCH modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics*. 1992. 52. P. 5-59.
223. Bollerslev T., Engle R. F., Nelson D. ARCH models. *Handbook of Econometrics*:/ eds. R.Engle and D.McFadden. North-Holland, Amsterdam. 1993. Vol. 4. P. 2959 – 3038.
224. Bollerslev T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*. 1986. Vol. 31. P. 307 – 327.
225. Bollerslev T., Mikkelsen H. O. Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*. 1996. Vol. 73. P. 151–184.
226. Bollerslev T. On the correlation structure of the generalized autoregressive conditional heteroskedastic process. *Journal of Times Series Analysis*. 1988. Vol. 9. P. 121–131.

227. Bolstad W. M. Understanding computational Bayesian statistics. Hoboken (New Jersey): John Wiley & Sons, Ltd, 2010. 334 p.
228. Bouchaud J.-Ph. Elements for a Theory of Financial Risks. *Physica A*. 2000. Vol.285. – No.1. – pp. 18–28.
229. Bouchaud J.-Ph., Potters M. Theory of Financial Risk. *Science & Finance* .1999. 237p. URL: <http://web.math.ku.dk/~rolf/Klaus/bouchaud-book.ps.pdf> (last accessed 25.09.2018).
230. Bramer M. Principles of Data Mining . Springer–Verlag. 2016. 526 p.
231. Breheny P. Robust regression. URL: <http://web.as.uky.edu/statistics/users/pbreheny/764-F11/notes/12-1.pdf>. (last accessed 25.09.2018).
232. Breiman L. Bagging predictors. *Machine learning*. 1996. 24(2). P.12 –140.
233. Breiman L., Random forests. *Machine learning*. 2001. 45(1) P. 5–32.
234. Breiman L., Friedman J., Stone C. J., Olshen R. A. Classification and regression trees. New York: CRC press, 1984. 368 p.
235. Brockwell P. J., Davis R. A. Time Series: Theory and Methods:2nd ed. New York : Spring-Verlag, 2009. 580p.
236. Bruno G. R. Lean Compendium. Introduction to Modern Manufacturing Theory. Springer International Publishing, 2018.149 p. DOI 10.1007/978-3-319-58601-4. (last accessed 25.09.2018).
237. Bunker R.P., Naeem M.A., Zhang W. Improving a Credit Scoring Model by Incorporating Bank Statement Derived Features. Cornell University Library, 2016. URL: <https://arxiv.org/ftp/arxiv/papers/1611/1611.00252.pdf>. (last accessed 25.09.2018).
238. Burstein F., Holsapple C.W. Handbook of Decision Support Systems. Berlin: Springer-Verlag, 2008. 908 p.
239. Cao R., Vilar J.M., Devia A. Modelling consumer credit risk via survival analysis *SORT* 33 (1) January-June 2009, p.3–30.

240. Capon N. Credit scoring systems: a critical analysis. *Journal of Marketing*. 1982. Vol. 46. No. 1. P. 82–91.
241. Carling K., Jacobson T., Roszbach K. Duration of consumer loans and bank lending policy: dormancy versus default risk. *Working Paper Series in Economics and Finance*. 1998. No. 280. P. 109–122.
242. Carnero M. A., Pea D., Ruiz E. Is stochastic volatility more flexible than garch? *Working paper, Departamento de Estadística y Econometría*. Universidad Carlos III de Madrid. 39p. URL: <https://e-archivo.uc3m.es/bitstream/handle/10016/152/ws010805.pdf?sequence=1> (last accessed 25.09.2018).
243. Casolaro L., Gobbi G. Information Technology and Productivity Changes in the Banking Industry. *Economic Notes*. 2007. Vol.36. No.1. P.43–76.
244. Chen B., Gel Y. R., Balakrishna. N., Bovas A. Computationally efficient bootstrap prediction intervals for returns and volatilities in ARCH and GARCH processes. 2011. 25p. URL: http://www.academia.edu/21503708/Computationally_efficient_bootstrap_prediction_intervals_for_returns_and_volatilities_in_ARCH_and_GARCH_processes (last accessed 25.09.2018).
245. Chen G., Thomas A. Bound and Collapse Bayesian Reject Inference When Data are Missing not at Random. *Proc. Conference Banff International Research Station for Mathematical Innovation and Discovery*. 2003. 30 p. URL: <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=C9C9955D3639951308F47C274A7254FA?doi=10.1.1.115.3430&rep=rep1&type=pdf> (last accessed 25.09.2018).
246. Cheng T., Carlos C., “XGBoost. A Scalable Tree Boosting System”. Retrieved May 17, 2018. URL: <https://arxiv.org/pdf/1603.02754.pdf>. (last accessed 25.09.2018).
247. Chou R.Y. Volatility persistence and stock returns – some empirical evidence using GARCH. *Journal of Applied Econometrics*. 1987. No. 3, p. 279 – 294.

248. Connor G. A Structured GARCH Model of Daily Equity Return Volatility 2001. URL: https://www.researchgate.net/publication/5055210_A_Structured_GARCH_Model_of_Daily_Equity_Return_Volatility (last accessed 25.09.2018).
249. Conway Drew, White John Myles. *Maschile Learning for Hackers*. O'Reiley. 2012. 322 p.
250. Cooman G., Zaffalon M. Updating beliefs with incomplete observations. *Artificial Intelligence*. 2004. Vol. 159, Iss.1–2. P. 75–125.
251. Cox D. R. Regression Models and Life-Tables. *Journal of the Royal Statistical Society, Series B*. 1972. Vol. 34, No2. P. 187–220.
252. Cox D. R., The regression analysis of binary sequences (with discussion). *Journal of the Royal Statistical Society, Series B*. 1958. Vol.20: P. 215 – 242.
253. Dabrowska D. Non-parametric regression with censored survival time data. *Scandinavian Journal of Statistics*. 1987. Vol. 14, No. 3. P. 181–197.
254. Dempster A. P., Laird N. M., Rubin D. B. Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1977. Vol. 39, N. 1. P. 1– 38.
255. Ding Z., Granger C. W. J., Engle R. F. A Long Memory Property of Stock Returns . *Journal of Empirical Finance*. 1993. Vol. 1. P. 83–106.
256. Dobson A. *An Introduction to Generalized Linear Models*. New York: CRC Press Company, 2013. 407 p.
257. Downey A. B. *Think Bayes. Bayesian Statistics in Python* . O'Reilly, 2013. 210 p.
258. Durand D. *Risk Elements in Consumer Installment Financing*. New York: National Bureau of Economy Research, Inc., 1941. 101p.
259. Engle R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*. 1982. Vol. 50. P. 987–1007.
260. Engle R. F., Bollerslev T. Modeling the persistence of conditional variance. *Econometric Reviews*. 1986. Vol. 5. P. 1–50.
261. Fleming T. R., Harrington D. P. *Counting Processes and Survival Analysis*. New York: John Wiley & Sons, 1991. 429p.

262. Gartner. Homepage. URL: <https://www.gartner.com/en>. (last accessed 25.09.2018).
263. Gilks W. R., Richardson S., Spiegelhalter D. J. Markov Chain Monte Carlo in Practice. New York: Chapman & Hall/CRC, 2000. 486 p.
264. Glennon D., Nigro P. Measuring the default risk of small business loans: a survival analysis approach. *Journal of Money, Credit, and Banking*. 2005. Vol. 37. P. 923–947.
265. Gogishvili P. Determination of the Vehicle Location in Case of Incomplete GPS Data. *Інформаційні технології та комп'ютерна інженерія*. 2012, №3. С.19 – 23.
266. Granger C., Joyeux R. An Introduction to Long Memory Time Series Models and Fractional Differencing. *Journal of Time Series Analysis*. 1980. Vol. 1. P. 15–39.
267. Guo X., Zhu Z., Shi J. A Corporate Credit Rating Model Using Support Vector Domain Combined with Fuzzy Clustering Algorithm. *Mathematical Problems in Engineering*, 2012, Article ID 302624, 20 pages doi:10.1155/2012/302624. (last accessed 25.09.2018).
268. Hamilton J.D. A new approach for the economic analysis of nonstationary time series and the business cycle. *Econometrica*. 1989. Vol. 57. P. 357–384.
269. Han J., Kamber M. Data Mining Concepts and Techniques. Morgan Kaufman, 2011. 744 p.
270. Hand D. J., Henley W. E. Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 1997. Vol. 160. No. 3. P. 523–541.
271. Hand D., Mannile H., Smyth P. Preexamples of data mining. A Bradford Book, 2001. 578 p.
272. Harris R., Sollis R. Applied Time Series Modelling and Forecasting. West Sussex: John Wiley & Sons Ltd., 2005. 313 p.

273. Hodeghatta U. R., Nayak U. Business Analytics Using R – A Practical Approach. New York: Apress, 2017. 280 p.
274. Hollsapple C.W. Winston A.B. Decision Support Systems. Saint Paul: West Publishing Company, 1996. 860 p. URL: <http://arxiv.org/ftp/arxiv/papers/1306/1306.2025.pdf>. (last accessed 25.09.2018).
275. Hosking J. Fractional Differencing . *Biometrika*. 1981. Vol. 68. P. 165–76.
276. How does Beta reflect systematic risk? | InvestopediaL. [Online]. URL: <https://www.investopedia.com/ask/answers/031715/how-does-beta-reflect-systematic-risk.asp#ixzz5GMuVIGCm> (last accessed 25.09.2018).
277. Hyper-V Technology Overview. URL: <https://docs.microsoft.com/en-us/windows-server/virtualization/hyper-v/hyper-v-technology-overview> (last accessed 25.09.2018).
278. IBM Cognos Business Intelligence. URL: https://www.ibm.com/support/knowledgecenter/ru/SSEP7J_10.1.1/com.ibm.swg.ba.cognos.wig_cr.10.1.1.doc/c_gtstd_c8_bi.html (last accessed 25.09.2018).
279. IBM SPSS Statistics. URL: <http://www.predictivesolutions.ru/software/statistics.htm> (last accessed 25.09.2018).
280. International Convergence of Capital Standards. A revised Framework . Basel Committee of Banking Supervision. Basel. Updated November 2005. URL: <https://www.bis.org/publ/bcbsca.htm>. (last accessed 25.09.2018).
281. Jensen F.V., Nielsen Th. Bayesian Networks and Decision Graphs. New York: Springer–Verlag, 2009. 457 p.
282. Jilek O. Mathematical applications in credit risk modeling. *Journal of Applied Mathematics*. 2008. Vol. 1. No. 1. P. 432–438.
283. JPMorgan Chase & Co. URL: <https://www.jpmorganchase.com/>(last accessed 25.09.2018).
284. Kantardzic Mehmed. Data Mining: Concepts, Models, Methods and Algorithms. Wiley-IEEE Press, 2011. 552 p.

285. Kaplan E.L., Meier P. Non-parametric estimation for incomplete observations. *J. Am. Stat. Assoc.* 1958. N. 53. P. 457–81. URL: <http://www.jstor.org/stable/2281868> (last accessed 25.09.2018).
286. Kim S., Shephard N., Chib S. Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies*. 1998. Vol. 65. P. 361–393.
287. Kiss F. Credit scoring processes from a knowledge management perspective. *Periodica Polytechnica Ser. Soc. Man. Sci.* 2003. Vol. 11. №1. P. 95–110.
288. Kjaerulff U. dHugin: A Computational System for dynamic time-sliced Bayesian networks. *International Journal of Forecasting*. 1995. No. 11. P. 89–111.
289. Kleinbaum David G., Klein M. Survival Analysis. A self-Learning Text (Statistics for Biology and Health). 3rd ed. Springer, 2011. 700 p.
290. Korb K. B., Nicholson A. E. Bayesian artificial intelligence. New York: Chapman & Hall/CRC, 2004. 365 p.
291. Kuznyetsova N. Informational technologies for financial data analysis. *Proc. IV Int. Scientific Conf. "Intelligence, Integration, Reliability"*. Kyiv-Warsaw, April 21-22, 2011). K.: NTUU "KPI". 2011. P. 32–33.
292. Kuznietsova N. V. Information Technologies for Clients' Database Analysis and Behaviour Forecasting. *CEUR Workshop Proceeding (ISSN 1613-0073)* 2017. Vol. 2067. P.56-62 [Online]. Available: <http://ceur-ws.org/Vol-2067/>. (last accessed 25.09.2018).
293. Kuznyetsova N. Integrated approach to credit rating. *Int. Scientific Conf. "Intelligence, Integrity, Reliability"*. K.: NTUU "KPI", 2010. P. 30–31.
294. Kuznietsova N. Project management. Students hackathon for mobile application *Summer School IT-SEA 5-20 July'14. Report*. Odessa National Polytechnic University. Odessa. 2014. P. 122–130.

295. Kuznietsova N. Requirements management. *Summer School IT-SEA 5-20 July'14. Report*. Odessa National Polytechnic University. Odessa, 2014. P. 98–121.
296. Kuznietsova N. V. Scoring technology for risk assessment of fraud in banking *CEUR Workshop Proceeding (ISSN 1613-0073)*. 2016. Vol. 1813. P.54-61 [Online]. Available: <http://ceur-ws.org/Vol-1813/paper8.pdf>. (last accessed 25.09.2018).
297. Kuznietsova N. V., Bidyuk P. I. Business Intelligence Techniques For Missing Data Imputation. *Research bulletin of NTUU "KPI"*. 2015. N5. P. 47–56.
298. Kuznietsova N., Bidyuk P. Forecasting of Financial Risk Users' Outflow. *IEEE First International Conference on System Analysis & Intelligent Computing (SAIC)*, (Kyiv: 08-12 October). 2018. P.250-255. DOI: <https://ieeexplore.ieee.org/abstract/document/8516782>.
299. Kuznietsova N.V., Bidyuk P.I. Modeling of credit risks on the basis of the theory of survival. *Journal of Automation and Information Sciences*.2017.Vol. 49. Is. 11. P. 11–24.
300. Kuznietsova N.V., Bidyuk P.I. Systemic approach to estimation of financial risks. *Науково-технічна конференція «Інформатика, математика, автоматика. ІМА:2015 (Суми, 20–25 квітня 2015 року. Сумський державний університет)*. 2015. С. 46–47.
301. Kuznietsova N., Seebauer M., Zabelin S. Some Methods for Estimating Financial Risks in Banking. *IEEE First International Conference on System Analysis & Intelligent Computing (SAIC)*, (Kyiv: 08-12 October, 2018) 2018. P.271-274. DOI: <https://ieeexplore.ieee.org/document/8516873>.
302. Lehmann B. How Good Is „Good“ ? -Making Better Use of Subjective Information in Bank Internal Credit Scoring Systems:Dissertation zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften:Doctor rerum politicarum (Dr.rer.pol.)am Fachbereich Wirtschaftswissenschaften der Universität Konstanz, 2008. 238p.

303. Lightning Fast Data Science Platform for Teams | RapidMiner©. URL: <https://rapidminer.com/> (last accessed 25.09.2018).
304. Lin Shu-Min SMEs Credit Risk Modelling for Internal Rating Based Approach in Banking Implementation of Basel II Requirement: Thesis for Doctor of Philosophy, The University of Edinburgh, 2007. 270p.
305. Lobato I. N., Savin N. E. Real and Spurious Long-Memory Properties of Stock-Market Data. *Journal of Business and Economics Statistics*. 1998. Vol. 16. P. 261–68.
306. Malik M., Thomas T.L. Modelling credit risk of portfolio of consumer loans. *School of Management Working Paper Series*. 2006. Vol. 7. No. 12. P. 1–2.
307. Marimo M. Survival analysis of bank loans and credit risk prognosis master of science mathematical statistics URL: http://wiredspace.wits.ac.za/jspui/bitstream/10539/18597/1/Mercy%20Marimo%20Thesis_Survival%20Analysis_28.03.%202015_v1.pdf. (last accessed 25.09.2018).
308. Marwala T. Flexibly-bounded Rationality and Marginalization of Irrationality Theories for Decision Making. 2013. URL: https://www.researchgate.net/publication/237082530_Flexibly-bounded_Rationality_and_Marginalization_of_Irrationality_Theories_for_Decision_Making. (last accessed 25.09.2018).
309. Mays E. (Ed.) Handbook of Credit Scoring. Chicago: Glenlake Publishing Company, Ltd., 2001. 460.
310. McDonald E., “The Global Games Market Will Reach \$108.9 Billion in 2017 With Mobile Taking 42%”. Retrieved May 17, 2018 from: <https://newzoo.com/insights/articles/the-global-games-market-will-reach-108-9-billion-in-2017-with-mobile-taking-42/>. (last accessed 25.09.2018).
311. McLachlan G. J., Krishnan T. The EM algorithm and extensions. Hoboken: John Wiley & Sons, Inc., 2008. 359 p.

312. McNeil A. J., Frey R., Embrechts P. Quantitative Risk Management. Concepts, Techniques and Tools – Revised Edition. Princeton: Princeton University Press, 2015. 720p.
313. Murphy K. A Brief Introduction to Graphical Models and Bayesian Networks. University of British Columbia, Faculty of Science. URL: <https://www.cs.ubc.ca/~murphyk/Bayes/bayes.html/> (last accessed 25.09.2018).
314. Murray I. Advances in Markov chain Monte Carlo methods. London: University of London, 2007. 176 p.
315. Narain B. Survival analysis and the credit granting decision. *Credit Scoring and Credit Control*. 1992. No. 1. P. 1–2.
316. Neil M., Fenton N. E., Tailor M. Using Bayesian networks to model expected and unexpected operational losses. *Risk Analysis*. 2005. P. 34–57.
317. Neil M., Tailor M., Marquez D. Inference in hybrid Bayesian networks using dynamic discretization. *Statistics and Computing*. 2007. Vol. 17. Is. 3. P. 219 – 233.
318. Nelson D. B. Conditional heteroscedasticity in asset returns: a new approach. *Econometrica*. 1991. Vol. 59. No. 2. P. 347 – 370.
319. Nisbet R., Elder J., Miner G. Handbook of Statistical Analysis & Data Mining Applications. Academic Press, 2009. 864p.
320. Oracle Named a Leader of the Sixth Consecutive Year. URL: <https://www.oracle.com> (last accessed 25.09.2018).
321. Overview of Online Analytical Processing (OLAP). URL: <https://support.office.com/en-us/article/overview-of-online-analytical-processing-olap-15d2cdde-f70b-4277-b009-ed732b75fdd6?ui=en-US&rs=en-US&ad=US> (last accessed 25.09.2018).
322. Owen M. Tukey's Biweight Correlation and the Breakdown. 2005. URL: <http://pages.pomona.edu/~jsh04747/Student%20Theses/MaryOwen10.pdf> (last accessed 25.09.2018).

323. Pearl J. Probabilistic reasoning in intelligent systems. San Francisco: Morgan Kaufmann Publishers, 1988. 552 p.
324. Perianez A. Churn Prediction in Mobile Social Games: Towards a Complete Assessment Using Survival Ensembles. Retrieved May 17, 2018 from: <https://arxiv.org/pdf/1710.02264.pdf> (last accessed 25.09.2018).
325. Peters J-P. Estimating and forecasting volatility of stock indices using asymmetric GARCH models and (Skewed) Student-t densities / Jean-Philippe. Ecole d'Administration des Affaires, University of Liege, Belgium, 2001. 20p. URL:<https://pdfs.semanticscholar.org/f445/493c46b3b4d52e958f69d502ccf8f03d51c3.pdf>. (last accessed 25.09.2018).
326. Poncelet P., Teisseire M., Maseglia F. Data Mining Patterns: New Methods and Application. Idea Group Referens, 2007. 307p.
327. Prado H. A, Ferneda E. Emerging Technologies of Text Mining Techniques and Applications. Idea Group Referens, 2007. 358 p.
328. Principles For the Management of Credit Risk. URL: <http://www.bis.org/publ/bcbs54.htm> (last accessed 25.09.2018).
329. ProZorro: публічні закупівлі. URL: <https://prozorro.gov.ua/>. (last accessed 25.09.2018).
330. Qlik: Бизнес-анализ. Инструменты визуализации данных. URL: <https://www.qlik.com/ru-ru> (last accessed 25.09.2018).
331. Ramos F.T., Cozman F. G. Selecting Exact and Approximate algorithms in decomposed Bayesian Networks. *Proc. Workshop*. 2002. P. 151–155.
332. Roszbach K. Bank lending policy, credit scoring and the survival of loans. *Sverriges Riksbank Working Paper Series*. 2003. No. 154. P. 1-2.
333. Rubin D. B. An Overview of Multiple Imputation. *Proc. Survey Research Methods Section of the American Statistical Association*. 1988. P. 79–84.
334. Russell M. A. Mining the Social Web (Data Mining Facebook, Twitter, LinkedIn, Google, GitHub and More). O'Reilly, 2013. 448 p.

335. SAP Software Solutions: Business Applications and Technology. URL: <https://www.sap.com/index.html> (last accessed 25.09.2018).
336. Saradhi V., Palshikar G. K. Employee churn prediction. *Expert Systems with Applications*, 2011. 38(3). P.1999–2006.
337. SAS Enterprise Miner. URL: https://www.sas.com/en_us/software/enterprise-miner.html (last accessed 25.09.2018).
338. SAS Viya. URL: https://www.sas.com/en_us/software/viya.html (last accessed 25.09.2018).
339. Schoderbek P. P., Schoderbek C. G., Kefalas A. G. Management Systems: Conceptual Considerations. New York: Irwin Inc. Company, Custom Publishing, USA, 1990. 280 p.
340. Shi F., Zhang D., Chen J., Karimi H. R. Missing Value Estimation for Microarray Data by Bayesian Principal Component Analysis and Iterative Local Least Squares. *Mathematical Problems in Engineering* . Article ID 162938. – 2013. URL: <http://www.hindawi.com/journals/mpe/2013/162938/>. (last accessed 25.09.2018).
341. Shumway R. H., Stoffer D. S. Time Series Analysis and its Applications. New York: Springer Verlag, 2006. 588 p.
342. Shutt R, O’Neil C. Doing Data Science. Straight Talk From the Frontline. O’Reilly, 2013. 408 p.
343. Siddiqi N. Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring . Hoboken : John Wiley & Sons, 2005. 208 p.
344. Size of the online gambling market from 2009 to 2020. Retrieved May 17, 2018 from URL: <https://www.statista.com/statistics/270728/market-volume-of-online-gaming-worldwide/>.
345. Smaranda, C. Scoring Functions and Bankruptcy Prediction Models – Case Study for Romanian Companies. *Procedia Economics and Finance*. 2014. Vol. 10. P. 217 – 226.

346. Smarter Analytics for Connected Intelligence. Accelerate innovation with data science. URL: <https://www.tibco.com/products/tibco-statistica> (last accessed 25.09.2018).
347. Sorournejad S., Zojaji Z., Atani R.E., Amir Hassan Monadjemi. A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective. Cornell University Library, 2016. URL: <https://arxiv.org/ftp/arxiv/papers/1611/1611.06439.pdf> (last accessed 25.09.2018).
348. Sowell F. B. Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models. *Journal of Econometrics*. 1992. Vol. 53. P. 165–188.
349. Stepanova M., Thomas L. C. Survival analysis methods for personal loan data. *Operations Research*. 2002. Vol. 50. No. 2. P. 277–289.
350. Stokey N.L., Lucas R. E. Jr. with the collaboration of Prescott E.C. Recursive Methods in Economic Dynamics. New York: Harward College, 1989. 26p.
351. Svolba G. Data Quality for Analytics Using SAS. Cary, NC: SAS Institute Inc., 2012. 340 p.
352. Tableau. URL: <https://www.tableau.com/> (last accessed 25.09.2018).
353. Tamaddoni J. A., Sepehri M. M., Teimourpour B., and Choobdar S. Modeling customer churn in a non-contractual setting: the case of telecommunications service providers. *Journal of Strategic Marketing*. 2010. 18(7). P.587–598.
354. Taffler R. J. Forecasting company failure in UK using discriminant analysis and financial ratio data. *Journal of the Royal Statistical Society*, 1982. 145 (3). P. 342–358.
355. Taylor S. J. Financial returns modeled by the product of two stochastic processes – a study of the daily sugar prices: 1961 – 1975. *Time Series Analysis: Theory and Practice/ In Anderson, O. D. (ed.), 1. – Amsterdam: North-Holland, 1982. P. 203 – 226.*

356. Taylor S. J., *Modelling Financial Time Series*. Chichester: John Wiley, 1986. 268 p.
357. Telenyk S., Bukasov M. Data Center Resource Management for SAAS. *Technical Transactions Electrical Engineering. 3-E/ 2016*. P.229-235. DOI: 104467/2353737XCT.16.279.6078.
358. The Bank for International Settlements (BIS), Basel Committee on Banking Supervision, *Quantitative Impact, Study 3, Technical Guidance*, October 2002.
359. TIBCO Software: Global Leader in Integration and Analytics Software. URL: <https://www.tibco.com/> (last accessed 25.09.2018).
360. Trofymchuk O. M., Bidiuk P.I., Gozhyj O. P., Bidiuk O. P. Decision support system for implementing systemic approach to forecasting. *International Journal of Computers and Technology*. 2015. Vol. 14. No. 5.P. 5769 – 5778.
361. Tsay R.S. *Analysis of financial time series*. New York: John Wiley & Sons, Inc., 2010. 715 p.
362. Tuffery S. *Data mining and statistics for decision making*. New York: John Wiley & Sons, Inc., 2011.704 p.
363. Vapnik Vladimir N. *The Nature of Statistical Learning Theory*. 2ed. Springer-Verlag, 2000. 314 p.
364. Verbeke W., Martens D., Mues C., Baesens B. Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications*. 2011. 38(3). P. 2354–2364.
365. vSphere Hypervisor. URL: <https://www.vmware.com/products/vsphere-hypervisor.html> (last accessed 25.09.2018).
366. Witten I. H., Frank E. *Data Mining. Practical Machine Learning Tools and Techniques with Java Implementations*. Elsevier Sciens, 2016. 620 p.
367. Ye Nong. *Data Mining Theories, Algorithms and Examples*. CRC Press, 2013. 349 p.

368. Yu X., Guo S., Guo J., Huang X. An extended support vector machine forecasting framework for customer churn in e-commerce. *Expert Systems with Applications*. 2011. 38(3). P.1425–1430.
369. Yoon S., Koehler J., Ghobarah A. Prediction of advertiser churn for google adwords. *JSM proceedings*, 2010. URL: <https://storage.googleapis.com/pub-tools-public-publication-data/pdf/36678.pdf>. (last accessed 25.09.2018).
370. Zaychenko Yu. The fuzzy Group Method of Data Handling and its application for economical processes forecasting. *Scientific Inquiry*. 2006. 7. №1. P. 83–96.
371. Zgurovsky M. Z., Zaichenko Yu. P. An Introduction to Computing Intelligence. Kyiv: Naukova Dumka, 2013. 406 p.
372. Zhao Yanchang. R and Data Mining Examples and Case Studie. Academic Press, 2012. 256 p.

ДОДАТОК А

ФІНАНСОВІ РИЗИКИ: КЛАСИФІКАЦІЯ, МОДЕЛІ І МЕТОДИ ОЦІНЮВАННЯ, ПРИКЛАДИ

А.1. Класифікації фінансових ризиків

Існує досить багато класифікацій фінансових ризиків за різними критеріями. У роботі [193] запропоновано критерії класифікації ризиків і фінансових систем за: *типом невизначеності, охопленням об'єктів ризику, точністю вимірювання.*

Основними причинами невизначеності, тобто джерелами ризику є: обсяг, зміст, корисність інформації щодо об'єкту, що аналізується, і випадковість. З ними пов'язані чотири аспекти ризику: ризик кількості, адекватності, якості і випадковості [193]. Ризик кількості виникає через нестачу інформації, неповноту знань. Ризик адекватності пов'язаний з недосконалістю людського мислення і виникає в процесі інтерпретації інформації та проявляється у вигляді багатозначності і неадекватного відображення суті даних. Ризик якості виявляється в результаті можливих розбіжностей в процесі аналізу між інформацією, що використовується, і її корисністю з точки зору зниження невизначеності щодо об'єкту. Ризик випадковості пов'язаний з можливістю виникнення ситуації, яка в подібних умовах проявляється неоднаково.

За охопленням об'єктів ризику можуть бути глобальними, частковими і одиничними. Глобальні ризики пов'язані зі зміною ринкової кон'юнктури і його впливом на більшість учасників ринку. Часткові ризики пов'язані з економічним становищем лише деяких груп учасників ринку. Поодинокі ризики стосуються результатів діяльності окремих учасників ринку [133, 193]. Наприклад, кредитний ризик пов'язаний перш за все з фінансовим становищем одного з суб'єктів кредитної угоди – позичальника, тому може бути або одиничним, або частковим (в разі, якщо кредит видається групі пов'язаних між собою позичальників). Процентний ризик не пов'язаний з якою-небудь конкретною операцією, одиничним об'єктом. Він – результат кредитної, інвестиційної та процентної політики компанії, а також макроекономічної ситуації в країні, а тому його відносять до глобальних ризиків.

За точністю вимірювання виділяють відносний і абсолютний ризики. Відносний ризик – це ризик виникнення якої-небудь події без урахування зовнішніх чинників, що прямим чином не відносяться до цієї події. Абсолютний ризик – це сукупний ризик, який складається з відносних ризиків.

За часом виникнення виділяють: ретроспективні, поточні, перспективні ризики.

За чинниками виникнення: політичні (зумовлені зміною політичної обстановки, що впливає на підприємницьку діяльність) та економічні ризики (обумовлені несприятливими змінами в економіці підприємства або в економіці країни).

За характером обліку: зовнішні ризики (безпосередньо не пов'язані з діяльністю підприємства або його контактної аудиторії) та внутрішні ризики (обумовлені діяльністю самого підприємства і його контактної аудиторії).

За характером наслідків ризики підрозділяються на чисті і спекулятивні. Якщо ризик передбачає наявність як позитивних, так і негативних результатів, він відноситься до спекулятивних ризиків. Чисті ризики (прості або статичні) характеризуються тим, що вони несуть в собі втрати для підприємницької діяльності, тобто негативні відхилення, або відсутність результату.

За об'єктом, що характеризується виділяють наступні групи фінансових ризиків: ризик окремої фінансової операції, різних видів фінансової діяльності (наприклад, ризик інвестиційної чи кредитної діяльності підприємства), фінансової діяльності підприємства в цілому [193]. Комплекс різних видів ризиків, властивих фінансовій діяльності підприємства, визначається специфікою організаційно-правової форми його діяльності, структурою капіталу, складом активів, співвідношенням постійних і змінних витрат і т.п.

За сукупністю досліджуваних інструментів виділяють індивідуальний та портфельний фінансовий ризик.

За джерелами виникнення виділяють наступні групи фінансових ризиків:

➤ зовнішні, систематичні або ринкові ризики (всі терміни визначають цей ризик як незалежний від діяльності підприємства). Цей вид ризику характерний для всіх учасників фінансової діяльності і всіх видів фінансових операцій. Він виникає при зміні окремих стадій економічного циклу, зміні кон'юнктури фінансового ринку і в ряді інших випадків, на які підприємство в процесі своєї діяльності вплинути не може. До цієї групи ризиків відносять інфляційний ризик, процентний ризик, валютний ризик, податковий ризик і частково інвестиційний ризик (при зміні макроекономічних умов інвестування);

➤ внутрішні, несистематичні або специфічні ризики (всі терміни визначають цей фінансовий ризик як залежний від діяльності конкретного підприємства). Він може бути пов'язаний з некваліфікованим фінансовим менеджментом, неефективною структурою активів і капіталу, надмірної прихильністю до ризикових (агресивних) фінансових операцій з високою нормою прибутку, недооцінкою господарських партнерів та іншими факторами, негативними наслідками яких значною мірою можна запобігти при ефективному управлінні фінансовими ризиками.

За рівнем фінансових втрат ризики поділяються на такі групи [159]:

- допустимий фінансовий ризик – фінансові втрати за ризиком не перевищують розрахункової суми прибутку по здійснюваній фінансовій операції;
- критичний фінансовий ризик – фінансові втрати за ризиком не перевищують розрахункової суми валового доходу по здійснюваній фінансовій операції;
- катастрофічний фінансовий ризик – фінансові втрати за ризиком визначаються частковою або повною втратою власного капіталу (цей вид ризику може супроводжуватися втратою і позикового капіталу).

За можливістю передбачення ФР поділяються на такі дві групи:

- прогнозовані фінансові ризики, пов'язані з циклічним розвитком економіки, зміною стадій кон'юнктури фінансового ринку, передбачуваним розвитком конкуренції і т.п. Прикладом прогнозованих фінансових ризиків є інфляційний ризик, процентний ризик і деякі інші їх види;
- непрогнозовані фінансові ризики, які характеризуються повною непередбачуваністю прояви. Прикладом таких ризиків виступають ризики форс-мажорної групи, податковий ризик і деякі інші.

За можливістю страхування фінансові ризики бувають: страхові (тобто ті, які можуть бути передані в порядку зовнішнього страхування відповідним страховим організаціям) та нестрахові (ті види, за якими відсутня пропозиція відповідних страхових продуктів на страховому ринку).

А.2. Огляд моделей оцінювання волатильності фінансових процесів і приклади моделювання

Волатильність – основний параметр для процесів ціноутворення, портфельної оптимізації, VaR аналізу та інших методів оцінювання і менеджменту ризиків. Тому «адекватне» моделювання і, за можливості, точніше оцінювання волатильності фінансових процесів є однією із центральних задач, які виникають сьогодні у фінансовій сфері. Проаналізуємо сучасні математичні моделі гетероскедастичних процесів, зокрема моделі регресійного типу: авторегресія з умовною гетероскедастичністю (АРУГ(p)), узагальнена модель такого ж типу (УАРУГ(p, q)) і модель стохастичної волатильності (МСВ), яка здобула популярність завдяки високій адекватності опису умовної дисперсії досліджуваних процесів, та деякі інші. Властивості кожної моделі розглянуто стосовно адекватності опису волатильності, виконано порівняльний аналіз моделей різної структури. Запропоновано модифіковану структуру МСВ, що враховує значення волатильності у минулі проміжки часу (модель авторегресійного типу).

Моделі зі змінною волатильністю

Розвиток фінансових процесів за допомогою часових рядів представляється моделлю, що описує змінну волатильність доходів y_t як:

$$y_t = \mu_t + \sigma_t \varepsilon_t, \quad \varepsilon_t \sim N(0, 1), \quad t = 1, 2, \dots, T,$$

$$\mu_t = \alpha + \sum_{i=1}^k b_i x_{i,t},$$

де μ_t – середнє значення, що залежить від константи α та коефіцієнтів регресії b_1, \dots, b_k ; σ_t – значення змінної у часі волатильності. Серед змінних $x_{1,t}, \dots, x_{k,t}$ можуть бути також несистемні змінні, введені із затримкою в часі. Випадковий

процес збурення $\{\varepsilon_t\}$ має нормальний розподіл з нульовим середнім та одиничною дисперсією, тобто $\{\varepsilon_t\} \sim N\{0, 1\}$.

Зазвичай приймається, що моделі зі змінною волатильністю можна поділити на класи: спостережувані та параметричні [222, 223]. Обидва класи моделей можна представити, використовуючи таку структуру:

$$y_t | z_t \sim N(\mu_t, \sigma_t^2).$$

Найпростішим прикладом моделей першого класу є модель авторегресії з умовною гетероскедастичністю (АРУГ або англійською ARCH – AutoRegressive Conditional Heteroskedasticity), наведена у роботі [259], та моделі узагальненої авторегресії з умовною гетероскедастичністю (УАРУГ або GARCH – Generalized AutoRegressive Conditional Heteroskedasticity) [248, 325, 355].

У другому класі моделей z_t є функцією неспостережуваних або прихованих компонент. Лог-нормальна модель стохастичної волатильності, запропонована Тейлором у 1986 р. [356], є найпростішим та найвідомішим прикладом:

$$y_t | h_t \sim N(0, \exp(h_t)),$$

$$h_t = \alpha + \beta h_{t-1} + \eta_t, \quad \eta_t \sim NID(0, \sigma^2),$$

де h_t є лог-волатильністю, яка є неспостережуваною, але може бути оцінена за допомогою спостережуваних даних; скорочення *NID* означає нормальні однаково розподілені значення (normally identically distributed).

За останні роки розроблена велика кількість узагальнень та модифікацій моделей УАРУГ і МСВ, що пояснюється розмаїтістю характеру нестационарностей досліджуваних процесів у фінансах, технологіях, медицині, біології та інших галузях досліджень.

Регресійні моделі для прогнозування волатильності

Моделі АРУГ та УАРУГ спрямовані на опис кластерів волатильності та пов'язаних з цим процесом ефектів, наприклад таких, як ексцес. Головна ідея моделей полягає у тому, що поточна волатильність залежить від попередніх реалізацій процесу, який описує поведінку фінансових активів та пов'язаних з ними гетероскедастичних процесів.

Модель УАРУГ

Для прогнозування волатильності серед моделей авторегресії еталонною є модель узагальненої авторегресії з умовною гетероскедастичністю (УАРУГ) [224]. Модель УАРУГ корисна при моделюванні волатильності, залежної від часу. Більшість фінансових часових рядів мають волатильність, з явно вираженим кластерним характером. Відповідно до цього волатильність буде високою, якщо на попередніх проміжках часу вона була високою, та волатильність буде низькою, якщо нещодавно вона була низькою. В умовах кластеризації волатильності моделі УАРУГ описують умовну дисперсію, яка

лінійно залежить від минулої поведінки квадратів різниць та ковзного середнього минулих умовних дисперсій. Модель УАРУГ(p, q) має вигляд:

$$y_t = \sigma_t \varepsilon_t, \quad t = 1, 2, \dots, T$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \dots + \alpha_p y_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_q \sigma_{t-q}^2,$$

де y_t – доходність фінансових активів на момент часу t ; σ_t – миттєва волатильність доходності на час t ; T – об'єм вибірки; α_i ($i=0, 1, \dots, p$) та β_i ($i=0, 1, \dots, q$) – параметри моделі повинні задовольняти обмеженню $\alpha_i > 0$, $\beta_i > 0$, зумовленому тим, що дисперсія завжди повинна мати додатне значення; p – максимальний порядок складової АРУГ; а q – максимальний порядок узагальнюючої складової.

Якщо $\sum_{i=1}^p \alpha_i < 1$ і $\sum_{i=1}^q \beta_i < 1$, то процес стаціонарний. Величину σ_t можна визначити, якщо відомі спостережувані значення $y_{t-1} \dots y_{t-n}$ та параметри моделі. Випадкова змінна ε_t розподілена нормально, а саме: $(\varepsilon_t \sim N(0, 1))$; при цьому $E_{t-1}(\varepsilon_t) = E_{t-1}(\varepsilon_t^2 - 1) = 0$. Для практичного використання запропоновано різноманітні варіації цього розподілу, наприклад: розподіл Стюдента [221], узагальнений розподіл похибок (generalized error) [318], узагальнений t -розподіл (generalized t -distribution) [223].

Модель ЕУАРУГ(p, q) (або EGARCH(p, q))

Експоненціальна модель УАРУГ (ЕУАРУГ) – це модель АРУГ, яка враховує асиметричний імпульс та волатильність хороших і поганих новин стосовно результатів виконання операцій на біржі. Модель ЕУАРУГ зі спеціальною змінною, яка розрізняє волатильність хороших та поганих новин, представлена у роботі [318]. Модель ЕУАРУГ(p, q) має наступний вигляд:

$$y_t = \sigma_t \varepsilon_t, \quad t = 1, 2, \dots, T$$

$$\log(\sigma_t^2) = \alpha_0 + \left[1 - \sum_{j=1}^q \beta_j B^j \right]^{-1} \left[1 + \sum_{i=1}^p \alpha_i B^i \right] g(\varepsilon_{t-1}),$$

де всі змінні і параметри визначаються так само, як у моделі АРУГ; крім того, B^j та B – лінійні оператори зсуву, для яких: $x_t B^j = x_{t-j}$, та $g(\varepsilon_t) = \lambda_1 \varepsilon_t + \lambda_2 (|\varepsilon_t| - E(|\varepsilon_t|))$, де λ_1 і λ_2 додаткові параметри.

У моделях ЕУАРУГ(p, q) при від'ємному коефіцієнті λ_1 імпульси поганих новин будуть мати більший вплив на волатильність, ніж імпульси хороших новин відповідної змінної.

Частково інтегрована узагальнена АРУГ (ЧІУАРУГ(p, q) або FIGARCH(p, q))

Для багатьох фінансових часових рядів характерна висока інертність (незмінність) волатильності на вибраних часових інтервалах. Важливим тут є

існування автокореляції різноманітних мір волатильності. Ця властивість називається «довгою пам'яттю» волатильності [209, 255, 305]. Розглянемо модель з «довгою пам'яттю», а саме частково інтегровану узагальнену АРУГ (FIGARCH = Fractionally Integrated GARCH) модель, запропоновану Бейлі у 1996 році [209]:

$$y_t = \sigma_t \varepsilon_t, \quad t = 1, 2, \dots, T$$

$$\sigma_t^2 = \alpha_0 \left[I - \sum_{j=1}^q \beta_j B^j \right]^{-1} + \left\{ I - \left[I - \sum_{j=1}^q \beta_j B^j \right]^{-1} \sum_{i=1}^p \alpha_i B^i (I - B)^d \right\} y_t^2,$$

де всі змінні та параметри визначенні як у моделях АРУГ, за винятком того, що B^j та B^i – оператори, для яких: $x_t B^j = x_{t-j}$ та d – дробовий параметр різниці.

Модель ЧІУАРУГ(p, q) описує особливість «довгої пам'яті» фінансової волатильності за рахунок параметра d . При $0 < d < 1$ умовна волатильність σ_t^2 буде повільно затухати за гіперболічною нормою, що є типовою прикметою «хорошої пам'яті» [209].

Частково інтегрована експоненціальна узагальнена АРУГ (ЧІЕУАРУГ(p, q) або FIEGARCH(p, q))

Для врахування асиметричних ефектів хороших та поганих новин і «довгої пам'яті» вперше була запропонована Боллерслевим та Мікельсеном модель ЧІЕУАРУГ(p, q) (FIEGARCH = Fractionally Integrated Exponential GARCH) [225]. Модель FIEGARCH має вигляд:

$$y_t = \sigma_t \varepsilon_t, \quad t = 1, 2, \dots, T$$

$$\log(\sigma_t^2) = \alpha_0 + \left[I - \sum_{j=1}^q \beta_j B^j \right]^{-1} (I - B)^{-d} \left[I + \sum_{i=1}^p \alpha_i B^i \right] g(\varepsilon_{t-1}),$$

де всі змінні та параметри визначені як і у моделях ЧІУАРУГ, за винятком того, що d – це дійсне число на проміжку $(-0,5; 0,5)$, та $g(\varepsilon_t) = \lambda_1 \varepsilon_t + \lambda_2 (|\varepsilon_t| - E(|\varepsilon_t|))$ з новими параметрами λ_1 та λ_2 . Варто відмітити, що на відміну від ЧІУАРУГ, в цій моделі параметр d не повинен задовольняти умові невід'ємності для того щоб відобразити особливість «довгої пам'яті». Параметр d має задовольняти умовам: $d < 1$ але $d \neq 0$ [225]. При $d = 0$ модель приймає вигляд звичайного ЕУАРУГ.

Частково інтегрована авторегресія з ковзним середнім (ЧІАРКС або ARFIMA)

У роботах [266, 275] представлена модель ЧІАРКС (ARFIMA = Fractionally Integrated ARMA). Це параметрична альтернатива для моделювання «довгої пам'яті» фінансових часових рядів. Модель ЧІАРКС розширює модель АРКС, застосовуючи дробовий параметр різниці для розрахунку високої стійкості часового ряду; прототип цієї моделі такий:

$$\Phi(B)(1 - B)^d y_t = \theta(B)\varepsilon_t,$$

де y_t – величина фінансового часового ряду на момент часу t ; d – дробовий параметр різниці; $\Phi(B) = 1 - \phi_1 B - L - \phi_p B^p$ та $\theta(B) = 1 - \theta_1 B - L - \theta_p B^p$ – відповідно авторегресійні та ковзні середні поліноми в операторі затримки B , та ε_t – білий шум.

Властивості процесу ЧІАРКС залежать від величини параметра d . Процес стаціонарний і має властивість оберненості при $|d| < 0,5$ та нестаціонарний при $|d| \geq 0,5$, оскільки має необмежену дисперсію. При $0 < d < 0,5$ процес демонструє «довгу пам'ять» або довгострокову залежність. При $-0,5 < d < 0$ процес має нестійкий характер або довгострокову негативну залежність. При $d = 0$ – коротку пам'ять (у цьому випадку ЧІАРКС редукується до АРКС). Параметри моделі ЧІАРКС можуть бути оцінені за методом максимальної правдоподібності [348], а k -кроковий майбутній прогноз y_t – можна розрахувати за авторегресійним представленням процесу [235].

Модель ЧІАРКС відрізняється від моделей ЧІУАРУГ та ЧІЕУАРУГ тим, що це модель з умовним середнім, в якій задане тільки одне рівняння, в той час як в ЧІУАРУГ та ЧІЕУАРУГ задано два рівняння (одне для умовного середнього, а друге для дисперсії). Для моделі ЧІАРКС необхідно спочатку згенерувати міру волатильності, а потім включити її у рівняння моделі. За міру волатильності використовують квадрат доходності. Ця стратегія моделювання відрізняється від моделей типу УАРУГ, в яких волатильність розглядається як неспостережувана величина, яка оцінюється за доходами відповідних активів.

Модель стохастичної волатильності

Розглянемо дві найбільш відомі моделі стохастичної волатильності: лог-нормальну модель та параметричну модель стохастичної волатильності.

Лог-нормальна модель стохастичної волатильності Тейлора [356] має вигляд:

$$\begin{aligned} y_t &= e^{\frac{h_t}{2}} \varepsilon_t, \quad \varepsilon_t \sim N(0, 1), \quad t = 1, \dots, T \\ h_t &= \alpha + \beta h_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2), \quad t = 1, \dots, T \end{aligned} \quad (A2.1)$$

де ε_t та η_t – незалежні гаусові процеси білого шуму з дисперсіями 1 та σ_η^2 , відповідно.

Модель називають лог-нормальною моделлю стохастичної волатильності, оскільки η_t – гаусів процес. Однією із інтерпретацій прихованої змінної h_t є її представлення як випадкового та нечіткого потоку нової інформації (частіше всього збурень), яку складно моделювати безпосередньо на фінансових ринках.

Іншою можливою інтерпретацією h_t є характеристика стану, в якому працює фінансовий ринок і який може бути описаний дискретною змінною величиною. Поширеним підходом до моделювання змін є клас марковських моделей, наведених Гамільтоном [268]. Модель тоді має вигляд:

$$y_t = e^{\frac{h_t}{2}} \varepsilon_t, \quad \varepsilon_t \sim N(0, 1), \quad t = 1, \dots, T,$$

$$h_t = \alpha + \beta s_t, \quad t = 1, \dots, T$$

де s_t – ланцюг Маркова з двома станами, який може приймати одне з двох значень 0 і 1 та не залежить від ε_t . Значення часового ряду s_t , для всіх моментів часу t , залежить тільки від останнього значення s_{t-1} , для $i, j = 0, 1$:

$$P(s_t = j | s_{t-1} = i, s_{t-2}, \dots) = P(s_t = j | s_{t-1} = i) = p_{ij}.$$

Ймовірності $(p_{ij})_{i,j=0,1}$ називають ймовірностями переходу з одного стану до іншого (із попереднього в наступний), їх подають у вигляді перехідної матриці P :

$$\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix},$$

яка повністю описує ланцюг Маркова; при цьому виконується рівність:

$$p_{00} + p_{01} = p_{10} + p_{11} = 1.$$

Ланцюг Маркова з двома можливими станами можна представити простим процесом авторегресії AP(1):

$$s_t = (1 - p_{00}) + (-1 + p_{00} + p_{11})s_{t-1} + v_t,$$

де $v_t = s_t - E(s_t | s_{t-1}, s_{t-2}, \dots)$. Рівняння волатильності можна записати:

$$h_t = \alpha + \beta s_t = \alpha + \beta[(1 - p_{00}) + (-1 + p_{00} + p_{11})s_{t-1} + v_t]$$

або

$$h_t = (2 - p_{00} - p_{11}) + \beta(1 - p_{00}) + (-1 + p_{00} + p_{11})h_{t-1} + \beta v_t = \alpha + bh_{t-1} + \omega_t, \quad (\text{A2.2})$$

яке має таку саму структуру МСВ, але з шумом, який може приймати лише скінчену множину значень.

Модель стохастичної волатильності у неперервній формі

Ідея МСВ полягає у тому, що на волатильність впливає прихований процес, представлений потоком інформації стосовно цін. Лог-нормальна МСВ, запропонована Тейлором [355], є альтернативою моделям авторегресії з умовною гетероскедастичністю (АРУГ), запропонованим Енглором [259]. МСВ є реалістичнішою та гнучкішою на відміну від моделей АРУГ, оскільки вона містить два випадкових процеси: один для спостережуваних даних y_t і один для прихованої волатильності h_t . Розглянемо МСВ з незалежними гаусовими процесами білого шуму ε_t та η_t , що описані формулою (A.2.1), де y_t є спостережуваною в момент t . Лог-волатильність h_t відповідає стаціонарному процесу авторегресії AP(1) при $|\beta| < 1$. В основній моделі ε_t та η_t – це незалежні нормально розподілені випадкові процеси. Збурення спостережень ε_t зумовлені похибками формування вибірки, а збурення η_t впливає на зміни у

динаміці волатильності, де β – міра сталості прихованого процесу волатильності h_t . Нестандартна інтерпретація МСВ розглянута в роботі [225], в якій волатильність поділяють на основну, представлену щоденною волатильністю, та волатильністю, спричиненою коливаннями кількості інформації, яку отримують ринки кожний день. З (А.2.1) видно, що, на відміну від моделі УАРУГ, волатильність на один період вперед – це не спостережувана величина.

Властивості моделі (А.2.1) розглянуто в роботі [255]. У порівнянні з відповідними моделями АРУГ властивості цієї моделі вирізнити легше, але оцінювання її параметрів виконується значно важче; розглянемо деякі з них.

Оскільки η_t – гаусів процес, то і h_t описується стандартною гаусовою авторегресією. Вона буде стаціонарною, якщо $|\beta| < 1$ та

$$\mu_h = E(h_t) = \frac{\alpha}{1-\beta},$$

$$\sigma_h^2 = Var(h_t) = \frac{\sigma_\eta^2}{1-\beta^2}.$$

Оскільки процес ε_t завжди стаціонарний, то y_t буде стаціонарним тоді і тільки тоді, коли процес h_t – стаціонарний. Процес y_t , який є добутком двох стаціонарних процесів, також стаціонарний. Використовуючи властивості лог-нормального розподілу, можна показати, що всі моменти існують тоді, коли h_t стаціонарний процес та ексцес

$$\frac{E(y_t^4)}{(E(y_t^2))^2} = 3 \exp(\sigma_h^2) \geq 3,$$

який показує, що МСВ має товщі «хвости», ніж у відповідному нормальному розподілі, а всі непарні моменти дорівнюють нулю.

Динамічні властивості y_t легко визначити. Оскільки ε_t – незалежна випадкова величина, то y_t – маргінальна різниця та білий шум, якщо $|\beta| < 1$.

Оскільки h_t – гаусів процес АР(1), то

$$\begin{aligned} Cov(y_t^2, y_{t-r}^2) &= E(y_t^2 y_{t-r}^2) - (E(y_t^2))^2 \\ &= E(\exp(h_t + h_{t-r})) - (E(\exp(h_t)))^2 \\ &= \exp(2\mu_h + \sigma_h^2) (\exp(\sigma_h^2 \beta^r) - 1) \end{aligned}$$

та

$$\rho_{y_t^2}(r) = \frac{Cov(y_t^2, y_{t-r}^2)}{Var(y_t^2)} = \frac{\exp(\sigma_h^2 \beta^r) - 1}{3 \exp(\sigma_h^2) - 1} \cong \frac{\exp(\sigma_h^2) - 1}{3 \exp(\sigma_h^2) - 1} \beta^r.$$

Пам'ять величини y_t визначається пам'яттю прихованої величини h_t стосовно процесу авторегресії АР(1). Якщо $\beta < 0$, то $\rho_{y_t^2}(r)$ може бути

від'ємною на відміну від моделей АРУГ; оскільки це автокореляційна функція процесу АРКС(1,1), то МСВ поводить себе аналогічно до моделі УАРУГ(1,1). Немає ніякої необхідності ні в обмеженнях на невід'ємність, ні в обмеженнях ексцесу на коефіцієнти, тобто це велика перевага відносно моделей УАРУГ.

Динамічні властивості МСВ краще спостерігати при піднесенні до квадрату та знаходження логарифму; маємо лінійний процес

$$\begin{cases} \log(y_t^2) = h_t + \log(\varepsilon_t^2) \\ h_t = \alpha + \beta h_{t-1} + \eta_t \end{cases},$$

у якому до авторегресійного процесу h_t додаються незалежні випадкові величини $\log(\varepsilon_t^2)$. В результаті маємо, що $\log(y_t^2) \sim ARMA(1,1)$. Якщо ε_t – гаусів процес, то $\log(\varepsilon_t^2)$ має середнє $-1,27$ та дисперсію $4,93$ [286], але цей розподіл далекий від того, щоб бути нормальним, оскільки він має велику асиметрію з довгим лівим «хвостом», обумовленим взяттям логарифмів маленьких чисел. Автокореляційна функція для $\log(y_t^2)$ має такий вигляд:

$$\rho_{\log(y_t^2)}(r) = \frac{\beta^r}{1 + \frac{4,93}{\sigma_h^2}}.$$

Модифікована лог-нормальна модель стохастичної волатильності

У дослідженні автором запропонована модифікована лог-нормальна модель стохастичної волатильності такого вигляду:

$$\begin{aligned} y_t &= e^{\frac{h_t}{2}} \varepsilon_t, \quad \varepsilon_t \sim N(0,1), \quad t = 1, \dots, T \\ h_t &= \alpha + \beta_1 h_{t-1} + \dots + \beta_p h_{t-p} + \gamma y_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2), \quad t = 1, \dots, T \end{aligned}$$

де $h_t = \frac{1}{w} \sum_{l=t-\frac{w-1}{2}}^{t+\frac{w-1}{2}} [y_l - \bar{y}]^2$ – умовна дисперсія, яка характеризує динаміку

дисперсії основного досліджуваного ряду фінансових даних; \bar{y} – вибіркове середнє для вікна даних розміром w . Запропонована модель враховує необхідну глибину пам'яті процесу, сформованого із значень умовної дисперсії h_t , а також значення основної змінної y_t , які забезпечують врахування впливу різнознакових збурень на умовну дисперсію. Апробація запропонованої моделі свідчить про її високу ступінь адекватності і можливість оцінювання прогнозів умовної дисперсії необхідної для оцінювання фінансових ризиків якості.

МСВ з дискретною волатильністю

Розглянемо ланцюг Маркова з двома можливими станами S_t , який не залежить від гаусової величини білого шуму ε_t . Припустимо, що процес

стаціонарний, тоді безумовні ймовірності у станах 0 ($P(s_t=0)=\pi_0$) та 1 ($P(s_t=1)=\pi_1$) визначаються таким чином:

$$\begin{cases} \pi_0 = p_{00}\pi_0 + (1-p_{11})\pi_1 \\ \pi_1 = (1-p_{00})\pi_0 + p_{11}\pi_1 \end{cases}$$

З $\pi_0 + \pi_1 = 1$, або у векторній формі:

$$\begin{cases} \pi = P\pi \\ 1'\pi = 1 \end{cases}$$

де $1 = (1,1)'$; тому отримуємо:

$$\begin{aligned} \pi_0 &= \frac{1-p_{11}}{2-p_{00}-p_{11}}; \\ \pi_1 &= \frac{1-p_{00}}{2-p_{00}-p_{11}}. \end{aligned}$$

При $p_{00} + p_{11} > 1$ з рівняння (А.2.2) випливає, що процес h_t залишається у поточному стані і має бути корельованим, тоді:

$$E(h_t) = \alpha + \beta E(s_t) = \alpha + \beta\pi_1$$

$$Var(h_t) = \beta^2 \pi_1(1-\pi_1)$$

За умови стаціонарності процесу, як і для МСВ з неперервною волатильністю, всі моменти існують, всі непарні моменти дорівнюють нулю, а ексцес визначається за виразом:

$$\frac{E(y_t^4)}{(E(y_t^2))^2} = 3 \frac{(\pi_0 + \exp(2\beta)\pi_1)}{(\pi_0 + \exp(\beta)\pi_1)^2} \geq 3.$$

Крім того, оскільки ε_t – незалежна випадкова величина, то y_t – це маргінальна різниця, а її динамічні властивості описуються коваріацією квадратів:

$$\begin{aligned} Cov(y_t^2, y_{t-r}^2) &= E(y_t^2 y_{t-r}^2) - (E(y_t^2))^2 \\ &= E(\exp(h_t + h_{t-r})) - (\exp(\alpha)\pi_0 + \exp(\alpha + \beta)\pi_1)^2 \\ &= \exp(2\alpha)P(s_t=0, s_{t-r}=0) + \exp(2\alpha + \beta)P(s_t=0, s_{t-r}=1) + \\ &+ \exp(2\alpha + \beta)P(s_t=1, s_{t-r}=0) + \exp(2\alpha + 2\beta)P(s_t=1, s_{t-r}=1) - \\ &- (\exp(\alpha)\pi_0 + \exp(\alpha + \beta)\pi_1)^2 \end{aligned}$$

де вектор безумовних спільних ймовірностей $P(s_t, s_{t-r})$ можна розрахувати таким чином:

$$P(s_t, s_{t-r}) = P(s_t | s_{t-r})P(s_{t-r}) = P^r \pi,$$

де

$$P^r = \begin{bmatrix} \frac{(1-p_{11}) + \lambda^r(1-p_{00})}{2-p_{00}-p_{11}} & \frac{(1-p_{11}) + \lambda^r(1-p_{11})}{2-p_{00}-p_{11}} \\ \frac{(1-p_{00}) + \lambda^r(1-p_{00})}{2-p_{00}-p_{11}} & \frac{(1-p_{00}) + \lambda^r(1-p_{11})}{2-p_{00}-p_{11}} \end{bmatrix},$$

та $\lambda = -1 + p_{00} + p_{11}$.

Волатильність h_t є сама по собі марковським ланцюгом, який може приймати значення α та $\alpha + \beta$ з такою самою матрицею переходу P .

Порівняння особливостей моделі стохастичної волатильності та моделі УАРУГ та результати моделювання

На відміну від моделей типу УАРУГ, які обмежують залежну від часу волатильність детерміністичною залежністю від минулої інформації, МСВ трактує волатильність як не спостережувану компоненту, яка супроводжує стохастичний процес. Крім минулої інформації, велика кількість інших факторів, зокрема економічна і політична інформація та поведінка інвесторів, також впливають на волатильність. Тому допускають, що волатильність містить деякі випадкові компоненти. МСВ явно включає не спостережуваний випадковий імпульс в описі динаміки волатильності і дає можливість волатильності відповідати деякому прихованому ймовірнісному процесу.

Теоретичне порівняння моделі УАРУГ та МСВ представлено у роботі [242]. Розглядається можливість цих моделей описувати три особливості зазвичай присутніх у фінансових часових рядах:

- високий ексцес;
- малий перший порядок автокореляції квадратів спостережень;
- висока стійкість автокореляції квадратів залишків.

Автокореляція квадратів доходностей в моделі УАРУГ(1,1) показує той самий шаблон як і процес АРКС(1,1) з параметрами авторегресії $\alpha + \beta$ [226]. У роботі [242] наведено співвідношення між автокореляцією першого порядку та стійкістю моделі УАРУГ. Виявилось, що для даного ексцесу велика стійкість пов'язана з меншим першим порядком автокореляції. Також показано, що в моделі УАРУГ(1,1) неможливо одночасно отримати низькі значення автокореляції першого порядку та високий ексцес, і що високий ексцес можна отримати тільки тоді, коли $\alpha + \beta$ прямує до одиниці. Стосовно МСВ показано, що вона є більш гнучкою, ніж модель УАРУГ(1,1) у тому сенсі, що вона може поєднувати високий ексцес та перший порядок автокореляції без наближення волатильності до нестационарної області [242]. Показано, що використання t -розподілу Стьюдента збільшує гнучкість МСВ, що дозволяє згенерувати вищий ексцес без збільшення порядку автокореляції вище першого.

Побудова функції прогнозування на основі МСВ

Різницеве рівняння, яке описує зміну значення волатильності у часі, дає можливість прогнозувати на один крок вперед, якщо у рівняння підставити

поточне значення волатильності та оцінки відповідних параметрів. Для прогнозування на декілька кроків вперед потрібно побудувати функцію прогнозування.

Змінну у часі волатильність описує друге рівняння МСВ, яке формально є авторегресією першого порядку:

$$h(k+1) = \mu + \phi(h(k) - \mu) + \sigma_{\eta}\eta(k),$$

або

$$h(k+1) = \mu(1 - \phi) + \phi h(k) + \sigma_{\eta}\eta(k). \quad (\text{A.2.3})$$

Оцінки параметрів моделі вважаємо відомими, докладно процедура їх оцінювання описана у [15,16]. Оцінка волатильності визначається умовним математичним сподіванням, яке визначається на основі відомої інформації на момент часу k , включно:

$$\hat{h}(k+1) = E_k[h(k+1)] = E[\mu(1 - \phi) + \phi h(k) + \sigma_{\eta}\eta(k)] = \mu(1 - \phi) + \phi h(k).$$

Аналогічно запишемо рівняння (A.2.3) для моменту $k+2$:

$$h(k+2) = \mu(1 - \phi) + \phi h(k+1) + \sigma_{\eta}\eta(k)$$

і знайдемо умовне математичне сподівання:

$$\begin{aligned} \hat{h}(k+2) &= E_k[h(k+2)] = \mu(1 - \phi) + \phi E[h(k+1)] = \\ &= \mu(1 - \phi) + \phi E[\mu(1 - \phi) + \phi h(k)] = \\ &= \mu(1 - \phi) + \mu\phi(1 - \phi) + \phi^2 h(k) = \\ &= \mu(1 - \phi)(1 + \phi) + \phi^2 h(k). \end{aligned}$$

Для наступного моменту часу отримаємо вираз:

$$\begin{aligned} \hat{h}(k+3) &= E_k[h(k+3)] = \mu(1 - \phi) + \phi E[h(k+2)] = \\ &= \mu(1 - \phi) + \phi E[\mu(1 - \phi)(1 + \phi) + \phi^2 h(k)] = \\ &= \mu(1 - \phi) + \mu\phi(1 - \phi)(1 + \phi) + \phi^3 h(k) = \\ &= \mu(1 - \phi)(1 + \phi + \phi^2) + \phi^3 h(k). \end{aligned}$$

Таким чином, функція прогнозування на довільну кількість кроків має вигляд:

$$\hat{h}(k+t) = \mu(1 - \phi)(1 + \phi + \phi^2 + \dots + \phi^{t-1}) + \phi^t h(k).$$

Зазначимо, що h_t – це логарифм квадрату волатильності, тобто $h_t = \ln \sigma_t^2$.

Узагальнена модель авторегресії з умовною гетероскедастичністю

МСВ відноситься до класу параметричних моделей зі змінною волатильністю. Іншим класом моделей зі змінною волатильністю є регресійні моделі, серед яких найбільш популярною є узагальнена модель авторегресії з УАРУГ. Модель УАРУГ у багатьох випадках достатньо ефективна стосовно моделювання та короткострокового прогнозування волатильності, яка залежить від часу. Модель УАРУГ(p, q) визначає залежність умовної дисперсії від минулої поведінки квадратів значень основної змінної та ковзного середнього минулих умовних дисперсій.

Побудова функції прогнозування для моделі УАРУГ

Можна представити модель УАРУГ(p, q) у вигляді процесу авторегресії з ковзним середнім (АРКС або ARMA в англійській термінології) [260]:

$$\varepsilon_t^2 = \omega + \sum_{j=1}^{\max(p, q)} (\alpha_j + \beta_j) \varepsilon_{t-j}^2 + \left(\nu_t - \sum_{i=1}^p \beta_i \nu_{t-i} \right).$$

Враховуючи, що $\varepsilon_t^2 = \sigma_t^2 + \nu_t$, де $E_{t-1}[\nu_t] = 0$, $\nu_t \in [-\sigma_t^2, \infty)$, тобто $\varepsilon_t^2 \sim ARMA(m, p)$, де $m = \max(p, q)$, тепер можна записати:

$$\sigma_{t+k}^2 = \omega + \sum_{i=1}^n (\alpha_i \varepsilon_{t+k-i}^2 + \beta_i \sigma_{t+k-i}^2) + \sum_{i=k}^m (\alpha_i \varepsilon_{t+k-i}^2 + \beta_i \sigma_{t+k-i}^2), \text{ де } n = \min\{m, k-1\}.$$

Оцінку прогнозу дисперсії на k кроків визначимо за умовним математичним сподіванням:

$$E_t[\sigma_{t+k}^2] = \omega + \sum_{i=1}^n ((\alpha_i + \beta_i) E_t[\sigma_{t+k-i}^2]) + \sum_{i=k}^m (\alpha_i \varepsilon_{t+k-i}^2 + \beta_i \sigma_{t+k-i}^2).$$

Для УАРУГ(1,1) при $k > 2$ маємо:

$$\begin{aligned} E_t[\sigma_{t+k}^2] &= \sum_{i=0}^{k-2} (\alpha_1 + \beta_1)^i \omega + (\alpha_1 + \beta_1)^{k-1} \sigma_{t+1}^2 = \\ &= \omega \frac{(1 - (\alpha_1 + \beta_1)^{k-1})}{(1 - (\alpha_1 + \beta_1))} + (\alpha_1 + \beta_1)^{k-1} \sigma_{t+1}^2 = \\ &= \sigma^2 (1 - (\alpha_1 + \beta_1)^{k-1}) + (\alpha_1 + \beta_1)^{k-1} \sigma_{t+1}^2 = \\ &= \sigma^2 + (\alpha_1 + \beta_1)^{k-1} (\sigma_{t+1}^k - \sigma^2). \end{aligned}$$

Як зазначено в [192], задача прогнозування майбутніх значень y_{n+m}^2 зводиться до задачі прогнозування «волатильності» σ_{n+m}^2 за результатами попередніх спостережень y_0, y_1, \dots, y_n .

Приклад побудови моделей МСВ та УАРУГ і результати моделювання

Вхідними даними для процедури прогнозування є параметри відповідної моделі та початкове значення волатильності. У роботі [15] докладно описано процедуру отримання оцінок МСВ та моделі УАРУГ. Оцінки параметрів МСВ та моделі УАРУГ отримані за допомогою процедури, реалізованої на мові Java, яка містить кілька алгоритмів. Всі оцінки параметрів МСВ та моделі УАРУГ отримані на основі даних обмінного курсу валют фунт/долар.

Алгоритми оцінювання параметрів МСВ ґрунтуються на методі Монте-Карло для марковських ланцюгів, в основу якого покладено алгоритм генерування псевдовипадкових послідовностей Гіббса [16]. Оцінки параметрів, отримані після виконання однокрокового алгоритму Гіббса та за методом «змішаного зсуву», наведені у таблиці А.2.1.

Оцінки параметрів моделі УАРУГ отримані за допомогою процедури, яка ґрунтується на двох алгоритмах: алгоритмах: Гастінгса-Метрополіса та

алгоритму «адаптивного відбраковування» (adaptive rejection). Реалізована програма повертає значення оцінок параметрів у такому вигляді: $\frac{\alpha_0}{1 - \alpha_1 - \alpha_2}$ – усталена умовна дисперсія; $\alpha_1 + \alpha_2$ – показник сталості (persistence); $\frac{\alpha_2}{\alpha_1 + \alpha_2}$ – коефіцієнт ковзного середнього (КС). Оцінки параметрів наведені у таблиці А.2.2.

Таблиця А.2.1

Значення параметрів $\beta = e^{\frac{\mu}{2}}$, ϕ та σ_v^2 за результатами моделювання (процес – курс валют фунт/долар)

	ϕ	$\beta = e^{\frac{\mu}{2}}$	σ_v^2
Однокроковий алгоритм Гіббса	0,97717	0,76300	0,16737
Алгоритм Гіббса на основі процедури «змішаного зсуву»	0,94040	0,68535	0,32299

Таблиця А.2.2

Значення параметрів $\frac{\alpha_0}{1 - \alpha_1 - \alpha_2}$, $\alpha_1 + \alpha_2$ та $\frac{\alpha_2}{\alpha_1 + \alpha_2}$ за результатами моделювання (процес – курс валют фунт/долар)

	$\frac{\alpha_0}{1 - \alpha_1 - \alpha_2}$	$\alpha_1 + \alpha_2$	$\frac{\alpha_2}{\alpha_1 + \alpha_2}$
Параметри моделі УАРУГ, отримані за процедурою AR	0,66585	0,90268	0,97506

За початкове значення волатильності взято останнє значення волатильності у вибірці, яка була використана для оцінювання параметрів моделей. Значення волатильності у попередні моменти часу розраховувались при визначенні оцінок параметрів МСВ, зокрема у алгоритмі Гіббса. Але, оскільки значення доходності у попередні проміжки часу відомі, то застосуємо простіший (за допомогою рухомого вікна) спосіб визначення волатильності у поточний момент часу за умови відомих значень доходності:

$$h_t = \frac{1}{w-1} \sum_{i=k-\frac{w-1}{2}}^{k+\frac{w-1}{2}} [y_i - \bar{y}]^2,$$

де y_i – доходність в момент часу i ; \bar{y} – середнє значення доходності, і нехай ширина рухомого вікна складає $w = 3$.

Результати прогнозування. Застосуємо отримані оцінки параметрів моделей для прогнозування майбутнього значення волатильності фінансового ринку за допомогою розробленої програми. Існує два підходи до процедури прогнозування: статичне прогнозування та динамічне прогнозування. На

першому кроці багатокрокового прогнозування ці підходи до прогнозування мають завжди давати ідентичні результати.

Статичне прогнозування представляє собою процедуру однокрокового прогнозування. Виконавши статичне прогнозування послідовно на кожному кроці, можна порівняти прогноз на основі моделі УАРУГ та реальні спостережувані значення дохідності. Нижче наведено результати статичного прогнозування для моделі УАРУГ (табл. А.2.3).

Таблиця А.2.3

Результати статичного прогнозування значень волатильності (процес – курс валют фунт/долар)

Реальні значення	Прогноз на основі моделі УАРУГ
1,11153	0,12694
1,13027	1,30501
0,64194	0,47261
2,07543	1,71567
1,57441	1,88263
0,43457	0,26463
0,90577	0,70775
0,56025	0,67875
0,03745	0,05704
1,54986	1,41828
1,95121	2,49299
0,71724	0,73137
0,43737	0,51395
0,17528	0,21662
0,21510	0,18647
...	...

Результати розрахунків показників якості однокрокового прогнозу та якості моделі для УАРУГ, Е-УАРУГ та МСВ наведено у таблиці А.2.4.

Таблиця А.2.4

Характеристики оцінок прогнозів та моделей

Тип моделі	Характеристики прогнозу			Характеристики моделі		
	СКП	САПП	Коеф-т Тейла	R^2	$\sum e^2$	DW
УАРУГ	0,7473	58,1307	0,3361	0,9136	509,975	2,622
Е-УАРУГ	0,3487	9,2195	0,1357	0,9321	239,384	2,217
МСВ	0,9538	11,7953	0,1254	–	–	–

Динамічне прогнозування представляє собою багатокрокове прогнозування. Нехай процедура прогнозування виконується покроково на 30 кроків вперед ($n_{forecast} = 30$). Прогнозування волатильності виконується для МСВ та моделі УАРУГ. Оцінювання параметрів МСВ виконано за двома методами: алгоритмом Гіббса та методом «змішаного зсуву». Результати

роботи програми для різних оцінок параметрів [12] наведені у таблиці А.2.5 для перших десяти кроків.

Таблиця А.2.5

Прогнозовані значення волатильності на 30 кроків вперед (курс фунт/долар)

Реальні значення	Прогноз на основі МСВ (алгоритм Гіббса)	Прогноз на основі МСВ («змішаного зсуву»)	Прогноз на основі моделі УАРУГ
0,42162	0,80763	0,89863	1,12763
0,51330	0,78393	0,73842	1,20763
2,06698	1,46128	1,57709	1,10737
2,58180	1,80963	2,32253	1,62360
0,74758	1,11891	0,97385	0,95361
0,42659	0,69900	0,63028	0,89512
0,48152	0,59010	0,73116	0,84625
0,24337	0,36191	0,45595	0,59542
0,59509	0,64448	0,72417	0,78130
0,82138	1,02777	0,79542	0,77280

На рис. А.2.1 зображено поведінку волатильності у часі, на основі якої виконувалось оцінювання параметрів моделей та прогнозована поведінка волатильності на $n_{forecast}$ кроків вперед.

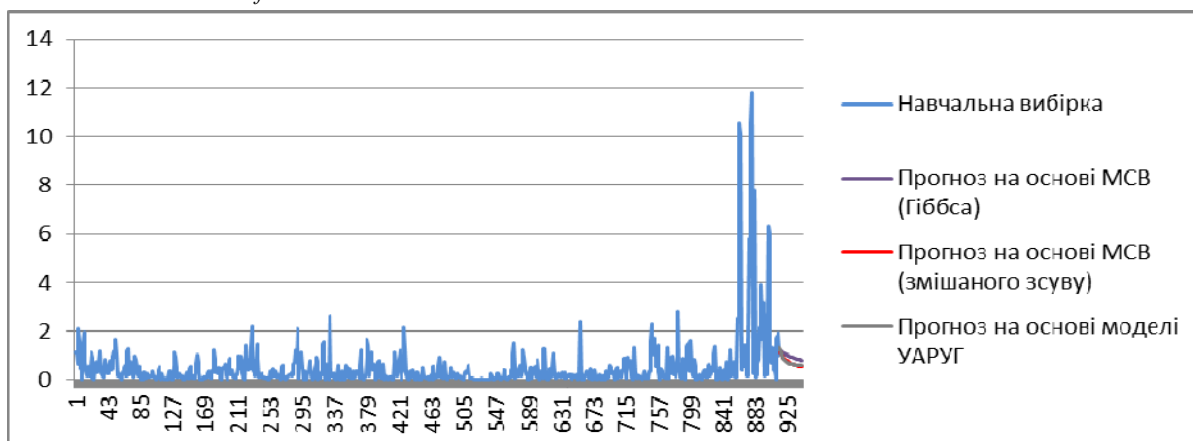


Рис. А.2.1 Значення волатильності у попередні моменти часу та прогноз на $n_{forecast}$ кроків вперед.

Для виконання обчислювальних експериментів [12] були також використані дані за цінами акції компанії Microsoft, отримані із бази даних системи Yahoo Finance (таблиця А.2.6).

Отже, найкращий результат прогнозування на один крок волатильності цін акцій компанії Microsoft був отриманий за допомогою експоненційної моделі авторегресії з умовною гетероскедастичністю (САПП = 4,75% на тестовій вибірці у 30 значень). Незадовільні за якістю були отримані результати для АРУГ, яка має найпростішу структуру і виявилась неадекватною для процесу моделювання.

Таблиця А.2.6

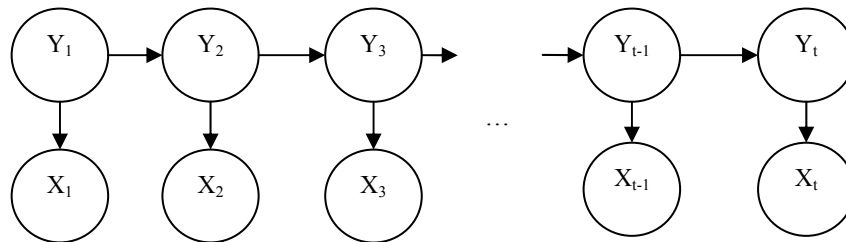
Результати прогнозування волатильності цін акцій компанії Microsoft

Тип моделі	Характеристики оцінок прогнозів на навчальній вибірці			Характеристики оцінок прогнозів на тестовій вибірці		
	СКП	САПП	Коеф-т Тейла	СКП	САПП	Коеф-т Тейла
АРУГ	1,2756	2387,7	0,389	1,754	3241,2	0,453
УАРУГ	0,549	24,73	0,115	0,691	33,35	0,138
Е-УАРУГ	0,436	3,45	0,017	0,073	4,75	0,022
МСВ	0,517	5,35	0,031	0,295	7,64	0,039

А.3. Динамічні мережі Байєса

Нехай приховані стани моделі представлені набором N_h випадкових дискретних змінних $Y_t^{(i)}$, $i \in \{1, \dots, N_h\}$ в момент часу t (рис. А.3.1). Аналогічно, спостереження представлені як N_o дискретних змінних $X_t^{(j)}$, $j \in \{1, \dots, N_o\}$. Динамічну мережу Байєса (ДМБ) можна визначити як пару звичайних МБ (B_1, B_2) [144].

Перша мережа B_1 описує модель спостережень в рамках одного часового про шарку $P(X_t / Y_t)$ і задає початковий розподіл $P(Y_1)$. Друга мережа B_2 описує модель часових переходів в рамках двох (або більше) часових про шарків $P(Y_t / Y_{t-1})$.

Рис. А.3.1 Приклад простої ДМБ для t часових про шарків

Обидві моделі спостережень і часових переходів задаються добутком своїх умовних розподілів:

$$P(Z_t / Z_{t-1}) = \prod_{i=1}^N P(Z_t^{(i)} / \pi(Z_t^{(i)})), \quad (\text{А.3.1})$$

де $Z_t^{(i)}$ – i -а вершина (що може бути прихованою, або спостережуваною), $N = N_h + N_o$, а $\pi(Z_t^{(i)})$ – батьківська множина вершини $Z_t^{(i)}$, елементи якої знаходяться або в поточному, або в попередньому часовому про шарку. Спільний розподіл ланцюжка довжиною T отримується розширенням мережі на T часових про шарків [288] і перемноженням всіх умовних ймовірностей:

$$P(Z_{1:T}^{(1:N)}) = \prod_{i=1}^N P_{B_1}(Z_1^{(i)} / \pi(Z_1^{(i)})) \times \prod_{t=2}^T \prod_{i=1}^N P_{B_2}(Z_t^{(i)} / \pi(Z_t^{(i)})) \quad (\text{А.3.2})$$

де $P_{B_1}(Z_1^{(i)} / \pi(Z_1^{(i)}))$ – умовна ймовірність в мережі B_1 , а $P_{B_{\rightarrow}}(Z_t^{(i)} / \pi(Z_t^{(i)}))$ – умовна ймовірність в мережі B_{\rightarrow} . Таким чином, мережі B_1 і B_{\rightarrow} вважаються незалежними одна від одної.

Для ДМБ, що зображена для прикладу на рис. А.3.1, спільний розподіл задається наступним чином:

$$P(Y_{1:T}, X_{1:T}) = P(Y_1)P(X_1 / Y_1) \times \prod_{t=2}^T P(Y_t / Y_{t-1})P(X_t / Y_t).$$

Ключовими моментами для ДМБ є навчання і логічний висновок.

Для динамічних мереж Байєса існує декілька проблем, пов'язаних із **логічним висновком**, зокрема:

- *Фільтрація*. Обчислення $P(Y_t / X_{1:t})$, контроль поточного стану із часом;
- *Прогноз*. Обчислення $P(Y_{t+h} / X_{1:t})$ для деякого $h > 0$ в майбутньому;
- *Фіксоване згладжування (в реальному часі)*. Обчислення $P(Y_{t-l} / X_{1:t})$, тобто оцінювання того, що сталося $l > 0$ кроків в минулому, маючи всі дані до тепер;
- *Згладжування за інтервалом*. Обчислення $P(Y_t / X_{1:T})$. Використовується як додаток при навчанні;
- *Декодування (за Вітербі)*. Обчислення $\arg \max_{Y_{1:t}} P(Y_{1:t} / X_{1:t})$, тобто пошук найбільш ймовірного пояснення даним, що спостерігається;
- *Класифікація*. Обчислення $P(X_{1:t}) = \sum_{Y_{1:t}} P(X_{1:t}, Y_{1:t})$. Може бути використане як обчислення правдоподібності послідовності спостережень для різних моделей.

Серед усіх можливих типів логічного висновку, перш за все, цікавою є фільтрація, тобто додаткове оцінювання поточного стану. Нехай Y_t – єдина прихована змінна, а $X_t^{(1)}, \dots, X_t^{(n)}$ – набір змінних, що спостерігаються. Прихована змінна має зв'язки лише із змінними, що спостерігаються у власному часовому прошарку, а самі спостережувані змінні пов'язані між собою у поточному часовому прошарку t і із спостережуваними змінними попереднього часового прошарку $t-1$. Тоді:

$$P(Y_T / X_{1:T}) = \frac{P(Y_T, X_{1:T})}{P(X_{1:T})} = \frac{\sum_{Y_{1:T-1}} P(Y_{1:T}, X_{1:T})}{\sum_{Y_{1:T}} P(Y_{1:T}, X_{1:T})} =$$

$$= \frac{\sum_{Y_{1:T-1}} P_{B_1}(Y_1 / \pi(Y_1)) \prod_{i=1}^n P_{B_1}(X_1^{(i)} / \pi(X_1^{(i)})) \prod_{t=2}^T (P_{B_{\rightarrow}}(Y_t / \pi(Y_t)) \prod_{i=1}^n P_{B_{\rightarrow}}(X_t^{(i)} / \pi(X_t^{(i)})))}{\sum_{Y_{1:T}} P_{B_1}(Y_1 / \pi(Y_1)) \prod_{i=1}^n P_{B_1}(X_1^{(i)} / \pi(X_1^{(i)})) \prod_{t=2}^T (P_{B_{\rightarrow}}(Y_t / \pi(Y_t)) \prod_{i=1}^n P_{B_{\rightarrow}}(X_t^{(i)} / \pi(X_t^{(i)})))}, \text{ де}$$

$\sum_{Y_{1:T-1}}(\cdot)$ означає маржиналізацію по всіх можливих значеннях змінної Y_t за проміжок часу $1:T-1$.

Застосування ДМБ для класифікації

Нехай o_1, \dots, o_m – множина об’єктів, що мають спільну множину ознак $X^{(1)}, \dots, X^{(n)}$. Кожна ознака $X^{(i)}$ спостережувана і має дискретне значення $\{x_{i,1}, \dots, x_{i,r_i}\}$. Також кожна ознака відповідає вершині МБ. Додаємо нову вершину $X^{(0)}$, що приймає наступні стани: $\{x_{0,1} = o_1, \dots, x_{0,r_0} = o_m\}$.

Нехай послідовність векторів $\vec{z}_1, \vec{z}_2, \dots, \vec{z}_T$, де $\vec{z}_i = \{z_i^{(1)}, \dots, z_i^{(n)}\}$ є послідовностями спостережень (у часі) за одним із об’єктів o_1, \dots, o_m . Тоді об’єкт, котрому належать ці спостереження, знаходиться наступним чином:

$$o_k = \arg \max_i P(X_T^{(0)} = o_i / X_1 = \vec{z}_1, \dots, X_T = \vec{z}_T) \tag{A.3.3}$$

На кожному кроці $l : T$ очікується побачити об’єкт o_k , тобто:

$$o_k = \arg \max_i P(X_l^{(0)} = o_i, \dots, X_T^{(i)} = o_i / X_l = \vec{z}_l, \dots, X_T = \vec{z}_T).$$

Має значення лише момент часу T , тому в А.3.3 проводиться маржиналізація по всіх об’єктах в моменти часу $l : T - 1$.

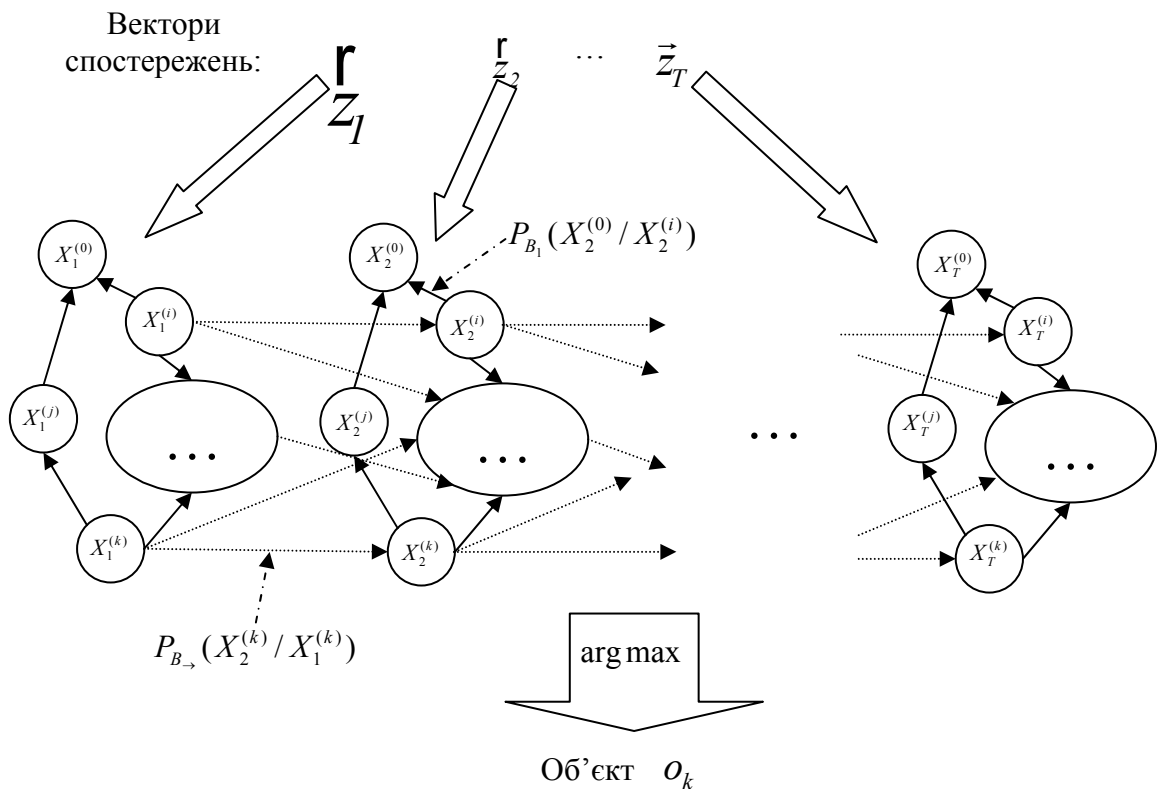


Рис. А.3.2 Схематичне зображення розв’язку задачі класифікації

Навчання ДМБ

Задача навчання МБ і ДМБ: із наборів спостережень необхідно визначити топологію мережі і усі таблиці умовних ймовірностей. Для ДМБ задача ускладнюється, оскільки існує часова компонента даних.

Для задачі класифікації нехай для кожного об'єкту $o_i \in \{o_1, \dots, o_m\} \in N_i$ наборів спостережень, що складаються із послідовності векторів ознак:

$$\forall o_i : D_i = D_i^1 \cup D_i^2 \cup \dots \cup D_i^{N_i}, D_i^k = (\bar{z}_{1,k_i}, \bar{z}_{2,k_i}, \dots, \bar{z}_{T_{k_i},k_i}) = \begin{pmatrix} z_{1,k_i}^{(1)} & z_{2,k_i}^{(1)} & \dots & z_{T_{k_i},k_i}^{(1)} \\ z_{1,k_i}^{(2)} & z_{2,k_i}^{(2)} & \dots & z_{T_{k_i},k_i}^{(2)} \\ \dots & \dots & \dots & \dots \\ z_{1,k_i}^{(n)} & z_{2,k_i}^{(n)} & \dots & z_{T_{k_i},k_i}^{(n)} \end{pmatrix} \quad (\text{A.3.4})$$

Сформуємо множину D для навчання мережі B_l наступним чином. До кожного вектора спостережень для кожного об'єкта додамо нульовий стан, що буде індикатором конкретного об'єкта і об'єднаємо всі вектори всіх об'єктів в одну множину:

$$D = \bigcup_{i=1}^m \bigcup_{k=1}^{N_i} \begin{pmatrix} o_i & o_i & \dots & o_i \\ \mathbf{r} & \mathbf{r} & \dots & \mathbf{r} \\ z_{1,k_i} & z_{2,k_i} & \dots & z_{T_{k_i},k_i} \end{pmatrix} = \begin{pmatrix} o_s & \dots & o_s & o_l & \dots & o_l \\ \dots & z_{1,k_s}^{(l)} & \dots & z_{T_{k_s},k_s}^{(l)} & \dots & z_{1,k_l}^{(l)} & \dots & z_{T_{k_l},k_l}^{(l)} & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ z_{1,k_s}^{(n)} & \dots & z_{T_{k_s},k_s}^{(n)} & z_{1,k_l}^{(n)} & \dots & z_{T_{k_l},k_l}^{(n)} \end{pmatrix}$$

Навчання B_l здійснюється із використанням міри середня відносна похибка (MRE) і отримана мережа здатна класифікувати об'єкти o_1, \dots, o_m без врахування часової динаміки.

Для навчання B_{\rightarrow} можна скористатися кількома підходами [13]. По-перше, її структуру можна задати вручну, як в прихованих марківських моделях (ПММ) [263, 314], стверджуючи, що теперішнє значення кожної змінної $X_t^{(i)}$, що спостерігається, залежить від її минулого значення (на 1 крок назад у часі) $X_{t-1}^{(i)}$, тобто $\forall i : \pi(X_t^{(i)}) = X_{t-1}^{(i)}$, і існує відповідний розподіл умовних ймовірностей $P(X_t^{(i)} / X_{t-1}^{(i)})$. Приклад структури мережі наведено на рис. А.3.3.

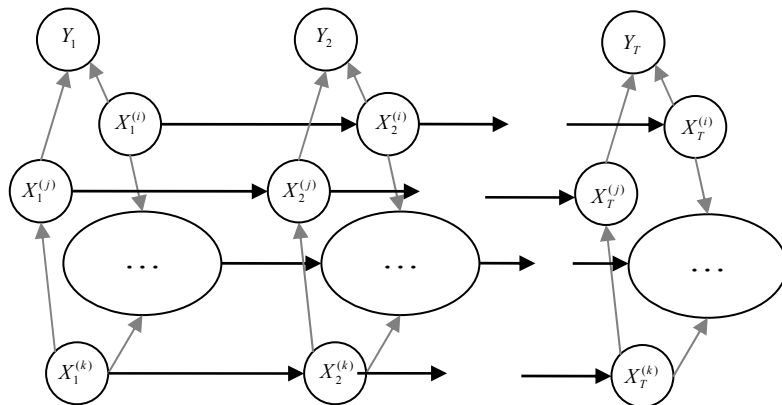


Рис. А.3.3 ДМБ у вигляді ПММ

За логікою, теперішні значення змінних не можуть залежати від майбутніх значень. Тому, будемо вважати, що мережа B_{\rightarrow} описує зв'язки змінної $X_t^{(i)}$ зі змінними попереднього часового про шарку $t - 1$.

Таким чином, для побудови B_{\rightarrow} також можна використати опис мінімальної довжини як міру (MDL). Використовувати міру MRE при цьому немає сенсу, оскільки завдання мережі B_{\rightarrow} описати часові переходи між змінними, тому міра MDL цілком для цього підходить.

Для того, щоб застосувати міру MDL [75, 77, 175], необхідно визначити, яким чином обчислювати $P_{B_{\rightarrow}}(X_t^{(j)} / \pi(X_t^{(j)}))$ і $P_{B_{\rightarrow}}(X_t^{(j)}, \pi(X_t^{(j)}))$. Нехай батьківська множина вершини $X_t^{(j)}$ має вигляд $\pi(X_t^{(j)}) = \{X_{t-1}^{s_1}, \dots, X_{t-1}^{s_j}\}$. Нехай вершина $X_t^{(j)}$ і її батьківська множина приймають певні значення: $X_t^{(j)} = y_0$, $X_{t-1}^{s_m} = y_m$, $\forall X_{t-1}^{s_m} \in \pi(X_t^{(j)})$. Тоді:

$$P_{B_{\rightarrow}}(X_t^{(j)} = y_0 / X_{t-1}^{s_1} = y_1, \dots, X_{t-1}^{s_j} = y_{q_j}) = \frac{\sum_{i=1}^m \sum_{k_i=1}^{N_i} \sum_{l=2}^{T_{k_i}} I(z_{t,k_i}^{(j)} = y_0, z_{t-1,k_i}^{(s_1)} = y_1, \dots, z_{t-1,k_i}^{(s_j)} = y_{q_j})}{\sum_{l=1}^{r_j} \sum_{i=1}^m \sum_{k_i=1}^{N_i} \sum_{l=2}^{T_{k_i}} I(z_{t,k_i}^{(j)} = x_{j,l}, z_{t-1,k_i}^{(s_1)} = y_1, \dots, z_{t-1,k_i}^{(s_j)} = y_{q_j})}, \quad (\text{A.3.5})$$

де $I(z_{t,k_i}^{(j)} = y_0, z_{t-1,k_i}^{(s_1)} = y_1, \dots, z_{t-1,k_i}^{(s_j)} = y_{q_j})$ – функція індикатор.

$$I(z_{t,k_i}^{(j)} = y_0, z_{t-1,k_i}^{(s_1)} = y_1, \dots, z_{t-1,k_i}^{(s_j)} = y_{q_j}) = \begin{cases} 1, & \text{якщо } z_{t,k_i}^{(j)} = y_0, z_{t-1,k_i}^{(s_1)} = y_1, \dots, z_{t-1,k_i}^{(s_j)} = y_{q_j} \\ 0, & \text{інакше} \end{cases}$$

Аналогічно отримуємо вираз для $P_{B_{\rightarrow}}(X_t^{(j)}, \pi(X_t^{(j)}))$:

$$P_{B_{\rightarrow}}(X_t^{(j)} = y_0, X_{t-1}^{s_1} = y_1, \dots, X_{t-1}^{s_j} = y_{q_j}) = \frac{\sum_{i=1}^m \sum_{k_i=1}^{N_i} \sum_{l=2}^{T_{k_i}} I(z_{t,k_i}^{(j)} = y_0, z_{t-1,k_i}^{(s_1)} = y_1, \dots, z_{t-1,k_i}^{(s_j)} = y_{q_j})}{\sum_{i=1}^m \sum_{k_i=1}^{N_i} \sum_{l=2}^{T_{k_i}} (k_i - 1)},$$

де $\sum_{i=1}^m \sum_{k_i=1}^{N_i} \sum_{l=2}^{T_{k_i}} (k_i - 1)$ – кількість міжчасових переходів у навчальній вибірці D ,

тобто аналог кількості навчальних векторів N .

Логічним довершенням побудови B_{\rightarrow} є ствердження того, що для різних об'єктів розподіли ймовірностей $P_{B_{\rightarrow}}(X_t^{(j)} / \pi(X_t^{(j)}))$ і $P_{B_{\rightarrow}}(X_t^{(j)}, \pi(X_t^{(j)}))$ відрізняються. Тобто кожен об'єкт має свої унікальні часові переходи станів. Формально це означає, що в міжчасові розподіли необхідно ввести параметр o_i , що буде маркувати вказані розподіли: $P_{B_{\rightarrow}}^{o_i}(X_t^{(j)} / \pi(X_t^{(j)}))$ і $P_{B_{\rightarrow}}^{o_i}(X_t^{(j)}, \pi(X_t^{(j)}))$. Щоб позбутися параметричних розподілів, параметр o_i можна внести як аргумент цих розподілів:

$$P_{B_{\rightarrow}}^{o_i}(X_t^{(j)} / \pi(X_t^{(j)})) = P_{B_{\rightarrow}}(X_t^{(j)} / \pi(X_t^{(j)}), Y_{t-1} = o_i)$$

$$P_{B_{\rightarrow}}^{o_i}(X_t^{(j)}, \pi(X_t^{(j)})) = P_{B_{\rightarrow}}(X_t^{(j)}, \pi(X_t^{(j)}), Y_{t-1} = o_i)$$

Самі ж набори умовних ймовірностей $P_{B_{\rightarrow}}^{o_i}(X_t^{(j)}/\pi(X_t^{(j)}))$ і $P_{B_{\rightarrow}}^{o_i}(X_t^{(j)},\pi(X_t^{(j)}))$ можна замінити на відповідні $P_{B_{\rightarrow}}(X_t^{(j)}/\pi'(X_t^{(j)}))$ і $P_{B_{\rightarrow}}(X_t^{(j)},\pi'(X_t^{(j)}))$, де $\pi'(X_t^{(j)}) = \pi(X_t^{(j)}) \cup Y_{t-1}$. Тобто до батьківської множини $\pi(X_t^{(j)})$ кожної спостережуваної вершини $X_t^{(j)}$ додається не спостережувана вершина Y_{t-1} за попередній часовий прошарок. Таким чином, отримана топологія мережі B_{\rightarrow} набуває вигляду, як зображено на рис. А.3.4.

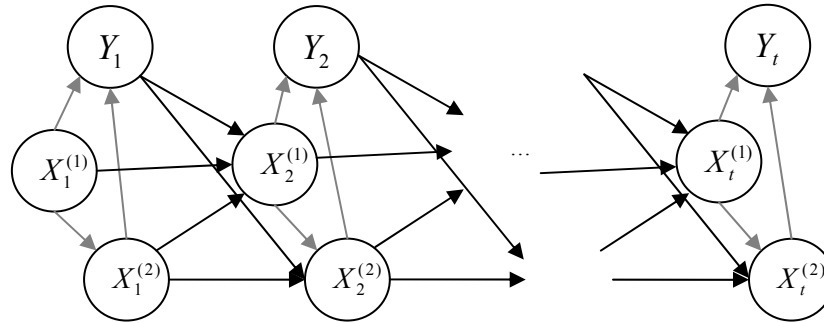


Рис. А.3.4 Приклад ДБМ, де умовні ймовірності ознак залежать від об'єкта

А.4. Реалізація ЕМ-алгоритму

Позначимо як матрицю неповних даних, і повний набір даних $\mathbf{Z}=[\mathbf{X}, \mathbf{Y}]$, яка може бути описано за допомогою функції спільного розподілу ймовірностей (щільності) у такому вигляді:

$$p(\mathbf{Z}|\theta) = p(\mathbf{X}, \mathbf{Y}|\theta) = p(\mathbf{Y}|\mathbf{X}, \theta) p(\mathbf{X}|\theta),$$

де θ – вектор параметрів для конкретного розподілу. Таким чином, ми визначили спільну функцію щільності (розподіл) між відсутніми \mathbf{Y} і доступними вимірами \mathbf{X} з відповідною функцією правдоподібності:

$$L(\theta|\mathbf{Z}) = L(\theta|\mathbf{X}, \mathbf{Y}) = p(\mathbf{X}, \mathbf{Y}|\theta).$$

Ця функція фактично є випадковою змінною через те, що відсутні дані \mathbf{Y} невідомі (випадкові); \mathbf{X}, θ розглядаються як відомі константи (значення) для поточного розрахункового кроку. ЕМ – алгоритм може бути представлений в два етапи, як показано нижче [311].

Крок 1 (очікування). ЕМ – алгоритм повинен обчислити очікуване значення повної функції даних логарифмічної правдоподібності $\text{Log}[p(\mathbf{X}, \mathbf{Y}|\theta)]$ щодо невідомих даних вимірювань \mathbf{Y} для даних \mathbf{X} , які спостерігаються, і поточних оцінок параметрів θ_k . Насправді на ітерації k ми зможемо використовувати оцінки параметрів, взяті з попередньої ітерації θ_{k-1} , тобто оптимізується наступна логарифмічна функція правдоподібності [311]:

$$L_1(\theta|\theta_{k-1}) = E[\text{Log}(p(\mathbf{X}, \mathbf{Y}|\theta)|\mathbf{X}, \theta_{k-1})], \quad (\text{A.4.1})$$

де θ – оптимальний вектор параметрів, які ми шукаємо. Цей вираз є умовним математичним очікуванням за умови $(\mathbf{X}, \theta_{k-1})$. Використовуючи маргінальний

розподіл відсутніх значень $f(\mathbf{Y}|\mathbf{X},\theta_{k-1})$, права частина (A.4.1) може бути записана у вигляді:

$$E\left[\text{Log}(p(\mathbf{X},\mathbf{Y}|\theta)|\mathbf{X},\theta_{k-1})\right]=\int_{y\in\Omega}\text{Log}(p(\mathbf{X},y|\theta)f(y|\mathbf{X},\theta_{k-1}))dy, \quad (\text{A.4.2})$$

де Ω є простором значень, які може приймати змінна y . Дуже часто щільність, яка насправді необхідна, визначається наступним чином:

$$f(y,\mathbf{X}|\theta_{k-1})=f(y|\mathbf{X},\theta_{k-1})f(\mathbf{X}|\theta_{k-1}),$$

З практичної точки зору корисно для обчислень на першому кроці максимізувати наступний детермінований вираз щодо θ :

$$q(\theta)=E_{\mathbf{Y}}[\varphi(\theta,Y)]=\int_y\varphi(\theta,Y)f_Y(y)dy,$$

де $\varphi(\theta, \mathbf{Y})$ – це функція з константою θ і \mathbf{Y} є випадковою величиною, що визначається за розподілом $f_{\mathbf{Y}}(y)$.

Крок 2 (максимізація). Тепер необхідно максимізувати очікування знайдене на першому етапі:

$$\theta_k=\arg\max_{\theta}L_1(\theta,\theta_{k-1}),$$

Ці два описані кроки, обчислюються ітераційно, до тих пір, поки деякий (вибраний) критерій зупинки не буде виконаний. На кожній ітерації виконання алгоритму правдоподібність (A.4.1) збільшується і ЕМ–алгоритм повинен збігатися до локального максимуму цієї функції. Параметри розподілу, обчислені таким чином, мають бути використані для створення додаткових даних зі спільного розподілу.

A.5. Приклад комбінованого використання мереж Байєса та нейронних мереж

У роботі [122] на основі фактичних статистичних даних стосовно позичальників кредитів банку побудовано математичні моделі у формі нейронних мереж зворотного розповсюдження та статичної байєсівської мережі. Виконано обчислювальні експерименти для прогнозування дефолту позичальників кредитів з використанням кожної побудованої моделі окремо, а також комбінованої (інтегрованої) моделі, яка складається з обох згаданих структур. Для перевірки доцільності використання у моделюванні кредитних ризиків запропонованого комбінованого підходу виконано експерименти, за якими на початковому етапі моделювання застосовувалась мережа Байєса, що встановлювала причинно-наслідкові зв'язки між змінними-характеристиками та результуючою змінною і визначала ті змінні, які доцільно використовувати у подальшому для моделювання. При встановленні причинно-наслідкових зв'язків між змінними розглянуто такі типи змінних: ті, що взагалі не впливають на результат (так звана «зайва» інформація у вибірці) [122]; змінні, які не впливають на результат, але зазнають впливу сукупності характеристик, які, у тому числі, впливають на ключову змінну; змінні, які опосередковано

впливають на результат («батьківські» змінні характеристик, що визначають результуючу змінну) та змінні, які безпосередньо впливають на результат.

Розглянемо конкретний приклад. Для включення у нейронну мережу було відібрано змінні, що за побудованою на попередньому етапі МБ взагалі не впливають на результат (дві змінні). Всі інші змінні подавалися на вхід нейронної мережі. В результаті час навчання нейромережі скоротився до 7 хвилин 28 секунд, що становить 3-4% від загального часу. З використанням інтегрованого підходу збільшилась точність моделі, зменшилась кількість помилок, та зріс індекс GINI. Отримані результати обчислювальних експериментів представлені у табл. А.5.1.

Таблиця А.5.1

Загальна точність моделі та помилки I-го та II-го роду для нейронної мережі

Факт повернення чи неповернення кредиту	Прогноз: Повернення кредиту (0)	Прогноз: Дефолт (1)	Процент точності
Cut-off=0,1			
Факт: Повернення кредиту (0)	111	39	0,74
Факт: Дефолт (1)	8	42	0,84
Загальна точність моделі			0,765
Cut-off=0,15			
Факт: Повернення кредиту (0)	111	39	0,74
Факт: Дефолт (1)	9	41	0,82
Загальна точність моделі			0,76
Cut-off=0,2			
Факт: Повернення кредиту (0)	113	37	0,753333
Факт: Дефолт (1)	9	41	0,82
Загальна точність моделі			0,77
Cut-off=0,25			
Факт: Повернення кредиту (0)	114	36	0,76
Факт: Дефолт (1)	9	41	0,82
Загальна точність моделі			0,775
Cut-off=0,3			
Факт: Повернення кредиту (0)	114	36	0,76
Факт: Дефолт (1)	9	41	0,82
Загальна точність моделі			0,775

Побудована за результатами експериментів ROC-крива представлена на рис. А.5.1. Значення площі під кривою становить: $AUC = 0,842$, а $GINI = 2 * AUC - 1 = 0,684$, що є вищим показником порівняно з обома методами – мережами Байєса та нейронними мережами, де значення відповідних характеристик становило: $GINI = 2 * AUC - 1 = 0,646$ для нейронних мереж та $GINI = 2 * AUC - 1 = 0,678$ для мереж Байєса.

Таким чином, кращі результати для використаної вибірки даних забезпечує комбінована модель із загальною точністю 0,775 та індексом GINI 0,684. Це підтверджує доцільність використання додаткових засобів інтелектуального аналізу даних для попередньої обробки вхідних даних. Як показали експериментальні дослідження, завдяки використанню МБ на попередньому етапі вдається на 3-4% зменшити час на обробку вхідних даних і навчання нейромережі, але найголовніше, вдається суттєво покращити результати класифікації і зменшити значення помилок другого роду та підняти загальну точність модель. Крім того, суттєво збільшився індекс GINI і в результаті модель забезпечує високу якість прогнозування ризику втрат.

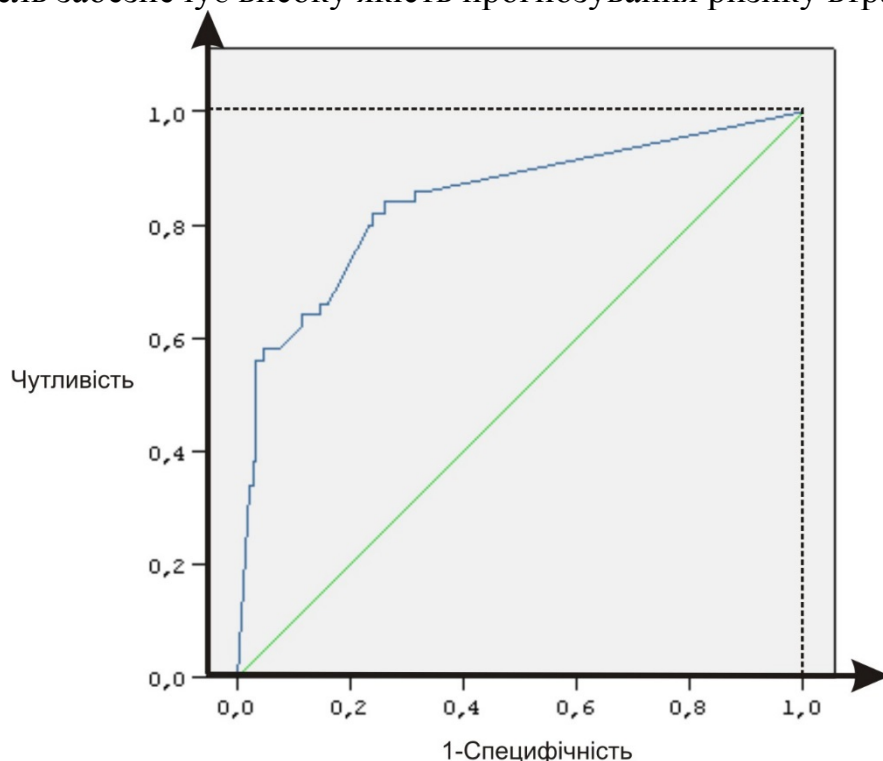


Рис. А.5.1 ROC-крива для інтегрованої моделі на основі мережі Байєса та нейронної мережі

Отже, комбіновані моделі [97, 122] для вирішення задач аналізу кредитних ризиків є доволі ефективним інструментом, що надає переваги у вигляді економії часу та підвищення якості побудованих моделей, тож доцільне подальше використання комбінованого підходу до аналізу ризиків різної природи.

ДОДАТОК Б

ПРИКЛАДИ ЗАСТОСУВАННЯ ЗАПРОПОНОВАНИХ У ДИСЕРТАЦІЙНІЙ РОБОТІ МОДЕЛЕЙ ТА МЕТОДІВ

Б.1. Приклад оцінювання ринкового ризику

Для моделювання зміни тенденції на ринку акцій було використано нормовані щоденні дані по абсолютних та процентних змінах цін закриття акцій відносно цін їх закриття попереднього дня для 100 компаній, які входять в S&P100 (2006 – 2017pp.) [124]. Для визначення набору акцій інвестиційного портфеля будувалась кореляційна матриця. Найбільші значення кореляції наявні для множини акцій: SPY, USB, JPM, MSFT, CAT (табл. Б.1.1), їх можна вибрати в якості інвестиційного портфеля для моделювання.

Таблиця Б.1.1

Кореляційна матриця пакету акцій

	SPY	USB	JPM	MSFT	CAT
SPY	1,000000	0,765861	0,759878	0,614661	0,601578
USB	0,765861	1,000000	0,829854	0,417971	0,497563
JPM	0,759878	0,829854	1,000000	0,422293	0,503763
MSFT	0,614661	0,417971	0,422293	1,000000	0,356486
CAT	0,601578	0,497563	0,503763	0,356486	1,000000

Для прогнозування ціни акції вибірку було розділено на навчальну та тестову (у відношенні 80 на 20). Прогнозування ціни здійснювалось за допомогою лінійної регресії, нейронної мережі та робастної лінійної регресії з критерієм Andrew Wave [26, 124], якість моделі перевірялась за критеріями R^2 та середньоквадратичної похибки (СКП або MSE) (табл. Б.1.2). Для лінійної регресії оцінки параметрів знаходились за звичайним МНК. Робастна лінійна модель додає до звичайного МНК ще деякі вагові коефіцієнти. Це дозволяє подолати викиди, оскільки в цій моделі вони отримують менші ваги. Також побудована нейронна мережа з трьома внутрішніми шарами, розмірами 60, 120, 20. Моделювання за методами Random Tree Regression та Random Forest Regression показало неприйнятні результати.

Результати моделювання свідчать про недостатню адекватність побудованих моделей. Це пов'язано з тим, що у вхідних даних не враховується зміна цін на акції протягом дня, оскільки ставилась задача прогнозування тенденції на наступні періоди. Існує певна лаговість (відставання) зміни ціни на акції, проте за всіма моделями чітко видно, що саме тенденція зміни ціни на кілька кроків вперед була змодельована із прийнятною точністю.

Покращення результатів можна досягти, якщо використати знання про взаємозв'язки між компаніями. Наприклад SPY – це ф'ючерс на індекс, в який входить 500 найбільших за капіталізацією компаній США. Оберемо 15 найбільших компаній з цього списку: "BAC", "LMT", "AMZN", "CAT", "MRK", "CVX", "WFC", "INTC", "AAPL", "MSFT", "GOOGL", "IBM", "SLB", "ACN", "PFE" та спрогнозуємо значення SPY за вибірками для цих 15 компаній. Оскільки ціни акцій суттєво відрізняються для різних компаній, то для моделі використано нормовані величини.

Таблиця Б.1.2

Порівняльний аналіз моделей за критеріями R^2 та СКП

Лінійна регресія (навчальна вибірка)					
	SPY	USB	JPM	MSFT	CAT
R^2	0,53519212	0,50873436	0,50988328	0,51256457	0,49582926
MSE	0,3166873	0,64740614	0,9021358	1,03702051	1,03702051
Нейронна мережа (навчальна вибірка)					
	SPY	USB	JPM	MSFT	CAT
R^2	0,54821077	0,49533378	0,50446581	0,51577144	0,47740322
MSE	0,30781731	0,66506585	0,91210749	1,0301979	1,17925373
Лінійна регресія з критерієм Andrew Wave (навчальна вибірка)					
	SPY	USB	JPM	MSFT	CAT
R^2	0,53016137	0,505368	0,50655844	0,50693544	0,49207475
MSE	0,3201149	0,65184243	0,90825569	1,04899652	1,14614702
Лінійна регресія (перевірочна вибірка)					
	SPY	USB	JPM	MSFT	CAT
R^2	0,39069642	0,46365834	0,48275756	0,46515445	0,47809016
MSE	0,11213232	0,37483548	0,48441264	0,44040342	0,92611484
Нейронна мережа (перевірочна вибірка)					
	SPY	USB	JPM	MSFT	CAT
R^2	0,43416615	0,49742729	0,51133462	0,48063778	0,46718419
MSE	0,10413243	0,35123522	0,45764939	0,42765411	0,94546719
Лінійна регресія з критерієм Andrew Wave (перевірочна вибірка)					
	SPY	USB	JPM	MSFT	CAT
R^2	0,4236834	0,48036966	0,49230044	0,46205395	0,48246044
MSE	0,10606161	0,36315636	0,47547545	0,44295644	0,91835989

Використання лінійної регресії для прогнозування значень на навчальній вибірці показало $R^2 = 0,917$, $MSE = 0,055$, та $R^2 = 0,814$ і $MSE = 0,035$ – на тестовій вибірці. Графіки прогнозованих і реальних значень SPY практично співпадають, а лінійна модель дає можливість обчислити прийнятні за якістю оцінки для прогнозування ціни акції SPY.

Попередній кореляційний аналіз портфеля акцій дав можливість відібрати акції, найбільш корельовані між собою та із ф'ючерсом на індекс S&P500 (SPY) (рис. Б.1.1).



Рис. Б.1.1 Тренд S&P500 на ринку США

За наявності додатної кореляції з показником ринку, з акцій, обраних для утворення портфеля, можна формувати довгу позицію. Для утворення портфеля обрано ті самі найбільш корельовані акції SPY, USB, JPM, MSFT, CAT.

Припущення методології VaR передбачають існування нормального розподілу. Для того щоб перевірити тип розподілу цін на акції побудовано гістограму розподілу для кожної акції з портфеля (рис. Б.1.2), яка показала, що вони дійсно розподілені за нормальним розподілом з близьким до нуля математичним сподіванням. Тому можна використати таблицю квантилів для нормального розподілу.



Рис. Б.1.2 Процентна зміна ціни закриття за акціями SPY

Для прогнозування часу настання інвестиційного ризику необхідно визначити волатильність [5, 247] як оцінку середньоквадратичного відхилення для кожного тікера акції:

$$\sigma = \sqrt{\frac{n}{n-1} s^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2},$$

де \bar{x} – вибіркове середнє.

Далі можна розрахувати час настання ризику (а точніше, кількість днів, на які очікувані втрати будуть покриватися інвестиційним капіталом) за формулою:

$$N = \left(\frac{VaR}{\alpha * \sigma * ВП} \right)^2,$$

де VaR – величина втрат (розмір резервного капіталу) на кожну компанію; σ – обчислена волатильність; ВП – сума інвестицій (задана користувачем).

Будемо вважати, що сума інвестицій всередині портфеля розподілена рівномірно. Якщо ні, то можна ввести певні ваги, задані користувачем, і здійснити оцінку втрат за кожною акцією та портфелем в цілому.

Оцінювання за допомогою інформаційної СППР дало такі результати [124]: якщо було вкладено 100 доларів у портфель в цілому (тобто по 20 доларів

в кожну акцію), то з ймовірністю 99% можна бути впевненим, що впродовж 5 робочих днів очікувані втрати не перевищать 7,8\$, а неочікувані 56,4\$.

Розроблена у дисертації ІСППР дозволяє вирішувати задачу визначення часу до настання інвестиційного ризику та прогнозувати обсяг втрат за обраним пакетом акцій та вказаною ймовірністю настання ризику.

Б.2. Задача аналізу електронних закупівель на платформі ProZorro

Електронні закупівлі на платформі ProZorro піддаються ризикам шахрайства. Потрібно якомога швидше виявляти ледь помітні випадки змови або шахрайства серед численних закупівель, договорів, що укладаються на електронній платформі. Для перевірки прозорості тендерних закупівель [72] були використані засоби інтелектуального аналізу даних, реалізовані у вигляді інформаційних технологій на платформі SAS із метою виявлення можливих порушень і зловживань [101].

Висувалось припущення, що існує певна взаємозалежність між тривалістю участі компанії в торгах та її характеристиками, а саме: чи залежить тривалість участі компанії в торгах від того, чи була вона запідозрена в неправомірній діяльності на платформі (у змовах з іншими компаніями). Емпіричним шляхом сформована така вибірка даних [101, 114]:

1. Wins – кількість вигравів певної компанії;
2. Losses – кількість програшів компанії в торгах;
3. Sum_of_deals – загальна сума виграних торгів;
4. Participations – загальна кількість участі в торгах;
5. Objections – кількість скарг, які подавала ця компанія;
6. Date_start – дата початку участі в системі торгів;
7. Date_finish – дата останньої участі в торгах;
8. IdTenderer – унікальний номер учасника тендеру;
9. Suspected – змінна, що показує, чи була фірма запідозрена в неправомірних змовах з іншими учасниками.

10. Churn out – цільова змінна, що дорівнює 1, якщо компанія переставала брати участь у торгах через короткий термін (вважається, що раптове припинення участі у торгах може бути ознакою фіктивної компанії для одного торгу).

Компанію вважають такою, що продовжує вести торги, якщо проміжок часу між початком торгів на платформі і часом останнього торгу складає більше 60 днів (середня статистична тривалість бізнес-циклів компаній на платформі – згідно з офіційними даними ProZorro [329]). Було відфільтровано компанії, що брали участь у конкурсі на платформі менше, ніж тричі.

Вхідна вибірка містила 3966 випадків, із них 1757 компаній [101], що перестали брати участь через досить короткий термін, а 2209 компаній продовжували брати участь у тендерах. Вибірку було розбито на навчальну та перевірючу методом стратифікації щодо цільової змінної у відношенні 70/30.

Оскільки постановка завдання передбачає розв'язання задачі класифікації, то було запропоновано використати такі методи інтелектуального аналізу

даних, як нейронні мережі, дерева рішень, логістичну регресію та Байєсівський класифікатор [101]. У якості критеріїв обрання кращої моделі можуть бути використані [26]: середньоквадратична похибка (MSE – mean squared error), сума квадратів похибок (SSE), інформаційний критерій Акаїке, критерій Байєса – Шварца, частка неправильної класифікації (Misclassification Rate).

Побудова описаних моделей та розрахунок статистичних критеріїв здійснювали на основі інформаційної технології SAS Enterprise Miner [7], обрання кращої моделі може бути здійснено автоматично або на основі заданого критерію якості. Послідовність аналізу тендерних закупівель представлена на рис. Б.2.1.

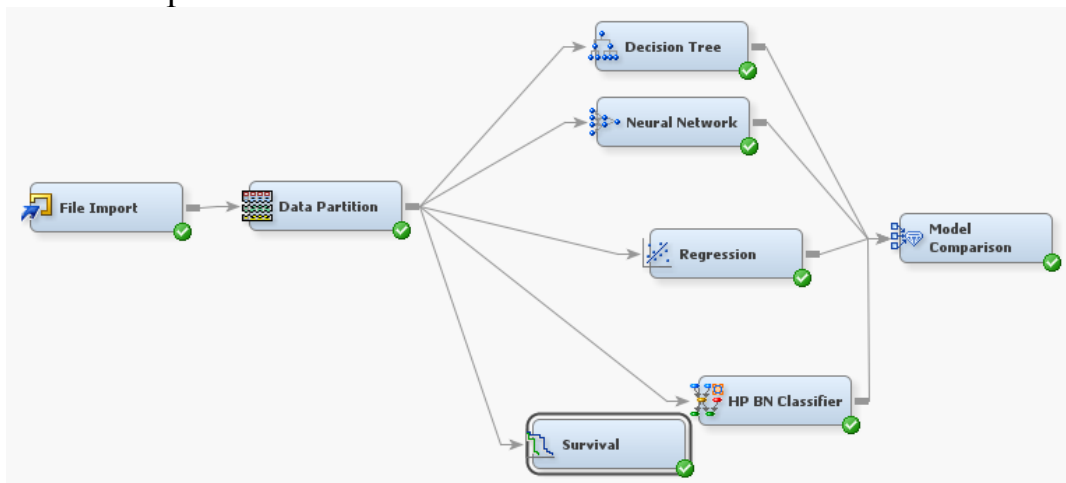


Рис. Б.2.1 Послідовність аналізу тендерних закупівель на основі інформаційної технології SAS Enterprise Miner

Нейронна мережа. Було побудовано різні види нейронних мереж [71, 370, 371], з різною кількістю шарів, активаційними функціями тощо. Критеріями якості було обрано інформаційний критерій Акаїке. Найкращою моделлю для вхідних даних виявилась проста перцептронна нейронна мережа із 20 прихованими шарами та стандартизацією входів на основі відхилення, радіальною комбінаційною функцією, логістичною активаційною функцією та критерієм навчання помилки класифікації. Статистичні критерії якості для найкращої нейронної мережі наведено в таблиці Б.2.1.

Таблиця Б.2.1

Статистичні характеристики найкращої нейронної мережі

Target	Fit Statistics	Statistics Label	Train	Validation
churn_out	DFT	Total Degrees of Freedom	2774	-
churn_out	DFE	Degrees of Freedom for Error	2743	-
churn_out	DFM	Model Degrees of Freedom	31	-
churn_out	NW	Number of Estimated Weights	31	-
churn_out	AIC	Akaike's Information Criterion	3227.794	-
churn_out	SBC	Schwarz's Bayesian Criterion	3411.564	-
churn_out	ASE	Average Squared Error	0.197105	0.193835
churn_out	MAX	Maximum Absolute Error	0.986715	0.984846
churn_out	DIV	Divisor for ASE	5548	2384

churn_out	NOBS	Sum of Frequencies	2774	1192
churn_out	RASE	Root Average Squared Error	0.443965	0.440267
churn_out	SSE	Sum of Squared Errors	1093.537	462.1033
churn_out	SUMW	Sum of Case Weights Times Freq	5548	2384
churn_out	FPE	Final Prediction Error	0.20156	-
churn_out	MSE	Mean Squared Error	0.199332	0.193835
churn_out	RFPE	Root Final Prediction Error	0.448954	-
churn_out	RMSE	Root Mean Squared Error	0.446467	0.440267
churn_out	AVERR	Average Error Function	0.570619	0.567532
churn_out	ERR	Error Function	3165.794	1352.996
churn_out	MISC	Misclassification Rate	0.317231	0.305369
churn_out	WRONG	Number of Wrong Classifications	880	364

Регресійна модель. Для прогнозування, чи перестане компанія брати участь у наступних тендерах (бінарний вихід: 0 – «ні» або 1 – «так») було обрано логістичну регресію stepwise з попарним введенням та виведенням характеристики з моделі. Були отримані такі характеристики: $AIC = 3301,941$ та $MisclassificationRate = 0,328767$.

Дерева рішень. Здійснювалось моделювання на основі методу дерев рішень [232 – 234] із різними налаштуваннями правил відсікання, кількості нащадків та алгоритмів формування піддерев. Найкращим виявилось дерево з мінімальним рівнем неправильної класифікації: на навчальній вибірці $MisclassificationRate = 0,327325$ та перевіірочній вибірці: $MisclassificationRate = 0,313758$. Структура дерева наведена на рис. Б.2.2.

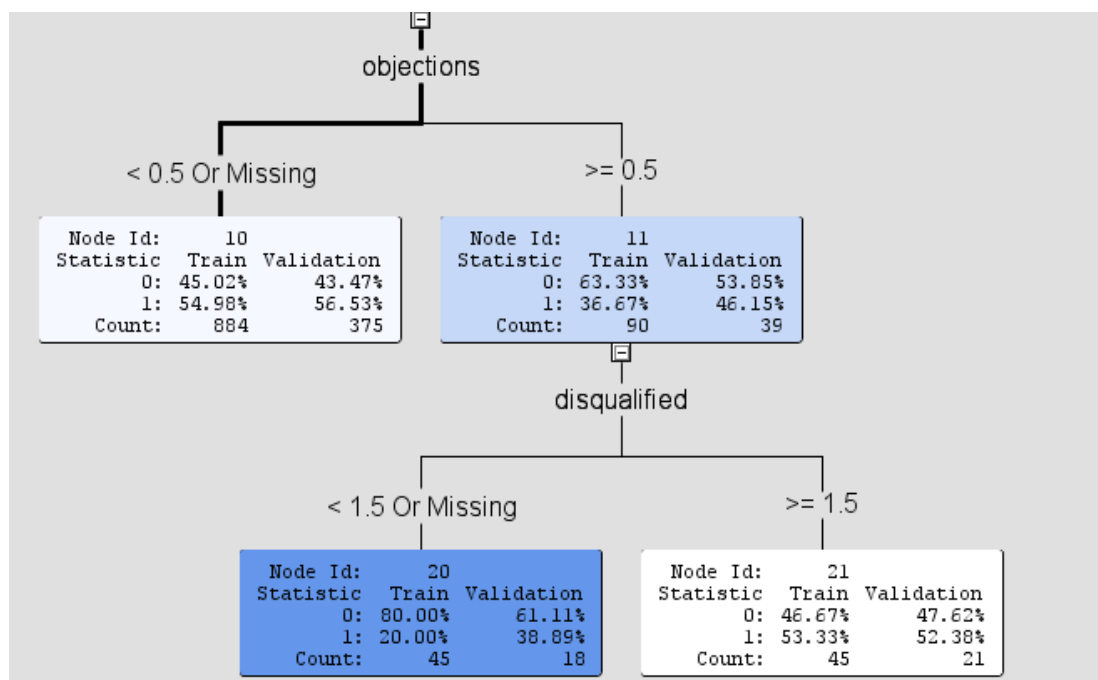


Рис. Б.2.2 Структура побудованого дерева рішень

Наївний Байєсівський класифікатор. За прямого використання наївного

Байєсівського класифікатора [257, 281, 290] відсоток неправильної класифікації становив близько 50%. При збільшенні кількості розбиттів отримано кращі результати (табл. Б.2.2), проте таку модель не рекомендовано для застосування.

Таблиця Б.2.2

Статистичні критерії для наївного Байєсівського класифікатора

Fit Statistics	Statistics Label	Train	Validation
ASE	Average Squared Error	0.240782	2.41E-01
DIV	Divisor for ASE	5548	2384
MAX	Maximum Absolute Error	0.849634	8.18E-01
NOBS	Sum of Frequencies	2774	1192
RASE	Root Average Squared Error	0.490695	4.91E-01
SSE	Sum of Squared Errors	1335.859	575.7147
DISF	Frequency of Classified Cases	2774	1192
MISC	Misclassification Rate	0.460707	0.463926
WRONG	Number of Wrong Classifications	1278	553

Порівняльний аналіз результатів та вибір кращої моделі

Розв'язувалась задача класифікації, вибір кращої моделі здійснювався на основі критерію кількості неправильно класифікованих прикладів на перевіірочній вибірці. Порівняння на перевіірочній вибірці є обґрунтованішим, оскільки певні методи схильні прилаштовуватись до навчальної вибірки. Результати моделювання наведено у таблиці Б.2.3.

Таблиця Б.2.3

Результати класифікації різними методами

Модель	Misclassification Rate
Нейронна мережа	0.305369
Логістична регресія	0.307047
Дерево рішень	0.313758
Наївний Байєсівський класифікатор	0.463926

Отже, найкращою моделлю для аналізу даних ProZotto виявилась нейронна мережа, яка з точністю 70 % дозволяє спрогнозувати, чи буде продовжувати компанія брати участь в публічних торгах. Було підтверджено відсутність дієвих механізмів вилучення недоброчесних учасників із системи торгів, а також знайдено дві потенційні групи компаній-привидів – одноденні та ті, що постійно діють.

Далі було виконано динамічне оцінювання на основі розробленого динамічного методу для прогнозування поведінки реальних учасників закупівель і виявлення шахрайських та нетипових учасників торгів.

Моделювання учасників тендерів за допомогою моделі Каплан-Майєра

Для моделювання задачі тривалості перебування компанії на платформі в умовах конкуренції використаємо деякі припущення з теорії виживання. Непараметрична модель Каплан-Майєра дозволяє оцінити, як змінюється функція виживання (подальша участі компанії в тендерах) з часом. Функція виживання залишається на високому рівні і більше половини компаній залишаються в системі торгів на період більше 60 днів (рис. Б.2.3).

Найбільша загроза виходу компаній спостерігається у перші 20 днів і після 40 дня (в середньому відповідає 1-2% торгів в цей період) (рис. Б.2.4).

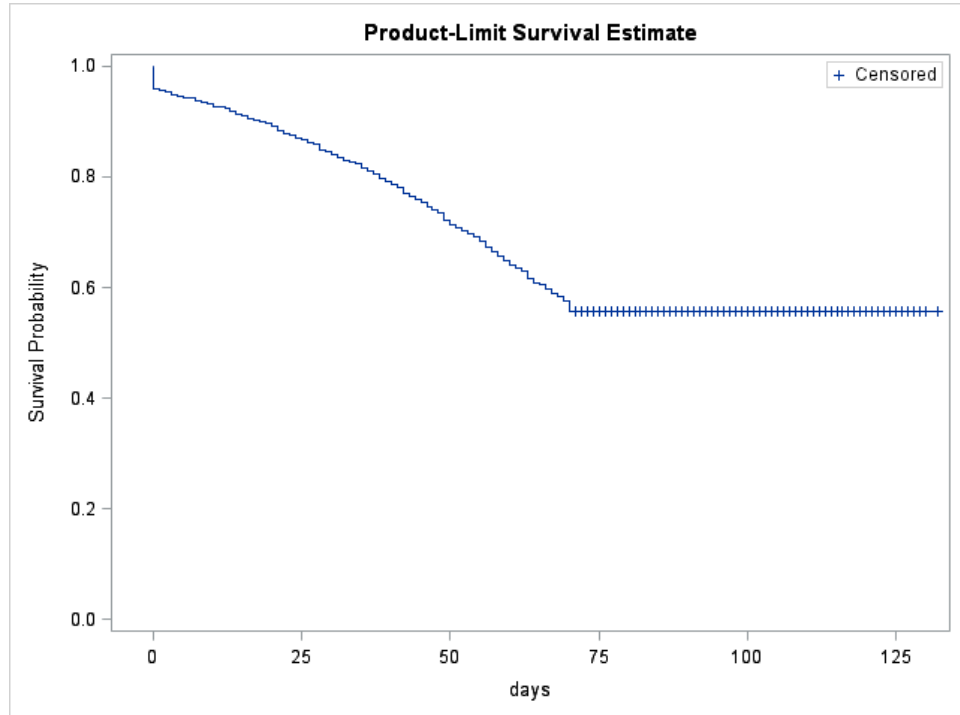


Рис. Б.2.3 Графік ймовірності виживання в цілому компаній на платформі ProZorro

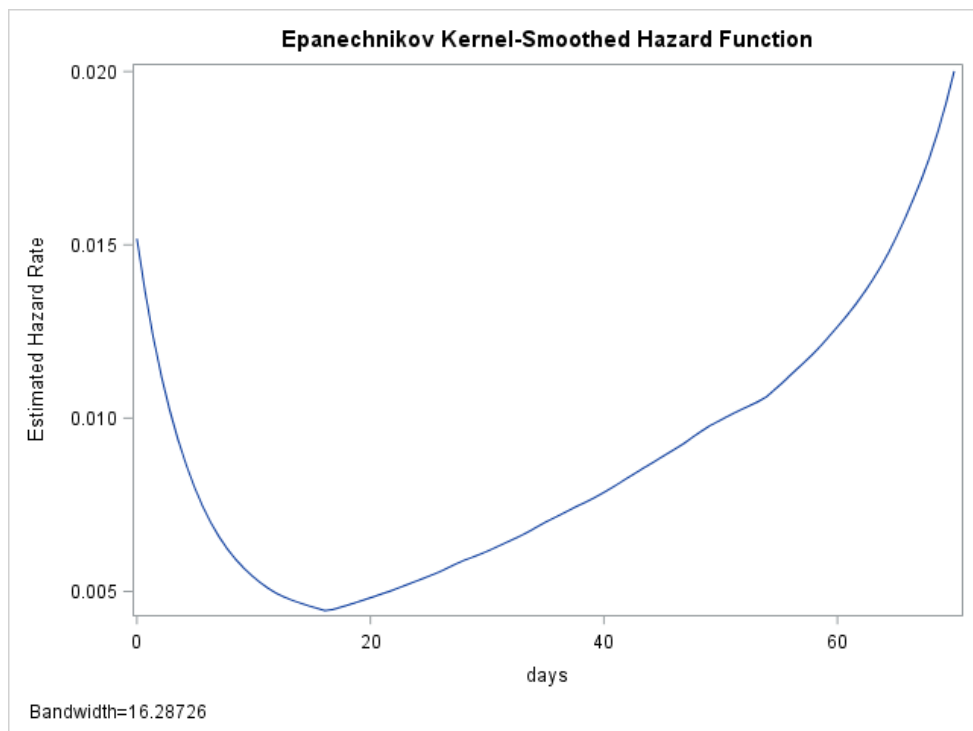


Рис. Б.2.4 Графік функції ризику

Для відповіді на ключове питання дослідження також були побудовані функції виживання для двох груп – підозрюваних в змові фірм ($\text{suspected} = 1$) і тих, що не були запідозрені. Кількісний розподіл фірм за стратам наведено у таблиці Б.2.4, а у таблиці Б.2.5 показані значення статистичних критеріїв для порівняння: WilCoxon і Logrank.

Таблиця Б.2.4

Кількісне співвідношення цензурованих і нецензурованих даних

Summary of the Number of Censored and Uncensored Values					
Страта	Підозрілі	Всього	Вибули	Цензуровані	Процент цензурованих
1	0	3872	1725	2147	55.45
2	1	94	32	62	65.96
Всього		3966	1757	2209	55.70

Таблиця Б.2.5

Порівняння за статистичними критеріями по стратах функцій виживання

Test of Equality over Strata			
Test	Chi-Square	DF	Pr >
Log-Rank	3,4498	1	0,0633
Wilcoxon	3,0405	1	0,0812
-2Log(LR)	5,7826	1	0,0162

Згідно побудованої моделі, компанії, що були запідозрені у змовах, в середньому мають більш високу ймовірність «виживання на платформі», тобто вони в меншій кількості залишають майданчик для торгів (рис. Б.2.5). Цей сумний факт підтверджує неієздатність поточної системи виявлення змов на майданчику.

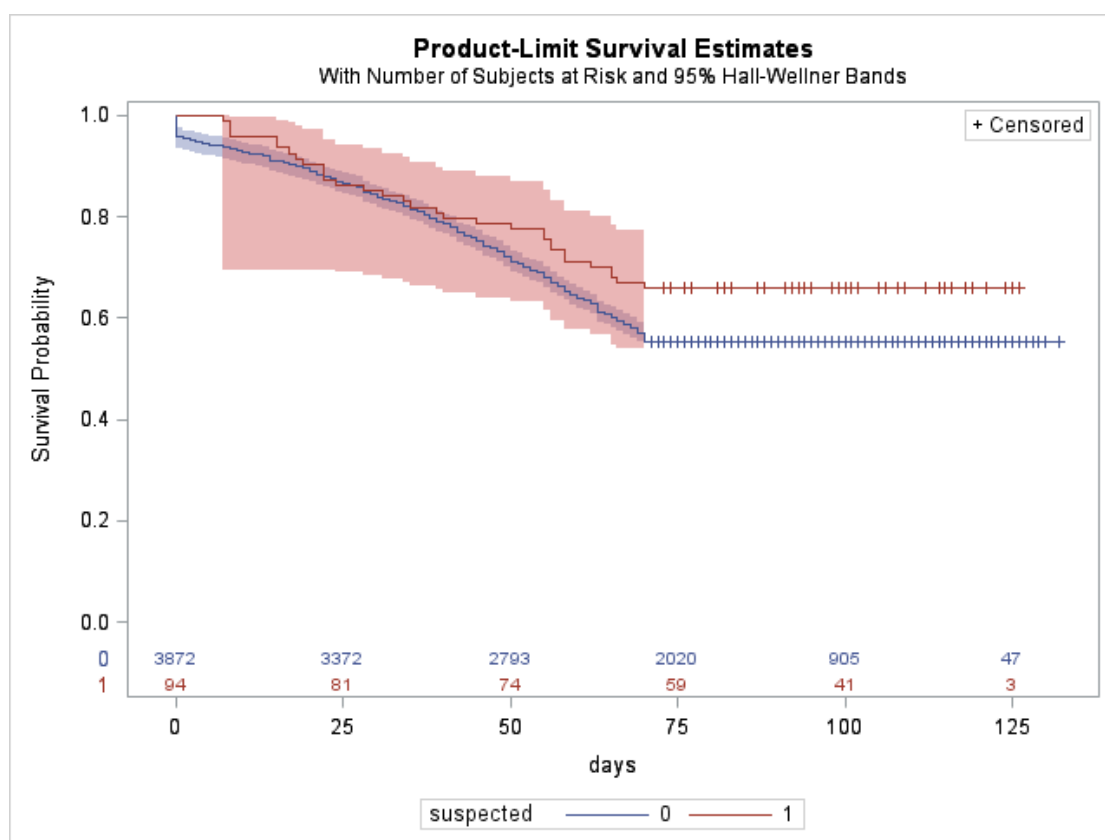


Рис. Б.2.5 Графік функції виживання за стратам (0 – нормальні компанії, 1 – запідозрені в шахрайстві (suspected), Survival Probability – функція виживання)

На графіках функцій ризику, побудованих для страт підозрілих і нормально класифікованих компаній (рис. Б.2.6), ймовірність виходу підозрюваної компанії з торгів зменшується після 20 денного періоду, і вона у 3-4 рази нижче, ніж для звичайної компанії.

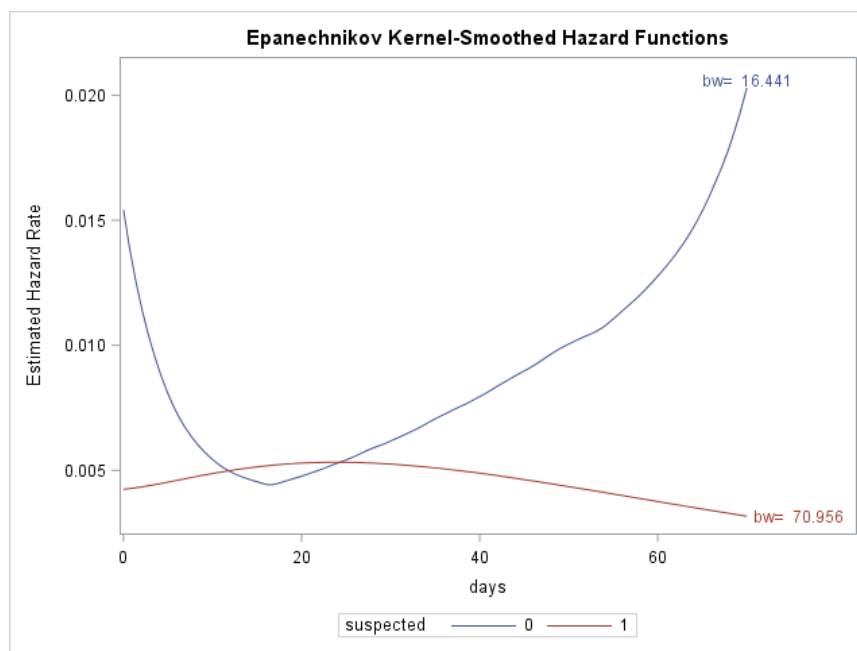


Рис. Б.2.6 Графіки функцій ризику по стратам

Далі проводився аналіз компаній, що беруть участь в торгах по маркеру дискваліфікації. Так, у вибірці присутні компанії, які не були дискваліфіковані жодного разу (значення *suspected* <1), були дискваліфіковані 2 рази через неправильно оформлені або невідповідні документи, були дискваліфіковані більше 3-х разів через те, що платформа запідозрила їх в шахрайських діях.

На графіках (рис. Б.2.7 і рис. Б.2.8) можна бачити різницю між поведінкою груп, що не були ніколи дискваліфіковані, і дискваліфікованих 2, 3 або більше разів.

За допомогою критерію Wilcoxon здійснено перевірку відмінностей в групах компаній (стратах), які були дискваліфіковані (табл. Б.2.6).

Таблиця Б.2.6

Порівняння статистик за тестом Wilcoxon

Adjustment for Multiple Comparisons for the Wilcoxon Test				
Strata Comparison			p-Values	
disqualified	disqualified	Chi-Square	Raw	Scheffe
1,0000	2,0	63,7528	<,0001	<,0001
1,0000	3,0000	109,7	<,0001	<,0001
2,0	3,0000	0,6076	0,4357	0,7380

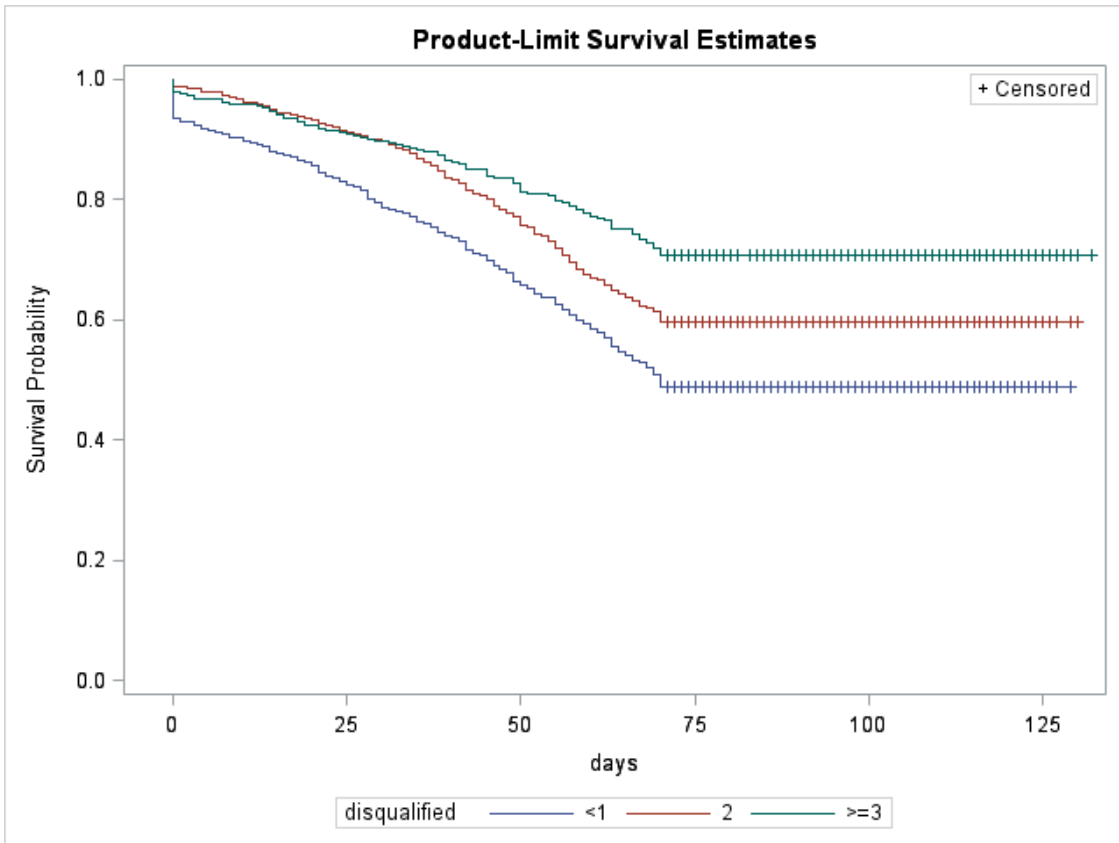


Рис. Б.2.7 Графік функції виживання компаній в залежності від мітки дискваліфікації (disqualified – кількість дискваліфікацій)

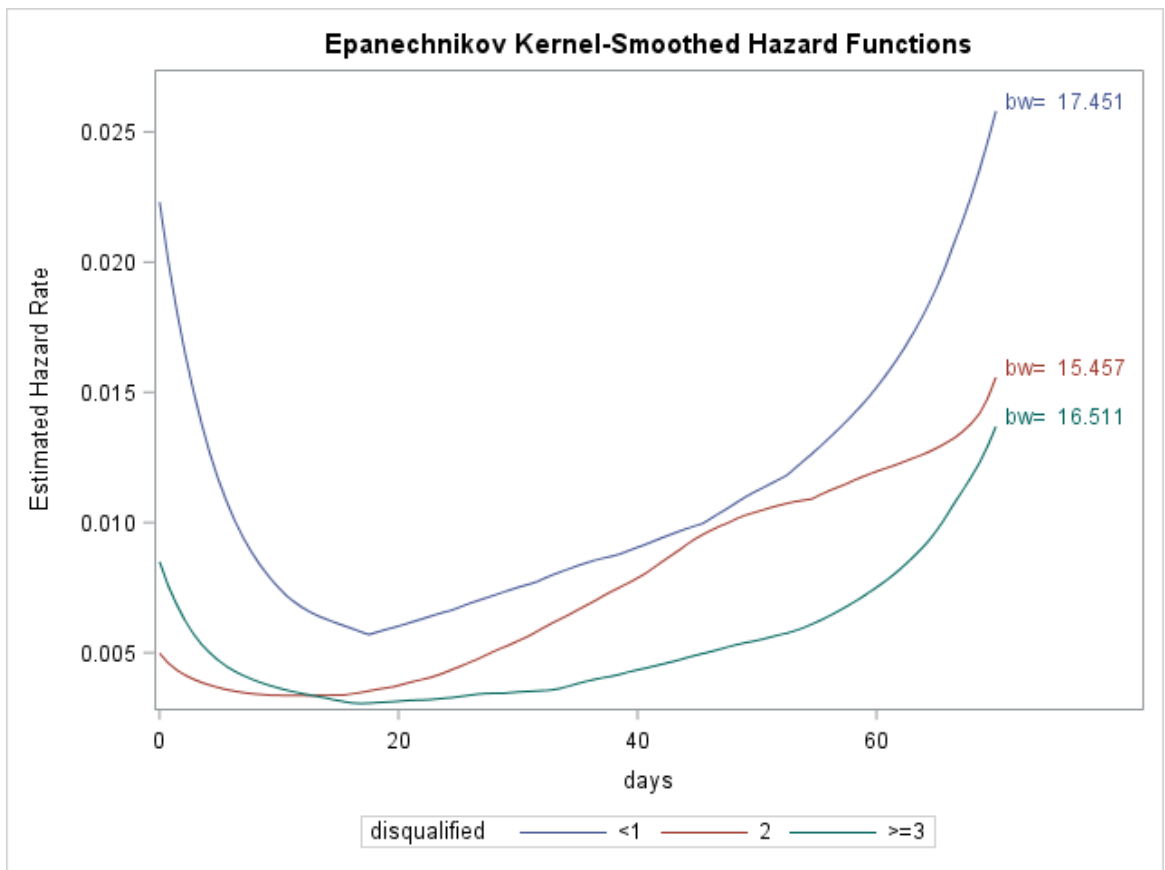


Рис. Б.2.8 Графік функції рівня ризику в залежності від класифікації за ознакою дискваліфікації з торгів

Таблиця Б.2.7

Аналіз оцінок максимальної правдоподібності

Параметр	DF	Оцінка параметру	Станд. похибка	Chi-Square	Pr > ChiSq	Рівень ризику
losses	1	0,05430	0,01203	20,3790	<,0001	1,056
objections	1	-0,13225	0,03631	13,2655	0,0003	0,876
participations	1	-0,15657	0,01051	222,0473	<,0001	0,855
wins	0	0

Також були побудовані моделі пропорційних ризиків Кокса в SAS Enterprise Guide [197] і порівняні функції ризику і функції виживання з плином часу (рис. Б.2.9). Графіки показують, що функція виживання для компаній істотно падає до 70 дня, це як раз і пов'язане зі зникненням з торгів компаній-одноденок. Хоча слід зазначити також і відтік компаній, які вперше взяли участь у торгах, зіткнулися з фактом непрозорості проведення тендерів або просто змовою, і прийняли для себе рішення більше не брати участь у торгах.

Графік функції ризику, який досить довго коливається протягом перших 60 днів, показує залежність і ризикованість для компаній, які встигли взяти участь тільки в одних торгах, а далі перестали брати участь у торгах. Тут також чітко видно пік для компаній, які були маріонетками і йшли після одноразового фіктивного участі в торгах. Цікавим фактом є зміна функцій виживання в часі для компаній, в залежності від факту дискваліфікації з торгів.

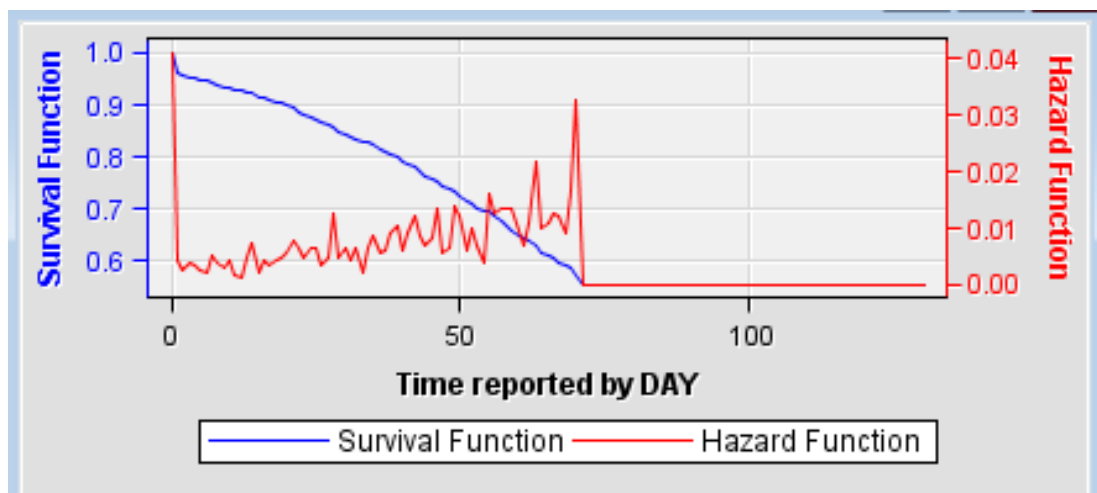


Рис. Б.2. 9 Зміна графіка функції ризику і функції виживання в залежності від кількості днів перебування в системі торгів (Survival Function – функція виживання, Hazard Function – функція ризику)

Отже, аналіз даних, проведений методами нейронних мереж, логістичної регресії і дерев рішень, показав результати помилкової класифікації тендерних заявок на предмет шахрайства на рівні 30%, що є досить високим значенням, і говорить про недосконалість застосовуваних на платформі алгоритмів, і в цілому нездатності системи автоматично виявляти і блокувати змови на етапі подачі заявок на тендери. Побудовані на основі динамічного методу функції ризику і функції виживання компаній в часі, дозволяють визначити і

класифікувати компанію на предмет шахрайських дій до її безпосередньої участі в наступних торгах. Це можна побачити по поведінці компанії, зміні функції ризику, а також поведінки функції ризику в залежності від характеристики зняття з торгів.

Застосування динамічного оцінювання поведінки компаній, що беруть участь в тендерах, є більш універсальним підходом для виявлення шахрайства, оскільки повторення шахраями дій реального поведінки чесних компаній є неможливим і недоцільним з точки зору ідеї шахрайства.

Б.3. Задача аналізу ризику втрати клієнтів в існуючій ERP-системі

Існуюча іноземна компанія Shop24Direct, яка спеціалізується на продажу товарів медійного характеру – CD, DVD, а також товарах елітного сегменту – люкс-парфумів та прикрас, розробила власну ERP-систему для збору статистичної інформації щодо клієнтів-користувачів інтернет-магазину, бази даних каталогу-товарів та бази знань у вигляді рекомендацій щодо супутніх товарів та нових придбань на основі попередніх придбаних товарів. Компанія має у своїй базі більше ніж 2 млн. унікальних клієнтів та більше 5 млн. замовлень за історію існування. З метою підвищення попиту серед власних користувачів та залучення нових клієнтів компанія проводить маркетингові компанії шляхом e-mail розсилок. Формування рекомендацій щодо супутніх товарів здійснювалась аналітиком, проте релевантність таких рекомендацій є досить низькою, Аналітик використовує власну ERP-систему компанії як автоматизоване робоче місце. Для підвищення ефективності роботи компанії доцільно провести аналіз поведінки клієнтів та автоматично розробляти рекомендації за допомогою запропонованих у дисертації методів і моделей.

База даних компанії включає користувацьку інформацію (стать клієнта та дата народження (ці характеристики мають пропуски; місто проживання, оскільки товари доставляються поштою)); інформацію щодо товару (назва, тип, характеристики товару); кількість придбаного товару і дату покупки.

Нехай існує матриця R розміру $u \times o$ з суб'єктами (клієнтами), об'єктами (товарами) та деякими даними щодо зворотнього зв'язку (попередні замовлення). Необхідно знайти спосіб перетворення її в одну матрицю з суб'єктами та їх профілями (прихованими вподобаннями) $P = (p_{iu})_{|T| \times |U|}$ і одну матрицю з об'єктами та їх профілями (прихованими вподобаннями, що вони задовольняють) $Q = (q_{io})_{|T| \times |O|}$. P та Q матриці містять ваги, які визначають як кожний суб'єкт/об'єкт відноситься до кожного з вподобань t . Задача розрахувати P , Q таким чином, щоб їх добуток наближався до R якомога точніше: $R \approx P \times Q$. На рис. Б.3.1 показана ідея задачі пошуку і надання релевантних рекомендацій.

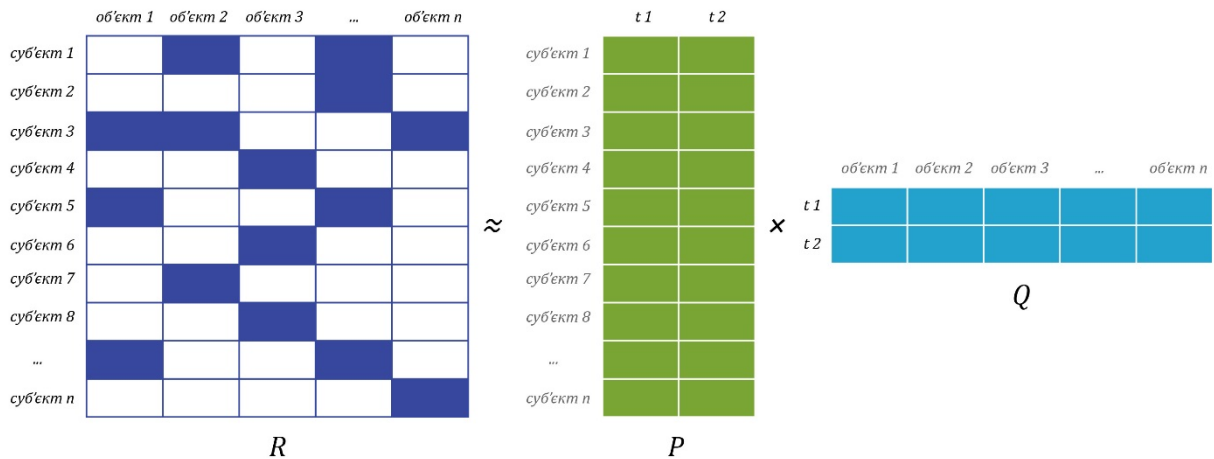


Рис. Б.3.1 Задача пошуку матриці релевантних рекомендацій

В процесі ітеративного присвоювання випадкових значень в матрицях P та Q , використовуючи метод найменших квадратів (МНК), маємо прийти до того значення ваг, які найбільш наближують матрицю R .

В алгоритмі МНК послідовно, на кожній ітерації по чергово змінюються наступні стани системи:

- фіксується P , оптимізується Q ;
- фіксується Q , оптимізується P ,

і продовжується ця операція до наближення до $R \approx P \times Q$.

Для оцінювання точності вирішення задачі аналізу та прогнозування уподобань клієнтів не можуть використовуватись такі стандартні критерії для оцінювання якості прогнозу, як середньоквадратична похибка (RMSE) або середнє абсолютне відхилення (MAE), оскільки важко оцінити, де помилка у прогнозі моделі уподобань клієнтів, а де – рішення клієнту не купувати цей товар тут і зараз. Доцільно було б організувати збір статистичної інформації щодо самого факту перегляду рекомендаційного товару, проте це теж непрямая характеристика, оскільки у клієнта може бути недостатньо часу, проте товар для нього цікавий, тобто прогноз і підбір є вдалим для цього клієнту. Тому для оцінки якості рекомендацій було використано критерії [44]:

- точність ($Precision@k$)

$$Precision@k = \frac{\xi}{k} \quad (Б.3.1)$$

де ξ – кількість рекомендованих об'єктів, з якими у суб'єкта відбулася взаємодія (кількість вірно передбачених уподобань); k – кількість рекомендацій. Критерій показує, яка доля рекомендацій відповідає уподобанням суб'єкту.

- повнота ($Recall@k$)

$$Recall@k = \frac{\xi}{N} \quad (Б.3.2)$$

де N – загальна кількість взаємодій, що була виконана суб'єктами. $Recall@k$ оцінює, яка частка взаємодій, що виконана суб'єктами, припадає на прогнозовані взаємодії, тобто скільки з прогнозованих підібраних товарів були цікавими для клієнтів.

Можна здійснити оцінювання цих показників у грошовому еквіваленті, встановивши вартість кожної взаємодії та штрафи за відсутність взаємодії.

Задачу пошуку супутніх товарів можна представити як пошук таких товарів o , що є схожими за уподобаннями, і які задовольняють клієнтів u :

$$\text{sim_score} = QQ_i^T. \quad (\text{Б.3.3})$$

Збираються такі характеристики: унікальний ідентифікатор клієнта в системі (Kunden_Id), категоріальна ознака статі клієнта (Geschlecht_Id – містить пропуски), назва (Ort) та індекс (Plz) міста клієнта (також неповні дані), дата народження клієнта (Geburtsdatum – містить пропуски), унікальний ідентифікатор товару в системі (Artikelnummer), ціна (Produkt_Preis) та дата продажу (Rechnungsdatum) продукту і кількість придбаного товару (Anzahl).

Тобто існує 4 характеристики з можливо невірними або пропущеними/втраченими даними. Для того, щоб їх коректно обробити, пропонується виконати глибинний аналіз причин появи пропусків та використати запропонований автором *комбінований метод обробки неповних та втрачених даних*, який був детально викладений у розділі 3.

Застосувавши запропонований метод до категоріальної змінної статі клієнту Geschlecht_Id на 2 кроці заповнюємо її як “Missing”, в даному випадку – 0, та будемо вважати, що це категорія «стать не вказано». Аналогічно для змінних місто та поштовий індекс (Ort, Plz), заповнивши всі пропуски значеннями «unknown», і видалимо всі символи за наявності, окрім чисел. Щодо змінної дати народження Geburtsdatum, то пропуски напряму відновити через прогноз регресійної моделі не вважається доцільним, оскільки ця характеристика є суттєвою, а пропуски можуть бути систематичними. Оскільки найчастіше характеристику «вік» сприймають у вигляді кількості повних років, то є сенс здійснити наступне перетворення. Заповнюємо пропуски значеннями, що відповідатимуть першому дню першого місяця, поточного року. Далі формуємо нову змінну «Age», яка розраховується як різниця між значеннями року у змінній Geburtsdatum та значенням року поточної дати. Ця змінна буде цілою числовою і прийматиме значення більше або дорівнює нулю. Нуль буде відображати окремий випадок пропущених даних.

На рис. Б.3.2 представимо візуалізацію даного набору даних, щоб отримати уявлення про те, як пов’язані між собою ознаки.

Графіки залежностей між характеристиками показують, що:

- база клієнтів компанії складається майже порівну з жінок та чоловіків, з незначною перевагою жінок;
- розподіл замовлень серед міст близький до рівномірного;
- розподіл значень віку клієнтів має середнє значення віком у 61 рік.

Набір ознак для імплементації колаборативної фільтрації має наступний набір ознак: Kunden_Id, Artikelnummer, Anzahl. Важливо зазначити, що тестування якості рекомендацій буде виконано у 2 етапи:

- експерт компанії задає різні ідентифікатори користувачів і суб’єктивно аналізує видачу рекомендацій;

- якщо перший етап проходить успішно, виконується другий етап, а саме – аналіз точності моделі через критерії Precision@k та MeanAveragePrecision@k.

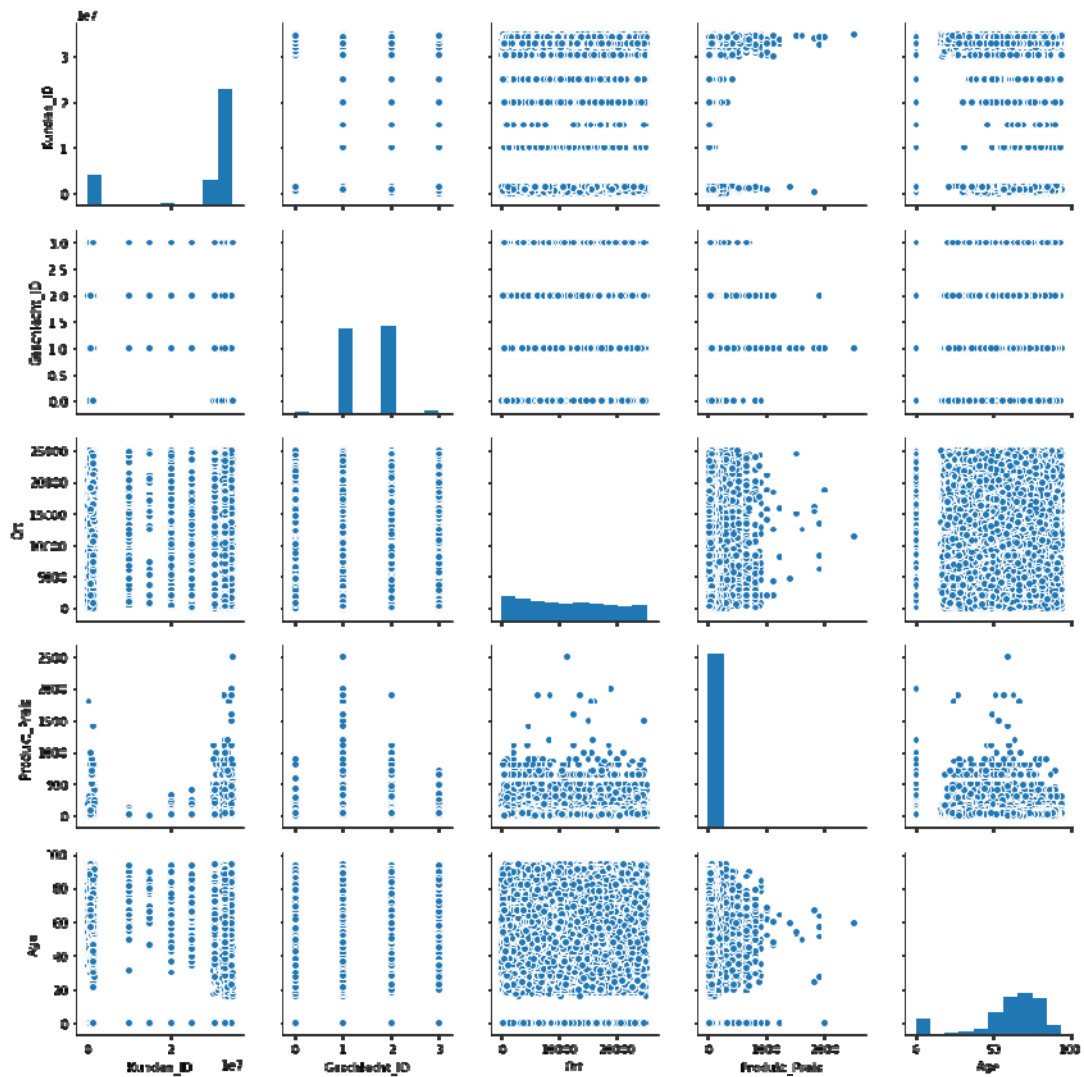


Рис. Б.3.2 Графіки залежності характеристик

Результати роботи моделі колаборативної фільтрації на базі методу Alternating Least Squares (ALS)

Крос-валідація, яку іноді називають перехресною перевіркою, це техніка для перевірки моделі на те, наскільки успішно статистичний аналіз, що застосовується у моделі, здатний працювати на незалежному наборі даних, тобто перевіряє, наскільки прогнозна модель здатна працювати на практиці. Один цикл крос-валідації включає розбиття набору даних на частини, потім побудова моделі на одній частині (тренувальному наборі), і валідація моделі на іншій частині (тестовому наборі). Щоб зменшити розкид результатів, різні цикли крос-валідації можуть проводитися на різних розбиттях, а результати валідації беруться як середній показник якості по всіх циклах.

Модель, яка запропонована в даній роботі, має такі гіперпараметри:

- Кількість ітерацій – i ;
- Кількість прихованих факторів – t ;
- Значення коефіцієнту регуляризації – λ .

В усіх експериментах, функціоналом якості виступатиме Precision@k. Формується «сітка» значень параметрів: i – перебирається від 5 до 50 з кроком в 5; $t = 50$; $\lambda = 0.01$.

Після виконання 10 ітерацій, змінюючи значення кількості ітерацій, будується графік залежності показника Precision@k від значення показника «кількість ітерацій ALS» та графік залежності показника MeanAveragePrecision@k від значення показника «кількість ітерацій ALS».

Отримавши перше наближення до оптимальної кількості ітерацій, його фіксують та знаходять оптимальне значення коефіцієнту регуляризації. Отримують наступну «сітку» значень параметрів: $i = 10$; $t = 50$; λ – перебирається від 0.01 до 1 з кроком 0.01. Після виконання 10 ітерацій, змінюючи значення коефіцієнта регуляризації, будується графік залежності показника Precision@k від значення показника λ .

Змінюючи значення коефіцієнту регуляризації, будують залежність показника MAP@k від значення λ . Формується наступна «сітка» значень параметрів: $i = 10$; t – перебирається від 100 до 1600 з кроком 100; $\lambda = 0.09$. Виконавши 10 ітерацій, змінюючи значення коефіцієнту кількості прихованих факторів, був отриманий графік залежності показника Precision@k від значення показника t (рис. Б.3.3).

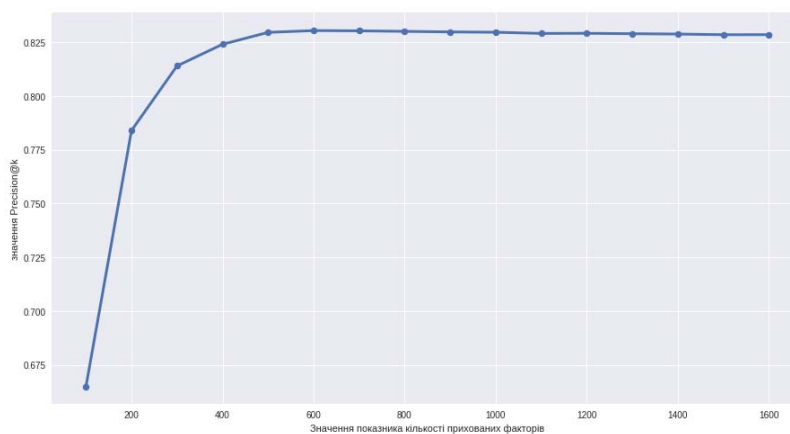


Рис. Б.3.3 Залежність Precision@k від значення кількості латентних факторів

Виконавши 10 ітерацій, змінюючи значення коефіцієнту кількості прихованих факторів, будують графік залежності MAP@k від t (рис. Б.3.4).

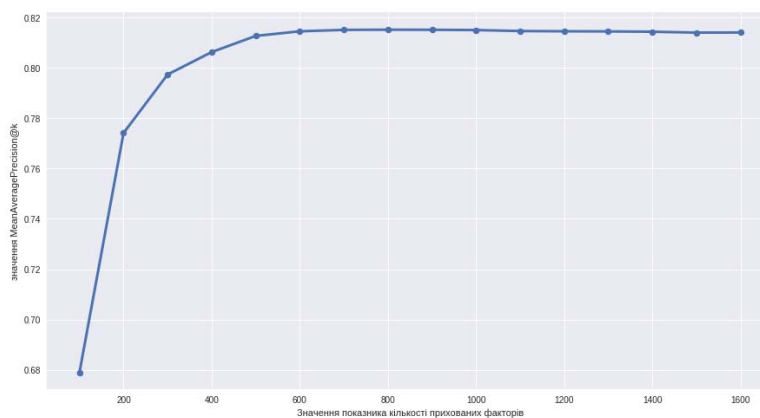


Рис. Б.3.4 Залежність MeanAveragePrecision@k від значення кількості латентних факторів

Аналіз результатів

Функції залежностей Precision@k та MeanAveragePrecision@k мають спадний характер, оскільки при збільшенні кількості ітерацій оптимізації кожної з компонент моделі, відбувається перенавчання моделі. Для Precision@k, показник оптимальної кількості ітерацій становить 15, для MeanAveragePrecision@k оптимальною є кількість ітерацій на значенні 10, як найменшого з цих значень.

Оптимальним значенням коефіцієнту регуляризації є 0.09, значення показників Precision@k та MeanAveragePrecision@k корелюють між собою. Оптимальним значенням кількості латентних факторів є 900. Характер функції залежності є зростаючим до певного рівня, а після цього рівня знаходиться майже в одних й тих самих межах. Це свідчить, про те що було визначено оптимальну кількість прихованих факторів для даного набору даних. Кількість латентних факторів є найважливішим показником даної системи. Всі значення гіперпараметрів моделі є актуальними тільки для набору даних, що були досліджені в ході експерименту. Для нових вибірок процес аналізу даних на інших наборах потрібно розпочинати знову з самого початку за описаним вище алгоритмом.

Якість передбачення для даної задачі зросла з рівня 67% до 83%, що свідчить про ефективність застосування запропонованих методів обробки даних та прогнозування уподобань клієнтів. Забезпечення лояльності клієнтів до компанії є наразі основним пріоритетом у формуванні взаємовідносин компанії з клієнтами, надаючи їм саме ці товари і послуги високої якості, які вони потребують.

ДОДАТОК В

ПРОГРАМНА РЕАЛІЗАЦІЯ ОКРЕМИХ МОДУЛІВ ІСППР

В.1. Реалізація динамічного методу для телекомунікаційної компанії в середовищі SAS

```

libname mobile "c:\mobile";
data mobile.Sas1;
set mobile.Sas;
dur = MONTH_ID - DATE_FROM - 304;
run;

proc sql;
create table mobile.Concat as
select MSISDN, sum(ONNET_MINS) as ONNET_MINS_SUM, sum(OMO_MINS) as
OMO_MINS_SUM, sum(PSTN_MINS) as PSTN_MINS_SUM,
sum(INTERN_MINS)as INTERN_MINS_SUM, sum(GPRS_USG_MB)as
GPRS_USG_MB_SUM, max(dur)as DUR_MAX, STATUS
from mobile.Sas1
group by MSISDN, STATUS
;

quit;
ods rtf FILE="C:\mobile\resultConcat.rtf";
proc means data = mobile.Concat;
var DUR_MAX STATUS;
run;
proc univariate data = mobile.Concat;
var DUR_MAX;
run;
proc lifetest data = mobile.Concat plots = (s) notable;
time DUR_MAX*STATUS(0);
run;

proc lifetest data = mobile.Concat method = act plots = (s(name=Actsurv), h(name=Acthaz))
notable;
time DUR_MAX*STATUS(0);
run;

proc lifereg data=mobile.Concat;
model DUR_MAX*STATUS(0)= ONNET_MINS_SUM GPRS_USG_MB_SUM / d=exponential;
run;

proc lifereg data=mobile.Concat;
model DUR_MAX*STATUS(0)= ONNET_MINS_SUM OMO_MINS_SUM PSTN_MINS_SUM
INTERN_MINS_SUM GPRS_USG_MB_SUM / d=exponential;
run;

proc lifereg data=mobile.Concat;
model DUR_MAX*STATUS(0)= ONNET_MINS_SUM OMO_MINS_SUM PSTN_MINS_SUM
INTERN_MINS_SUM GPRS_USG_MB_SUM / d=weibull;
run;

```

```
proc lifereg data=mobile.Concat;
  model DUR_MAX*STATUS(0)= ONNET_MINS_SUM OMO_MINS_SUM PSTN_MINS_SUM
  INTERN_MINS_SUM GPRS_USG_MB_SUM / d=llogistic;
run;
```

```
proc phreg data=mobile.Concat;
  model DUR_MAX*STATUS(0)= ONNET_MINS_SUM OMO_MINS_SUM PSTN_MINS_SUM
  INTERN_MINS_SUM GPRS_USG_MB_SUM;
run;
```

```
ods rtf close;
```

```
/*Part A*/
```

```
ods graphics off;
```

```
proc phreg data=sasuser.recurrent covs(aggregate) covm;
  class treatment(param=ref ref='0');
  model (time0 time1)*status(0)= age treatment / ties=exact;
  id subjectid;
  title1 'Analysis of Repeated Events';
  title2 'using the Counting Process Model';
run;
```

```
/*Part B*/
```

```
data mcf;
  input age treatment;
  datalines;
40 0
40 1
;
run;
```

```
ods graphics on;
ods select mcfplot;
```

```
proc phreg data=sasuser.recurrent covs(aggregate) covm
  plots(overlay cl)=mcf;
  class treatment(param=ref ref='0');
  model (time0 time1)*status(0)= age treatment / ties=exact;
  id subjectid;
  baseline covariates=mcf / rowid=treatment;
run;
```

```
title;
```

```
libname mobile "c:\mobile";
DATA mobile.Sas;
  INFILE "C:\mobile\new_data.txt"
  DELIMITER = "09"x
  MISSEVER DSD
  FIRSTOBS=2
  LRECL=32767;
```

```

INPUT MSISDN DATE_FROM ddmmyy10. A $ OBLAST_ACTIVATED $ ONNET_MINS
OMO_MINS PSTN_MINS INTERN_MINS GPRS_USG_MB MONTH_ID yymm6. B $
SUBSCRIPTION_TYPE_CODE $ STATUS;
RUN;

```

```

ods pdf FILE="c:\mobile\_result.pdf";
data mobile.Sas1;
set mobile.Sas;
dur = MONTH_ID - DATE_FROM - 304;
run;

```

```

proc means data = mobile.Sas1;
var dur STATUS;
run;
proc univariate data = mobile.Sas1;
var dur;
run;
proc lifetest data = mobile.Sas1 plots = (s) notable;
time dur*STATUS(0);
run;

```

```

proc lifetest data = mobile.Sas1 method = act plots = (s(name=Actsurv), h(name=Acthaz)) notable;
time dur*STATUS(0);
run;

```

```

proc lifereg data=mobile.Sas1;
model dur*STATUS(0)= ONNET_MINS GPRS_USG_MB / d=exponential;
run;

```

```

proc lifereg data=mobile.Sas1;
model dur*STATUS(0)= ONNET_MINS OMO_MINS PSTN_MINS INTERN_MINS
GPRS_USG_MB / d=exponential;
run;

```

```

proc lifereg data=mobile.Sas1;
model dur*STATUS(0)= ONNET_MINS OMO_MINS PSTN_MINS INTERN_MINS
GPRS_USG_MB / d=weibull;
run;

```

```

proc lifereg data=mobile.Sas1;
model dur*STATUS(0)= ONNET_MINS OMO_MINS PSTN_MINS INTERN_MINS
GPRS_USG_MB / d=llogistic;
run;

```

```

proc phreg data=mobile.Sas1;
model dur*STATUS(0)= ONNET_MINS OMO_MINS PSTN_MINS INTERN_MINS
GPRS_USG_MB;
run;
ods pdf close;

```

В.2. Динамічні моделі для аналізу ризиків шахрайства на платформі ProZorro

```

<?xml version="1.0" encoding="UTF-8"?>
<_ROOT_ ORIENTATION="HORIZONTAL"
EMVERSION="13.2"><Workspace><NODES><NODE TOOLPREFIX="EMCODE"
TOOLTYPE="UTILITY" PARENTID="" DIAGRAMID="_ROOT_" DESCRIPTION="SAS
Code" Y="65" X="732" ID="EMCODE" COMPONENT="SASCode"
CLASS="SASHELP.EMCORE.EMCODETOOL.CLASS"><PROPERTIES
Component="SASCode" RunAction="Train" ForceRun="N" MetaAdvisor="BASIC"
ScoreCodeFormat="DATASTEP" PublishCode="PUBLISH" DataNeeded="N"
ToolType="UTILITY"
UsePriors="Y"><VariableProperties/><EmTargetAttributes/><EmVariableAttributes/><EmFilePro
perties><USERTRAINCODE Extension="sas" Type="EXTERNAL"><Data>
<![CDATA[ proc sort data=&EM_IMPORT_SCORE. out=&EM_LIB..sorted; by
EM_SUBHZRD1_SURV; run; proc print data=&EM_LIB..sorted (obs=100) NOOBS split='*
; var acctno EM_SUBHZRD1_SURV; label EM_SUBHZRD1_SURV='Probability of
Upgrade:*====='; title &quot;Top 100 With Lowest Probability at Forecast
Interval&quot;; run; ]]>
</Data></USERTRAINCODE><EMNOTES Extension="txt"
Type="EXTERNAL"><Data/></EMNOTES></EmFileProperties></PROPERTIES></NODE><N
ODE TOOLPREFIX="EMCODE" TOOLTYPE="UTILITY" PARENTID=""
DIAGRAMID="_ROOT_" DESCRIPTION="SAS Code (Add _t_ variable)" Y="170" X="324"
ID="EMCODE2" COMPONENT="SASCode"
CLASS="SASHELP.EMCORE.EMCODETOOL.CLASS"><PROPERTIES
Component="SASCode" RunAction="Train" ForceRun="N" MetaAdvisor="BASIC"
ScoreCodeFormat="DATASTEP" PublishCode="PUBLISH" DataNeeded="N"
ToolType="UTILITY"
UsePriors="Y"><VariableProperties/><EmTargetAttributes/><EmVariableAttributes/><EmFilePro
perties><USERTRAINCODE Extension="sas" Type="EXTERNAL"><Data>
<![CDATA[ data &EM_EXPORT_SCORE; set &EM_IMPORT_DATA; by acctno; if
last.acctno; format _currentdate MMDDYY10.0 ;
_currentdate=input(&quot;31DEC2000&quot;;anydtdte10.);
_t_=intck(&quot;MONTH&quot;;initdate, _currentdate); drop Target _currentdate; run; ]]>
</Data></USERTRAINCODE><EMNOTES Extension="txt"
Type="EXTERNAL"><Data/></EMNOTES></EmFileProperties></PROPERTIES></NODE><N
ODE TOOLPREFIX="Ids" TOOLTYPE="SAMPLE" PARENTID="" DIAGRAMID="_ROOT_"
DESCRIPTION="Phone Plan" Y="66" X="122" ID="Ids" COMPONENT="DataSource"
CLASS="SASHELP.EMSAMP.IDS.CLASS"><PROPERTIES Component="DataSource"
ForceRun="N" MetaAdvisor="BASIC" DropMapVariables="N" ApplyMaxClassLevels="Y"
Segment="" NBytes="459776" NCols="13" NObs="4564" DBPassThrough="Y"
SampleSizeObs="10000" SampleSizePercent="20" SampleSizeType="PERCENT" Sample="D"
NewVariableRole="REJECT" VariableValidation="STRICT" IdentifyEmptyColumns="Y"
MaxClassLevels="20" MaxPercentMissing="50" ApplyMaxPercentMissing="Y"
IntervalLowerLimit="20" ApplyIntervalLevelLowerLimit="Y" OutputType="VIEW"
NewTable="SAMPPIO.PHONEPLAN_TV_C" DataSelection="USERTABLE" DataSource=""
ComputeStatistics="N" Table="PHONEPLAN_TV_C" Library="SAMPPIO" Role="RAW"
Description=""><EmVariableAttributes><DELTA AttributeValue="ID" AttributeName="ROLE"
Variable="acctno"/><DELTA AttributeValue="TARGET" AttributeName="ROLE"
Variable="event"/><DELTA AttributeValue="NOMINAL" AttributeName="LEVEL"
Variable="event"/><DELTA AttributeValue="NOMINAL" AttributeName="LEVEL"

```

```

Variable="eventdate"/><DELTA AttributeValue="NOMINAL" AttributeName="LEVEL"
Variable="initdate"/><DELTA AttributeValue="BINARY" AttributeName="LEVEL"
Variable="paywithcc"/></EmVariableAttributes><EmTargetAttributes><event><DecData
Table="EMWS7.Ids_event_DD"><Data><Obs DECISION3="0" DECISION2="0"
DECISION1="1" DECPRIOR="." TRAINPRIOR="0.0317703769"
DATAPRIOR="0.0317703769" COUNT="145" EVENT="2"/><Obs DECISION3="0"
DECISION2="1" DECISION1="0" DECPRIOR="." TRAINPRIOR="0.3393952673"
DATAPRIOR="0.3393952673" COUNT="1549" EVENT="1"/><Obs DECISION3="1"
DECISION2="0" DECISION1="0" DECPRIOR="." TRAINPRIOR="0.6288343558"
DATAPRIOR="0.6288343558" COUNT="2870" EVENT="0"/></Data><Metadata><Variable
TYPE="C" FORMAT="" LENGTH="32" LABEL="" NAME="event"/><Variable TYPE="N"
FORMAT="10." LENGTH="8" LABEL="Level Counts" NAME="COUNT"/><Variable
TYPE="N" FORMAT="" LENGTH="8" LABEL="Data Proportions"
NAME="DATAPRIOR"/><Variable TYPE="N" FORMAT="" LENGTH="8" LABEL="Training
Proportions" NAME="TRAINPRIOR"/><Variable TYPE="N" FORMAT="" LENGTH="8"
LABEL="Decision Priors" NAME="DECPRIOR"/><Variable TYPE="N" FORMAT=""
LENGTH="8" LABEL="2" NAME="DECISION1"/><Variable TYPE="N" FORMAT=""
LENGTH="8" LABEL="1" NAME="DECISION2"/><Variable TYPE="N" FORMAT=""
LENGTH="8" LABEL="0" NAME="DECISION3"/></Metadata></DecData><DecMeta
Table="EMWS7.Ids_event_DM"><Data><Obs EVENT="" TYPE="" FORMAT="" LABEL=""
USE="N" COST="" ORDER="" LEVEL="PROFIT" VARIABLE=""
_TYPE_="MATRIX"/><Obs EVENT="2" TYPE="N" FORMAT="" LABEL="" USE=""
COST="" ORDER="" LEVEL="NOMINAL" VARIABLE="event" _TYPE_="TARGET"/><Obs
EVENT="" TYPE="N" FORMAT="" LABEL="2" USE="Y" COST="" ORDER="" LEVEL=""
VARIABLE="DECISION1" _TYPE_="DECISION"/><Obs EVENT="" TYPE="N" FORMAT=""
LABEL="1" USE="Y" COST="" ORDER="" LEVEL="" VARIABLE="DECISION2"
_TYPE_="DECISION"/><Obs EVENT="" TYPE="N" FORMAT="" LABEL="0" USE="Y"
COST="" ORDER="" LEVEL="" VARIABLE="DECISION3" _TYPE_="DECISION"/><Obs
EVENT="" TYPE="N" FORMAT="" LABEL="Data Prior" USE="N" COST="" ORDER=""
LEVEL="" VARIABLE="DATAPRIOR" _TYPE_="DATAPRIOR"/><Obs EVENT=""
TYPE="N" FORMAT="" LABEL="Training Prior" USE="N" COST="" ORDER="" LEVEL=""
VARIABLE="TRAINPRIOR" _TYPE_="TRAINPRIOR"/><Obs EVENT="" TYPE="N"
FORMAT="" LABEL="Decision Prior" USE="N" COST="" ORDER="" LEVEL=""
VARIABLE="DECPRIOR" _TYPE_="DECPRIOR"/><Obs EVENT="" TYPE="N" FORMAT=""
LABEL="Predicted: event=2" USE="" COST="" ORDER="" LEVEL="2"
VARIABLE="P_event2" _TYPE_="PREDICTED"/><Obs EVENT="" TYPE="N" FORMAT=""
LABEL="Residual: event=2" USE="" COST="" ORDER="" LEVEL="2"
VARIABLE="R_event2" _TYPE_="RESIDUAL"/><Obs EVENT="" TYPE="N" FORMAT=""
LABEL="Predicted: event=1" USE="" COST="" ORDER="" LEVEL="1"
VARIABLE="P_event1" _TYPE_="PREDICTED"/><Obs EVENT="" TYPE="N" FORMAT=""
LABEL="Residual: event=1" USE="" COST="" ORDER="" LEVEL="1"
VARIABLE="R_event1" _TYPE_="RESIDUAL"/><Obs EVENT="" TYPE="N" FORMAT=""
LABEL="Predicted: event=0" USE="" COST="" ORDER="" LEVEL="0"
VARIABLE="P_event0" _TYPE_="PREDICTED"/><Obs EVENT="" TYPE="N" FORMAT=""
LABEL="Residual: event=0" USE="" COST="" ORDER="" LEVEL="0"
VARIABLE="R_event0" _TYPE_="RESIDUAL"/><Obs EVENT="" TYPE="C" FORMAT=""
LABEL="From: event" USE="" COST="" ORDER="" LEVEL="" VARIABLE="F_event"
_TYPE_="FROM"/><Obs EVENT="" TYPE="C" FORMAT="" LABEL="Into: event" USE=""
COST="" ORDER="" LEVEL="" VARIABLE="I_event"
_TYPE_="INTO"/></Data><Metadata><Variable TYPE="C" FORMAT="" LENGTH="32"
LABEL="Type" NAME="_TYPE_"/><Variable TYPE="C" FORMAT="" LENGTH="32"
LABEL="Variable" NAME="VARIABLE"/><Variable TYPE="C" FORMAT="" LENGTH="40"

```

```

LABEL="Label" NAME="LABEL"/><Variable TYPE="C" FORMAT="" LENGTH="32"
LABEL="Measurement Level" NAME="LEVEL"/><Variable TYPE="C" FORMAT=""
LENGTH="32" LABEL="Target Event" NAME="EVENT"/><Variable TYPE="C" FORMAT=""
LENGTH="10" LABEL="Order" NAME="ORDER"/><Variable TYPE="C" FORMAT=""
LENGTH="32" LABEL="Format" NAME="FORMAT"/><Variable TYPE="C" FORMAT=""
LENGTH="1" LABEL="Type" NAME="TYPE"/><Variable TYPE="C" FORMAT=""
LENGTH="32" LABEL="Cost" NAME="COST"/><Variable TYPE="C" FORMAT=""
LENGTH="1" LABEL="Use"
NAME="USE"/></Metadata></DecMeta></event></EmTargetAttributes><EmFileProperties><E
MNOTES Extension="txt"
Type="EXTERNAL"><Data></EMNOTES></EmFileProperties></PROPERTIES></NODE><N
ODE TOOLPREFIX="Ids" TOOLTYPE="SAMPLE" PARENTID="" DIAGRAMID="_ROOT_"
DESCRIPTION="Phone Plan for Scoring" Y="170" X="126" ID="Ids2"
COMPONENT="DataSource" CLASS="SASHELP.EMSAMP.IDS.CLASS"><PROPERTIES
Component="DataSource" ForceRun="N" MetaAdvisor="BASIC" DropMapVariables="N"
ApplyMaxClassLevels="Y" Segment="" NBytes="459776" NCols="13" NObs="4564"
DBPassThrough="Y" SampleSizeObs="10000" SampleSizePercent="20"
SampleSizeType="PERCENT" Sample="D" NewVariableRole="REJECT"
VariableValidation="STRICT" IdentifyEmptyColumns="Y" MaxClassLevels="20"
MaxPercentMissing="50" ApplyMaxPercentMissing="Y" IntervalLowerLimit="20"
ApplyIntervalLevelLowerLimit="Y" OutputType="VIEW"
NewTable="SAMPSIO.PHONEPLAN_TVC" DataSelection="USERTABLE" DataSource=""
ComputeStatistics="N" Table="PHONEPLAN_TVC" Library="SAMPSIO" Role="RAW"
Description=""><EmVariableAttributes><DELTA AttributeValue="ID" AttributeName="ROLE"
Variable="acctno"/><DELTA AttributeValue="NOMINAL" AttributeName="LEVEL"
Variable="event"/><DELTA AttributeValue="NOMINAL" AttributeName="LEVEL"
Variable="eventdate"/><DELTA AttributeValue="NOMINAL" AttributeName="LEVEL"
Variable="initdate"/><DELTA AttributeValue="BINARY" AttributeName="LEVEL"
Variable="paywithcc"/></EmVariableAttributes><EmTargetAttributes/><EmFileProperties><EM
NOTES Extension="txt"
Type="EXTERNAL"><Data></EMNOTES></EmFileProperties></PROPERTIES></NODE><N
ODE TOOLPREFIX="SURV" TOOLTYPE="APPS" PARENTID="" DIAGRAMID="_ROOT_"
DESCRIPTION="Survival" Y="66" X="329" ID="SURV" COMPONENT="Survival"
CLASS="SASHELP.EMCORE.EXTENSION.CLASS"><PROPERTIES Component="Survival"
RunAction="Train" ForceRun="N" SCORETIME="" CENSTIME="" TRUNCTIME=""
EXTRAPOINT="50" NTU="1" SCRTYPE="INT" Interactions="" NUMKNOTS="5"
NUMHIGHRISK="100" RiskPercentile="1" RiskTable="FIXEDNUMBER"
REPORTTABLES="NONE" ForecastTimeUnit="6" fctunitdefault="Y" MAXLIFE="60"
MAXMRLDEFAULT="Y" MRL="NONE" INTLEN="2" ModelValidation="DEFAULT"
KNOT_STEPWISE="N" SLSTAY="0.05" SLENTY="0.05" STEPWISE="N"
TIMECOVINTS="S" SampSeed="12345" SPARAM="0.2" SAMPLING="N"
TRUNCATION="N" TIMEUNIT="MONTH" DATAFORMAT="CHANGETIME"
Catalog="SASHELP.EMAPPS.SURVIVAL.SOURCE"
Location="CATALOG"><VariableProperties/><EmTargetAttributes/><EmVariableAttributes/><E
mFileProperties><USERTRAINCODE Extension="sas"
Type="EXTERNAL"><Data></USERTRAINCODE><TIMEIDTABLE
Type="DATA"><Data><Obs TIMEID="initdate" ROLE="Start Time Variable"/><Obs
TIMEID="eventdate" ROLE="End Time Variable"/><Obs TIMEID="changedate"
ROLE="Change-Time Variable"/></Data><Metadata><Variable TYPE="C" FORMAT=""
LENGTH="32" LABEL="Time ID Role" NAME="Role"/><Variable TYPE="C" FORMAT=""
LENGTH="32" LABEL="Time ID Variable"
NAME="TimeID"/></Metadata></TIMEIDTABLE><TRAINTIMESELECTION

```

```

Type="DATA"><Data><Obs    DATE="14487"    _INDEX_="1"/><Obs    DATE="14517"
_INDEX_="0"/><Obs    DATE="14548"    _INDEX_="0"/><Obs    DATE="14578"
_INDEX_="0"/><Obs    DATE="14609"    _INDEX_="0"/><Obs    DATE="14640"
_INDEX_="0"/><Obs    DATE="14669"    _INDEX_="0"/><Obs    DATE="14700"
_INDEX_="0"/><Obs    DATE="14730"    _INDEX_="0"/><Obs    DATE="14761"
_INDEX_="0"/><Obs    DATE="14791"    _INDEX_="0"/><Obs    DATE="14822"
_INDEX_="0"/><Obs    DATE="14853"    _INDEX_="0"/><Obs    DATE="14883"
_INDEX_="0"/><Obs    DATE="14914"    _INDEX_="0"/><Obs    DATE="14944"
_INDEX_="0"/><Obs    DATE="14975"    _INDEX_="2"/><Obs    DATE="15006"
_INDEX_="0"/></Data><Metadata><Variable    TYPE="N"    FORMAT=""    LENGTH="8"
LABEL=""    NAME="_INDEX_"><Variable    TYPE="N"    FORMAT="DATE9."    LENGTH="8"
LABEL=""
NAME="DATE"/></Metadata></TRAINTIMESELECTION><SCORETIMESELECTION
Type="DATA"><Data><Obs    DATE="14457"    _INDEX_="0"/><Obs    DATE="14488"
_INDEX_="0"/><Obs    DATE="14518"    _INDEX_="0"/><Obs    DATE="14549"
_INDEX_="0"/><Obs    DATE="14579"    _INDEX_="0"/><Obs    DATE="14610"
_INDEX_="0"/><Obs    DATE="14641"    _INDEX_="0"/><Obs    DATE="14670"
_INDEX_="0"/><Obs    DATE="14701"    _INDEX_="0"/><Obs    DATE="14731"
_INDEX_="0"/><Obs    DATE="14762"    _INDEX_="0"/><Obs    DATE="14792"
_INDEX_="0"/><Obs    DATE="14823"    _INDEX_="0"/><Obs    DATE="14854"
_INDEX_="1"/><Obs    DATE="14884"    _INDEX_="0"/><Obs    DATE="14915"
_INDEX_="0"/><Obs    DATE="14945"    _INDEX_="0"/></Data><Metadata><Variable    TYPE="N"
FORMAT="DATE9."    LENGTH="8"    LABEL=""    NAME="DATE"/><Variable    TYPE="N"
FORMAT=""    LENGTH="8"    LABEL=""
NAME="_INDEX_"></Metadata></SCORETIMESELECTION><INTERACTION
Type="DATA"><Data><Obs    TERM="cum_dial_add"    KEY="0"/><Obs    TERM="cum_dsl_add"
KEY="1"/><Obs    TERM="cum_isdn_add"    KEY="2"/><Obs    TERM="n_dial"    KEY="3"/><Obs
TERM="n_dsl"    KEY="4"/><Obs    TERM="n_isdn"    KEY="5"/></Data><Metadata><Variable
TYPE="N"    FORMAT=""    LENGTH="8"    LABEL=""    NAME="key"/><Variable    TYPE="C"
FORMAT=""    LENGTH="32"    LABEL=""
NAME="Term"/></Metadata></INTERACTION><EMNOTES    Extension="txt"
Type="EXTERNAL"><Data/></EMNOTES></EmFileProperties></PROPERTIES></NODE><N
ODE    TOOLPREFIX="Score"    TOOLTYPE="ASSESS"    PARENTID=""
DIAGRAMID="_ROOT_"    DESCRIPTION="Score"    Y="66"    X="534"    ID="Score"
COMPONENT="Score"    CLASS="SASHELP.EMUTIL.EMSCORE.CLASS"><PROPERTIES
Component="Score"    RunAction="Train"    ForceRun="N"    OutputType="VIEW"    GraphReports="Y"
OptimizedCode="Y"    PreferenceName=""    UserPackageName=""    PackageName="DEFAULT"
JScore="Y"    CScore="Y"    ScoreTest="N"    ScoreValidate="N"    HideOther="Y"    HideResidual="Y"
HideClassification="Y"    HideFreq="Y"    HidePredict="Y"    HideAssess="Y"    HideRejected="Y"
HideTarget="Y"    HideInput="Y"    HideVariables="N"
FixedOutputNames="Y"><VariableProperties/><EmTargetAttributes/><EmVariableAttributes/><
EmFileProperties><EMNOTES    Extension="txt"
Type="EXTERNAL"><Data/></EMNOTES></EmFileProperties></PROPERTIES></NODE><N
ODES><CONNECTIONS><CONNECTION    TO="EMCODE"
FROM="Score"/><CONNECTION    TO="Score"    FROM="SURV"/><CONNECTION
TO="SURV"    FROM="Ids"/><CONNECTION    TO="EMCODE2"
FROM="Ids2"/><CONNECTION    TO="Score"
FROM="EMCODE2"/></CONNECTIONS><NOTES><DATA>
<![CDATA[ ]]>
</DATA></NOTES></Workspace></_ROOT_>

```

В.3. Реалізація на мові R модулів інформаційної технології оцінювання фінансових ризиків онлайн-гемблінгу

fitting_models.R:

```

library(RODBC)
library(dplyr)
library(reshape2)
library(ROCR)
library(xgboost)
library(zoo)

con = odbcConnect('PostgreSQL35W')
df = sqlQuery(con, 'select
    t.*,
    coalesce(t.poker_hour_cnt, 0) > 0 is_poker,
    coalesce(t.casino_hour_cnt, 0) > 0 is_casino,
    date = max(case when poker_hour_cnt > 0 then date end) over (partition by user_id)
last_poker,
    date = max(case when casino_hour_cnt > 0 then date end) over (partition by user_id)
last_casino
    from public.vip_scoring_view t
    ;')

#FUNCTIONS
get_essential = function(absence_measure, lower_bound, target.quantile) {
  #function for target field
  total_diff = sum(absence_measure)
  diff_sum = ifelse(total_diff == 0, 1, cumsum(absence_measure)/total_diff)
  return(ifelse(diff_sum <= target.quantile & absence_measure > lower_bound, 1, 0))
}

prepare_data = function(df0) {
  #function for preparing data

  #Preparing lagging 1-day [1-week, 1-month (30-days)]
  df3 = df0 %>%
    select(user_id, date) %>%
    mutate(date1 = date - 1) %>%
    melt(id.vars = c('user_id'), value.name = 'date') %>%
    select(user_id, date) %>%
    distinct() %>%
    left_join(df, by = c('user_id' = 'user_id', 'date' = 'date'))

  #LOCF-ing lagged 1-day [1-week, 1-month (30-days)]
  df4 = df3 %>%
    select(user_id, date,
           cum_avg_rake,
           cum_avg_bonuses,
           cum_dep_in_sum_over_cum_dep_in_cnt,
           cum_dep_out_sum_over_cum_dep_out_cnt,
           cum_ng_sum,

```



```

    date0,
    age) %>%
arrange(user_id, date) %>%
group_by(user_id) %>%
mutate_all(funs(na.locf(., na.rm = F))) %>%
mutate(absence_weight = as.numeric(date - date0),
       cum_avg_rake = cum_avg_rake*age/(age + absence_weight),
       cum_avg_bonuses = cum_avg_bonuses*age/(age + absence_weight)) %>%
select(user_id, date,
       cum_avg_rake,
       cum_avg_bonuses,
       cum_dep_in_sum_over_cum_dep_in_cnt,
       cum_dep_out_sum_over_cum_dep_out_cnt,
       cum_ng_sum) %>%
ungroup()

df3 = df3 %>%
select(-c(cum_avg_rake,
         cum_avg_bonuses,
         cum_dep_in_sum_over_cum_dep_in_cnt,
         cum_dep_out_sum_over_cum_dep_out_cnt,
         cum_ng_sum,
         age)) %>%
filter(!is.na(date)) %>%
mutate_all(funs(na.fill(., 0, na.rm = F)))

df5 = df0 %>%
mutate(date1 = date - 1) %>%
inner_join(df3, by = c('user_id' = 'user_id', 'date1' = 'date'), suffix = c("", '.day')) %>%
inner_join(df4, by = c('user_id' = 'user_id', 'date1' = 'date'), suffix = c("", '.day')) %>%
filter(!is.na(cum_avg_rake.day))

return(df5)
}

createWatchlist = function(df, labels, train_id) {
#function for creating watchlist for xgboost
dtrain = xgb.DMatrix(data.matrix(df[train_id, ]), label = labels[train_id])
dtest = xgb.DMatrix(data.matrix(df[-train_id, ]), label = labels[-train_id])
list(eval = dtest, train = dtrain)
}

get_cutoff = function(model, newdata, labels) {
pred = prediction(predict(model, newdata = newdata), labels)
perf = performance(pred, 'tpr', 'fpr')
cutoffs = data.frame(cut = perf@alpha.values[[1]], fpr = perf@x.values[[1]], tpr =
perf@y.values[[1]])
cutoffs = cutoffs %>%
mutate(min_fpr = ceiling(fpr*100)/100) %>%
group_by(min_fpr) %>%
summarise(max_tpr = max(tpr),
         min_cut = min(cut))
}

```

```

)
return(list(cutoffs = cutoffs, perf = perf))
}

plot_feature_importance = function(xgb.model, df) {
  names <- dimnames(data.matrix(df))[[2]]
  importance_matrix <- xgb.importance(names, model = xgb.model)
  xgb.plot.importance(importance_matrix[1:20, ])
}

df = df %>% filter(cashout_sum >= 0) %>% select(-c(bets_cnt, wins_cnt, ng_sum,
  bonus_sum, is_only_bonus,
  cum_rake, cum_bonus_sum, cum_dep_in_sum, cum_dep_out_sum,
  avg_seats, avg_stakes, rub_minutes_share, usd_minutes_share,
  days_from_last_dep
))

df$date0 = df$date

for(i in 3:(dim(df)[2] - 1)) {
  df[[i]][is.na(df[[i]])] = 0
}

now = Sys.Date()

#FOR POKER
poker.df = df %>%
  filter(is_poker == 1) %>%
  group_by(user_id) %>%
  arrange(date) %>%
  mutate(absence_measure = as.numeric(lead(date, default = as.Date(now)) - date) - 1) %>%
  arrange(desc(absence_measure)) %>%
  mutate(is_bad = get_essential(absence_measure, 3, 0.975)) %>%

  mutate(max_absence = max(absence_measure, na.rm = T)) %>%
  filter(max_absence > 14) %>%
  select(-max_absence) %>%

  ungroup()

poker.df5 = prepare_data(poker.df)

# poker.df.eval = poker.df5 %>% filter(last_poker == 1)

poker.df.model = poker.df5 %>% filter(last_poker == 0)

poker.df.labels = poker.df.model$is_bad
poker.df.fin = poker.df.model %>%
  select(-c(user_id, date, is_poker, is_poker.day,
    is_casino, is_casino.day, date1,
    last_poker, last_poker.day,
    last_casino, last_casino.day,

```

```

  absence_measure, is_bad))

#FIT MODEL
set.seed(666)
poker.train_id = createDataPartition(poker.df.labels, p = 0.8, list = F)
poker.watchlist = createWatchlist(poker.df.fin, poker.df.labels, poker.train_id)
poker.param = list(max.depth = 5, eta = 0.05, objective = 'binary:logistic', eval_metric = 'auc')

poker.xgb.fit = xgb.train(
  params = poker.param,
  data = poker.watchlist$train,
  nrounds = 600,
  watchlist = poker.watchlist,
  seed = 666
)

saveRDS(poker.xgb.fit, 'poker.xgb.fit v2')

#POKER CUTOFFS
poker.cutoffs = get_cutoff(poker.xgb.fit, data.matrix(poker.df.fin), poker.df.labels)

sqlQuery(con, 'truncate forecast.cutoffs_poker_2')

sqlSave(con,
  poker.cutoffs$cutoffs,
  'forecast.cutoffs_poker_2',
  rownames = F,
  append = T)

#FOR CASINO
casino.df = df %>%
  filter(is_casino == 1) %>%
  group_by(user_id) %>%
  arrange(date) %>%
  mutate(absence_measure = as.numeric(lead(date, default = as.Date(now)) - date) - 1) %>%
  arrange(desc(absence_measure)) %>%
  mutate(is_bad = get_essential(absence_measure, 3, 0.975),
    date0 = date) %>%

  mutate(max_absence = max(absence_measure, na.rm = T)) %>%
  filter(max_absence > 14) %>%
  select(-max_absence) %>%

  ungroup()

casino.df5 = prepare_data(casino.df)

# casino.df.eval = casino.df5 %>% filter(last_casino == 1)

casino.df.model = casino.df5 %>% filter(last_casino == 0)

casino.df.labels = casino.df.model$is_bad

```

```

casino.df.fin = casino.df.model %>%
  select(-c(user_id, date, is_poker, is_poker.day,
            is_casino, is_casino.day, date0, date1, date0.day,
            last_poker, last_poker.day,
            last_casino, last_casino.day,
            absence_measure, is_bad))

#FIT MODEL
set.seed(666)
casino.train_id = createDataPartition(casino.df.labels, p = 0.8, list = F)
casino.watchlist = createWatchlist(casino.df.fin, casino.df.labels, casino.train_id)
casino.param = list(max.depth = 3, eta = 0.1, objective = 'binary:logistic', eval_metric = 'auc')

casino.xgb.fit = xgb.train(
  params = casino.param,
  data = casino.watchlist$train,
  nrounds = 200,
  watchlist = casino.watchlist,
  seed = 666
)

saveRDS(casino.xgb.fit, 'casino.xgb.fit v2')

#CASINO CUTOFF
casino.cutoffs = get_cutoff(casino.xgb.fit, data.matrix(casino.df.fin), casino.df.labels)

sqlQuery(con, 'truncate forecast.cutoffs_casino_2')

sqlSave(con,
  casino.cutoffs$cutoffs,
  'forecast.cutoffs_casino_2',
  rownames = F,
  append = T)

close(con)

eval_score.R:

# library(RODBC)
library(RPostgreSQL)
library(xgboost)
library(dplyr)
library(reshape2)
library(zoo)

#code for getting folder of the script
initial.options = commandArgs(trailingOnly = FALSE)
file.arg.name = "--file="
script.name = sub(file.arg.name, "", initial.options[grep(file.arg.name, initial.options)])
script.basename = dirname(script.name)
setwd(script.basename)

```

```

# con = odbcConnect('rdfraud')
drv = dbDriver('PostgreSQL')
con = dbConnect(drv, dbname = "cgames",
  host = "localhost", port = 5432,
  user = "rdfraud", password = 'IONjVGHzyvbsy12')

# df = sqlQuery(con, 'select
#   t.*,
#   coalesce(t.poker_hour_cnt, 0) > 0 is_poker,
#   coalesce(t.casino_hour_cnt, 0) > 0 is_casino,
#   date = max(case when poker_hour_cnt > 0 then date end) over (partition by user_id)
last_poker,
#   date = max(case when casino_hour_cnt > 0 then date end) over (partition by user_id)
last_casino
#   from vip_scoring.scoring_model_view t
#   ;')
#
# sqlQuery(con, 'truncate vip_scoring.vip_score')

df = dbGetQuery(con, 'select
  t.*,
  coalesce(t.poker_hour_cnt, 0) > 0 is_poker,
  coalesce(t.casino_hour_cnt, 0) > 0 is_casino,
  date = max(case when poker_hour_cnt > 0 then date end) over (partition by user_id)
last_poker,
  date = max(case when casino_hour_cnt > 0 then date end) over (partition by user_id)
last_casino
  from vip_scoring.scoring_model_view t
  ;')
dbSendQuery(con, 'truncate vip_scoring.vip_score')

prepare_data = function(df0) {
  #function for preparing data

  #Preparing lagging 1-day [1-week, 1-month (30-days)]
  df3 = df0 %>%
    select(user_id, date) %>%
    mutate(date1 = date - 1) %>%
    melt(id.vars = c('user_id'), value.name = 'date') %>%
    select(user_id, date) %>%
    distinct() %>%
    left_join(df, by = c('user_id' = 'user_id', 'date' = 'date'))

  #LOCF-ing lagged 1-day [1-week, 1-month (30-days)]
  df4 = df3 %>%
    select(user_id, date,
      cum_avg_rake,
      cum_avg_bonuses,
      cum_dep_in_sum_over_cum_dep_in_cnt,
      cum_dep_out_sum_over_cum_dep_out_cnt,
      cum_ng_sum,
      date0,

```

```

    age) %>%
  arrange(user_id, date) %>%
  group_by(user_id) %>%
  mutate_all(funs(na.locf(., na.rm = F))) %>%
  mutate(absence_weight = as.numeric(date - date0),
         cum_avg_rake = cum_avg_rake*age/(age + absence_weight),
         cum_avg_bonuses = cum_avg_bonuses*age/(age + absence_weight)) %>%
  select(user_id, date,
         cum_avg_rake,
         cum_avg_bonuses,
         cum_dep_in_sum_over_cum_dep_in_cnt,
         cum_dep_out_sum_over_cum_dep_out_cnt,
         cum_ng_sum) %>%
  ungroup()

df3 = df3 %>%
  select(-c(cum_avg_rake,
           cum_avg_bonuses,
           cum_dep_in_sum_over_cum_dep_in_cnt,
           cum_dep_out_sum_over_cum_dep_out_cnt,
           cum_ng_sum,
           age)) %>%
  filter(!is.na(date)) %>%
  mutate_all(funs(na.fill(., 0, na.rm = F)))

df5 = df0 %>%
  mutate(date1 = date - 1) %>%
  inner_join(df3, by = c('user_id' = 'user_id', 'date1' = 'date'), suffix = c("", '.day')) %>%
  inner_join(df4, by = c('user_id' = 'user_id', 'date1' = 'date'), suffix = c("", '.day')) %>%
  filter(!is.na(cum_avg_rake.day))

return(df5)
}

get_essential = function(absence_measure, lower_bound, target.quantile) {
  #function for target field
  total_diff = sum(absence_measure)
  diff_sum = ifelse(total_diff == 0, 1, cumsum(absence_measure)/total_diff)
  max_absence = max(absence_measure)
  return(ifelse(diff_sum <= target.quantile & absence_measure > ifelse(max_absence <= 3, 1,
lower_bound), 1, 0))
}

df = df %>% filter(cashout_sum >= 0) %>% select(-c(bets_cnt, wins_cnt, ng_sum,
        bonus_sum, is_only_bonus,
        cum_rake, cum_bonus_sum, cum_dep_in_sum, cum_dep_out_sum,
        avg_seats, avg_stakes, rub_minutes_share, usd_minutes_share,
        days_from_last_dep
))

df$date0 = df$date

```

```

for(i in 3:(dim(df)[2] - 1)) {
  df[[i]][is.na(df[[i])] = 0
}

now = Sys.Date()

#EVAL POKER
poker.df = df %>%
  filter(is_poker == 1) %>%
  mutate(date0 = date)

poker.df5 = prepare_data(poker.df)

poker.df.eval = poker.df5 %>% filter(last_poker == 1)

poker.fin = poker.df.eval %>%
  select(-c(user_id, date, is_poker, is_poker.day,
            is_casino, is_casino.day, date1,
            last_poker, last_poker.day,
            last_casino, last_casino.day))

poker.xgb.fit = readRDS('poker.xgb.fit v2')

poker.pred = predict(poker.xgb.fit, newdata = data.matrix(poker.fin))

poker.df.m = df %>%
  filter(is_poker == 1) %>%
  group_by(user_id) %>%
  arrange(date) %>%
  mutate(absence_measure = as.numeric(lead(date, default = as.Date(now)) - date) - 1) %>%
  arrange(desc(absence_measure)) %>%
  mutate(is_bad = get_essential(absence_measure, 3, 0.975)) %>%
  filter(is_bad == 1) %>%
  summarise(m = median(absence_measure))

tmp.db = data.frame(
  model_date = as.Date(now),
  type = 'poker',
  user_id = poker.df.eval$user_id,
  last_date = as.Date(poker.df.eval$date),
  score0 = poker.pred
) %>%
  inner_join(poker.df.m, by = 'user_id') %>%
  mutate(score1 = ifelse(score0 < 0.95, 0.95, score0 + (1 - score0)/2),
         mu = m*log((1-score0)/score0)/log(((1-score0)*score1)/(score0*(1-score1))),
         s = m/log(((1-score0)*score1)/(score0*(1-score1))),
         n = as.numeric(now - last_date) - 1,
         score = 1/(1 + exp(-(n - mu)/s))
  )

# sqlSave(con,
#   tmp.db,

```

```

# 'vip_scoring.vip_score',
# rownames = F,
# append = T
# )
dbWriteTable(con,
  c('vip_scoring', 'vip_score'),
  tmp.db,
  row.names = F,
  append = T)

#EVAL CASINO
casino.df = df %>%
  filter(is_casino == 1) %>%
  mutate(date0 = date)

casino.df5 = prepare_data(casino.df)

casino.df.eval = casino.df5 %>% filter(last_casino == 1)

casino.fin = casino.df.eval %>%
  select(-c(user_id, date, is_poker, is_poker.day,
    is_casino, is_casino.day, date0, date1, date0.day,
    last_poker, last_poker.day,
    last_casino, last_casino.day))

casino.xgb.fit = readRDS('casino.xgb.fit v2')

casino.pred = predict(casino.xgb.fit, newdata = data.matrix(casino.fin))

casino.df.m = df %>%
  filter(is_casino == 1) %>%
  group_by(user_id) %>%
  arrange(date) %>%
  mutate(absence_measure = as.numeric(lead(date, default = as.Date(now)) - date) - 1) %>%
  arrange(desc(absence_measure)) %>%
  mutate(is_bad = get_essential(absence_measure, 3, 0.975)) %>%
  filter(is_bad == 1) %>%
  summarise(m = median(absence_measure))

tmp.db = data.frame(
  model_date = now,
  type = 'casino',
  user_id = casino.df.eval$user_id,
  last_date = as.Date(casino.df.eval$date),
  score0 = casino.pred
) %>%
  inner_join(casino.df.m, by = 'user_id') %>%
  mutate(score1 = ifelse(score0 < 0.95, 0.95, score0 + (1 - score0)/2),
    mu = m*log((1-score0)/score0)/log(((1-score0)*score1)/(score0*(1-score1))),
    s = m/log(((1-score0)*score1)/(score0*(1-score1))),
    n = as.numeric(now - last_date) - 1,
    score = 1/(1 + exp(-(n - mu)/s))

```



```

)

# sqlSave(con,
#   tmp.db,
#   'vip_scoring.vip_score',
#   rownames = F,
#   append = T
# )

dbWriteTable(con,
  c('vip_scoring', 'vip_score'),
  tmp.db,
  row.names = F,
  append = T)

# close(con)
dbDisconnect(con)

scoring_model_view.sql:

create view scoring_model_view as
SELECT fs.user_id,
  fs.date,
  fs.bets_cnt,
  fs.wins_cnt,
  fs.bets_sum,
  fs.wins_sum,
  fs.ng_sum,
  fs.buyin_sum,
  fs.rebuyin_sum,
  fs.cashout_sum,
  fs.payout_sum,
  fs.pokerbet_sum,
  fs.rake_sum,
  fs.casino_hour_cnt,
  fs.poker_hour_cnt,
  fs.bonus_sum,
  fs.dep_in_cnt,
  fs.dep_in_sum,
  fs.dep_out_cnt,
  fs.dep_out_sum,
  fs.is_only_bonus,
  fs.avg_seats,
  fs.avg_stakes,
  fs.rub_minutes_share,
  fs.usd_minutes_share,
  COALESCE(sum(fs.rake_sum) OVER (PARTITION BY fs.user_id ORDER BY fs.user_id,
fs.date), (0)::double precision) AS cum_rake,
  ((fs.date - min(fs.date) OVER (PARTITION BY fs.user_id)) + 1) AS age,
  (COALESCE(sum(fs.rake_sum) OVER (PARTITION BY fs.user_id ORDER BY fs.user_id,
fs.date), (0)::double precision) / (((fs.date - min(fs.date) OVER (PARTITION BY fs.user_id)) +
1))::double precision) AS cum_avg_rake,

```

```

CASE
  WHEN ((fs.dep_out_sum IS NULL) AND (vp.id IS NULL)) THEN NULL::integer
  ELSE ntile(4) OVER (PARTITION BY
CASE
  WHEN ((fs.dep_out_sum IS NULL) AND (vp.id IS NULL)) THEN 1
  ELSE 2
END ORDER BY fs.dep_out_sum)
END AS quart_dep_out_sum,
CASE
  WHEN ((fs.dep_in_sum IS NULL) AND (vp.id IS NULL)) THEN NULL::integer
  ELSE ntile(4) OVER (PARTITION BY
CASE
  WHEN ((fs.dep_in_sum IS NULL) AND (vp.id IS NULL)) THEN 1
  ELSE 2
END ORDER BY fs.dep_in_sum)
END AS quart_dep_in_sum,
CASE
  WHEN (((fs.wins_sum IS NULL) AND (fs.bets_sum IS NULL)) OR ((fs.wins_sum =
(0)::double precision) AND (fs.bets_sum = (0)::double precision))) AND (vp.id IS NULL)) THEN
NULL::integer
  ELSE ntile(4) OVER (PARTITION BY
CASE
  WHEN (((fs.wins_sum IS NULL) AND (fs.bets_sum IS NULL)) OR ((fs.wins_sum =
(0)::double precision) AND (fs.bets_sum = (0)::double precision) AND (vp.id IS NULL))) THEN 1
  ELSE 2
END ORDER BY fs.wins_sum)
END AS quart_wins_sum,
CASE
  WHEN (((fs.wins_sum IS NULL) AND (fs.bets_sum IS NULL)) OR ((fs.wins_sum =
(0)::double precision) AND (fs.bets_sum = (0)::double precision))) AND (vp.id IS NULL)) THEN
NULL::integer
  ELSE ntile(4) OVER (PARTITION BY
CASE
  WHEN (((fs.wins_sum IS NULL) AND (fs.bets_sum IS NULL)) OR ((fs.wins_sum =
(0)::double precision) AND (fs.bets_sum = (0)::double precision) AND (vp.id IS NULL))) THEN 1
  ELSE 2
END ORDER BY fs.bets_sum)
END AS quart_bets_sum,
COALESCE(sum(fs.bonus_sum) OVER (PARTITION BY fs.user_id ORDER BY fs.user_id,
fs.date), (0)::double precision) AS cum_bonus_sum,
(COALESCE(sum(fs.bonus_sum) OVER (PARTITION BY fs.user_id ORDER BY fs.user_id,
fs.date), (0)::double precision) / (((fs.date - min(fs.date) OVER (PARTITION BY fs.user_id)) +
1))::double precision) AS cum_avg_bonuses,
COALESCE(sum(fs.dep_in_cnt) OVER (PARTITION BY fs.user_id ORDER BY fs.user_id,
fs.date), (0)::numeric) AS cum_dep_in_cnt,
COALESCE(sum(fs.dep_in_sum) OVER (PARTITION BY fs.user_id ORDER BY fs.user_id,
fs.date), (0)::double precision) AS cum_dep_in_sum,
(COALESCE(sum(fs.dep_in_sum) OVER (PARTITION BY fs.user_id ORDER BY fs.user_id,
fs.date), (0)::double precision) / (COALESCE(NULLIF(sum(fs.dep_in_cnt) OVER (PARTITION
BY fs.user_id ORDER BY fs.user_id, fs.date), (0)::numeric), (1)::numeric))::double precision) AS
cum_dep_in_sum_over_cum_dep_in_cnt,

```

```

COALESCE(sum(fs.dep_out_cnt) OVER (PARTITION BY fs.user_id ORDER BY fs.user_id,
fs.date), (0)::numeric) AS cum_dep_out_cnt,
COALESCE(sum(fs.dep_out_sum) OVER (PARTITION BY fs.user_id ORDER BY fs.user_id,
fs.date), (0)::double precision) AS cum_dep_out_sum,
(COALESCE(sum(fs.dep_out_sum) OVER (PARTITION BY fs.user_id ORDER BY fs.user_id,
fs.date), (0)::double precision) / (COALESCE(NULLIF(sum(fs.dep_out_cnt) OVER (PARTITION
BY fs.user_id ORDER BY fs.user_id, fs.date), (0)::numeric), (1)::numeric))::double precision) AS
cum_dep_out_sum_over_cum_dep_out_cnt,
COALESCE(sum(fs.ng_sum) OVER (PARTITION BY fs.user_id ORDER BY fs.user_id,
fs.date), (0)::double precision) AS cum_ng_sum,
(fs.date - max(
CASE
WHEN (fs.dep_in_cnt > 0) THEN fs.date
ELSE NULL::date
END) OVER (PARTITION BY fs.user_id ORDER BY fs.date)) AS days_from_last_dep
FROM (vip_scoring.scoring_model_sample fs
JOIN vips vp ON ((vp.id = fs.user_id)));

```

vip_score_view.sql:

```

create view vip_score_view as
WITH tmp_games AS (
SELECT vs_1.user_id,
(sum(sp.rakeusd) / (3)::double precision) AS rake,
(sum(sp.ngusd) / (3)::double precision) AS ng
FROM (vip_scoring.vip_score vs_1
JOIN pomodoro_spins sp ON ((sp.user_id = vs_1.user_id)))
WHERE (true AND (sp.date >= (vs_1.last_date - '21 days'::interval)))
GROUP BY vs_1.user_id
), tmp_dep AS (
SELECT vs_1.user_id,
percentile_disc((0.5)::double precision) WITHIN GROUP (ORDER BY dep.sumusd) AS
dep_median
FROM (vip_scoring.vip_score vs_1
JOIN aqua_curr_deps dep ON (((vs_1.user_id = dep.id) AND (dep.isin = 1) AND (dep.depdata
<= vs_1.last_date))))
GROUP BY vs_1.user_id
), tmp_vip AS (
SELECT vs_1.user_id,
array_agg(vs_1.type ORDER BY vs_1.type) AS type,
array_agg(vs_1.last_date ORDER BY vs_1.type) AS last_date,
array_agg(vs_1.score ORDER BY vs_1.type) AS score,
array_agg(
CASE
WHEN (vs_1.type = 'poker'::text) THEN (vip_scoring.get_poker_cutoff() < vs_1.score)
WHEN (vs_1.type = 'casino'::text) THEN (vip_scoring.get_casino_cutoff() < vs_1.score)
ELSE NULL::boolean
END ORDER BY vs_1.type) AS is_bad
FROM vip_scoring.vip_score vs_1
GROUP BY vs_1.user_id
)
)
SELECT vs.user_id,

```

```

vs.type,
vs.last_date,
vs.score,
vs.is_bad,
tg.rake,
tg.ng,
td.dep_median
FROM ((tmp_vip vs
LEFT JOIN tmp_games tg ON ((vs.user_id = tg.user_id)))
LEFT JOIN tmp_dep td ON ((tg.user_id = td.user_id)));

```

В.4. Реалізація на мові Python окремих мікросервісів для інтеграції в існуючу ERP-систему

Лістинг файлу load_data_als.py:

```

import pandas as pd
import numpy as np
import datetime as dt
import matplotlib.pyplot as plt
import matplotlib.pyplot as pylab
from sklearn.cluster import KMeans
from itertools import cycle
from sklearn import preprocessing
from mpl_toolkits.mplot3d import Axes3D
import pymysql
import seaborn as sb
import json as json
import scipy as sp
import scipy.sparse as sparse

df_sales_raw = pd.read_csv("./data_cf_fimin_bachelor.csv")
df_sales_raw.info()
data = df_sales_raw[['i_cnt', 'i_ext_id', 'u_id']].dropna()
print(data.shape)
X = data

X['I'] = X.i_ext_id.astype("category").cat.codes
X['U'] = X.u_id.astype("category").cat.codes
X.i_cnt.apply(lambda x: 1 if x > 0 else 0)
X['T'] = X.i_cnt.astype(float)
X.head()
R_iu = sparse.csr_matrix((X['T'], (X.I, X.U)))
R_ui = sparse.csr_matrix((X['T'], (X.U, X.I)))
item_lookup = X[['i_ext_id', 'I']].drop_duplicates()
item_lookup['i_ext_id'] = item_lookup.i_ext_id.astype(str)
user_lookup = X[['u_id', 'U']].drop_duplicates()
user_lookup['u_id'] = user_lookup.u_id.astype(str)

sparse.save_npz("../assets/sparse/R_ui.npz", R_ui)
sparse.save_npz("../assets/sparse/R_iu.npz", R_iu)
user_lookup.to_csv("../assets/csv/user_lookup.csv")

```

```
item_lookup.to_csv('../././assets/csv/item_lookup.csv')
```

Лістинг файлу analyze_data_als.py:

```
import implicit
import pandas as pd
import numpy as np
import scipy.sparse as sparse
import implicit.evaluation as ev
from implicit.als import AlternatingLeastSquares
import matplotlib.pyplot as plt
import matplotlib.pylab as pylab
import seaborn as s
from sklearn import metrics
import random
plt.style.use('seaborn')
k=4
R_ui = sparse.load_npz(file='../././assets/sparse/R_ui.npz')
R_iu = sparse.load_npz(file='../././assets/sparse/R_iu.npz')

I_lookup = pd.read_csv('../././assets/csv/item_lookup.csv')
U_lookup = pd.read_csv('../././assets/csv/user_lookup.csv')

train, test = ev.train_test_split((R_ui).tocoo().astype('double'))
train_iu, test_iu = ev.train_test_split(R_iu.tocoo().astype('double'))
print(train.shape, test.shape)
print(train_iu.shape, test_iu.shape)

fs = [5,10,15,20,25,30,35,40,45,50]
alpha = 80
k=4
p_iter_sens_0 = []
p_iter_sens_1 = []

for i,v in enumerate(fs):
    model = AlternatingLeastSquares(iterations = v,num_threads=16)
    model.fit((train_iu*alpha).astype('double'))
    p = ev.precision_at_k(model, train, test, K=k, num_threads=16)
    p1 = ev.mean_average_precision_at_k(model, train, test, K=k, num_threads=16)
    p_iter_sens_0.append(p)
    p_iter_sens_1.append(p1)

plt.style.use('seaborn')
q = plt.figure(figsize=(15,8))

fs = [5,10,15,20,25,30,35,40,45,50]

ax = plt.axes()
ax.plot(fs,p_iter_sens_0,marker='.',
        markersize=15,
        linewidth=3)
```

```

ax.set(xlabel= "Кількість ітерацій", ylabel="значення Precision@k")#, title="Залежність
Precision@k від кількості ітерацій")
plt.show()
print(fs[p_iter_sens_0.index(np.max(p_iter_sens_0))])

q = plt.figure(figsize=(15,8))

ax = plt.axes()
ax.plot(fs,p_iter_sens_1,marker='.',
        markersize=15,
        linewidth=3)
ax.set(xlabel= "Кількість ітерацій", ylabel="значення MeanAveragePrecision@k")#,
title="Залежність Precision@k від кількості ітерацій")
plt.show()

fs = [0.01,0.02,0.03,0.04,0.05,0.06,0.07,0.08,0.09,0.1]
alpha = 80
p_reg_sens_0 = []
p_reg_sens_1 = []

for i,v in enumerate(fs):
    model = AlternatingLeastSquares(regularization=v, iterations=10, num_threads=16)
    model.fit((train_iu*alpha).astype('double'))
    p = ev.precision_at_k(model, train, test, K=k, num_threads=16)
    p1 = ev.mean_average_precision_at_k(model, train, test, K=k, num_threads=16)
    p_reg_sens_0.append(p)
    p_reg_sens_1.append(p1)

fs = [0.01,0.02,0.03,0.04,0.05,0.06,0.07,0.08,0.09,0.1]
q = plt.figure(figsize=(15,8))
ax = plt.axes()
ax.plot(fs,p_reg_sens_0,marker='.',
        markersize=15,
        linewidth=3)
ax.set(xlabel= "Значення показника регуляризації", ylabel="значення Precision@k")#,
title="Залежність Precision@k від кількості ітерацій")
plt.show()
print(fs[p_reg_sens_0.index(np.max(p_reg_sens_0))])

q = plt.figure(figsize=(15,8))

ax = plt.axes()
ax.plot(fs,p_reg_sens_1,marker='.',
        markersize=15,
        linewidth=3)
ax.set(xlabel= "Значення показника регуляризації", ylabel="значення
MeanAveragePrecision@k")#, title="Залежність Precision@k від кількості ітерацій")
plt.show()

fs = [100,200,300,400,500,600,700,800,900,1000,1100,1200,1300,1400,1500,1600]

p_factors_sens_0 = []

```

```

p_factors_sens_1 = []
for i,v in enumerate(fs):
    model = AlternatingLeastSquares(iterations = 10, regularization=0.09, factors=v,num_threads=16)
    model.fit((train_iu*alpha).astype('double'))
    p = ev.precision_at_k(model, train, test, K=k, num_threads=16)
    p1 = ev.mean_average_precision_at_k(model, train, test, K=k, num_threads=16)
    p_factors_sens_0.append(p)
    p_factors_sens_1.append(p1)

fs = [100,200,300,400,500,600,700,800,900,1000,1100,1200,1300,1400,1500,1600]
q = plt.figure(figsize=(15,8))
ax = plt.axes()
ax.plot(fs,p_factors_sens_0,marker='.',
        markersize=15,
        linewidth=3)
ax.set(xlabel= "Значення показника кількості прихованих факторів", ylabel="значення Precision@k")#, title="Залежність Precision@k від кількості ітерацій")
plt.show()
print(fs[p_reg_sens_0.index(np.max(p_reg_sens_0))])

q = plt.figure(figsize=(15,8))

ax = plt.axes()
ax.plot(fs,p_factors_sens_1,marker='.',
        markersize=15,
        linewidth=3)
ax.set(xlabel= "Значення показника кількості прихованих факторів", ylabel="значення MeanAveragePrecision@k")#, title="Залежність Precision@k від кількості ітерацій")
plt.show()

```

Лістинг файлу k-means_clust.py:

```

import pandas as pd
import numpy as np
import datetime as dt
import matplotlib.pyplot as plt
import matplotlib.pyplot as pylab
from sklearn.cluster import KMeans
from itertools import cycle
from sklearn import preprocessing
from mpl_toolkits.mplot3d import Axes3D
import pymysql
from sklearn.externals import joblib
import dateutil.parser as parser
import sys
sys.path.insert(0, './')
df0 = pd.read_csv('research_data.csv')
df0['Age'] = df0.Age.apply(lambda x: 0 if x < 0 or x > 94 or x < 16 else x)
df0['Ort'] = df0.Ort.apply(lambda x: str(x).translate ({ord(c): "" for c in
""'0123456789!@#$%^&*()[]{};:./<>?|\`~=-_+"}))
df0['Ort'] = df0.Ort.apply(lambda x: 'unknown' if x == " or x == ' ' or len(str(x)) < 3 else
str(x).strip().lower())

```

```

df0['Land'] = df0.Land.apply(lambda x: str(x).translate ({ord(c): "" for c in
""'0123456789!@#%$%^&*()[]{};,:./<>?|\`~=-_+}))
df0['Land'] = df0.Land.apply(lambda x: 'unknown' if not x or x==" " or x == ' ' else
str(x).strip().lower())

olb = preprocessing.LabelEncoder()
llb = preprocessing.LabelEncoder()
ort = pd.unique(df0.Ort)
land = pd.unique(df0.Land)
ort_t = []
for i,v in enumerate(ort):
    ort_t.append(str(v))
ort_lbencoder = olb.fit(ort_t)
ort
df0.Ort = ort_lbencoder.transform(df0.Ort)

land_lbencoder = llb.fit(land)
df0.Land = land_lbencoder.transform(df0.Land)
X = df0[['Geschlecht_ID', 'Age','Ort','Land', 'Artikelnummer']]
train = X[['Age','Geschlecht_ID','Ort', 'Land']].values
train
np.random.seed(5)

estimators = [('kmeans_df_s24_sales_clean_grouped_7', KMeans(n_clusters=7,
random_state=52)),
('kmeans_df_s24_sales_clean_grouped_8', KMeans(n_clusters=8, random_state=52))]
labels = []
for name, est in estimators:
    est.fit(train)
    labels.append((est, name))
fignum = 1
titles = ['7 clusters', '8 clusters']
for labels_est, name in labels:
    fig = plt.figure(fignum, figsize=(13, 6))
    ax = Axes3D(fig, rect=[0, 0, .95, 1], elev=48, azim=150)
    lab = labels_est.labels_

    ax.scatter(X.Age, X.Geschlecht_ID, X.Ort, c=lab.astype(np.float), edgecolor='k')

# ax.w_xaxis.set_ticklabels([])
# ax.w_yaxis.set_ticklabels([])
# ax.w_zaxis.set_ticklabels([])
ax.set_xlabel('Age')
ax.set_ylabel('Geschlecht_ID')
ax.set_zlabel('Ort')
ax.set_title(titles[fignum - 1])
ax.dist = 12
fignum = fignum + 1
km_8 = labels[-1][0]
km_8
r_d_final_resultset = df0[['Age', 'Geschlecht_ID', 'Ort', 'Land', 'Artikelnummer', 'Anzahl',
'Produkt_Preis']]

```



```

r_d_final_resultset['Cluster_ID'] = km_8.predict(r_d_final_resultset[['Age','Geschlecht_ID', 'Ort',
'Land']])
r_d_final_resultset.head()
gr = r_d_final_resultset[['Cluster_ID', 'Anzahl']].groupby(['Cluster_ID']).sum()
gr.to_csv('gr.csv')
gr = pd.read_csv('gr.csv')
gr.head()
llabels = ['Cluster 0', 'Cluster 1', 'Cluster 2','Cluster 3','Cluster 4','Cluster 5','Cluster 6','Cluster 7']
sizes = gr.Anzahl.values
colors = ['green', 'red', 'yellow', 'cyan', 'gold', 'yellowgreen', 'lightcoral', 'lightskyblue']
explode = (0.0, 0.2, 0, 0.3, 0, 0, 0.2, 0) # explode slices

# Plot
plt.figure(figsize=(15,10))
plt.pie(sizes, explode=explode, labels=llabels, colors=colors,
        autopct='%1.1f%%', shadow=True, startangle=140)

plt.axis('equal')
plt.show()
plt.scatter(gr.Cluster_ID, gr.Anzahl)
r_d_final_resultset.where(r_d_final_resultset.Cluster_ID == 1).describe()
cluster_info = pd.DataFrame()
for i in r_d_final_resultset.Cluster_ID.unique():
    print(i)
    r = r_d_final_resultset[(r_d_final_resultset.Cluster_ID == i) & (r_d_final_resultset.Age > 0)]
    #print(r)
    cluster_info = cluster_info.append(pd.DataFrame([[i, r.Age.min(), r.Age.max(), r.Ort.min(),
r.Ort.max(), len(r.values)], columns=['Cluster ID','Age Min', 'Age Max', 'Ort Min','Ort Max',
'Count']))
grouped_set = r_d_final_resultset[['Cluster_ID','Artikelnummer',
'Anzahl']].groupby(['Cluster_ID','Artikelnummer']).sum()
grouped_set.to_csv('grouped_set.csv')
grouped_set = pd.read_csv('grouped_set.csv')
grouped_set.head()
clust_artikels = {}
for v in pd.unique(r_d_final_resultset.Cluster_ID):
    print("working on cluster id: ", v)
    clust_artikels[v] = grouped_set.where(grouped_set.Cluster_ID ==
v)[['Artikelnummer','Anzahl']].dropna()

clust_artikels[2].head()
filename = './s24_clust_km8_v-0.0.5.ml'
clusterAnalysis = CustomerGoodsSuggestionClassifier()
clusterAnalysis.init(km_8, clust_artikels, land_lbencoder, ort_lbencoder)

#print(clusterAnalysis._ortEncoder.classes_)
#print(clusterAnalysis._landEncoder.classes_)
clusterAnalysis.predict('Lietzow', 1, 'DE', '1968-07-15 00:00:00')
#pd.DataFrame.to_json(clust_artikels[2])
joblib.dump(clusterAnalysis, filename)
a = joblib.load(filename)

```

```
a.predict('Lietzow', 1, 'DE', '1968-07-15 00:00:00')
a.get_products_by_cluster_id(clust=[6])[0].Artikelnummer
```

Лістинг файлу flask.py:

```
import flask
from flask import request, jsonify, json
from sklearn.externals import joblib as srzr
from vst.classifiers import CustomerGoodsSuggestionClassifier
from vst.models import ProductItem
app = flask.Flask(__name__)

@app.route('/', methods=['GET'])
def api_test():
    return jsonify({"Title": "Willkommen!", "Message": "This is a prototype for artcom venture
Machine Learning assets", "API":
"/api/v1/customer/cluster/<int:Company_Id>/<Geburtsdatum>/<int:Geschlect_Id>/<Ort>/<Land>"
})

@app.route('/api/v1/customer/cluster/<int:Company_Id>/<Geburtsdatum>/<int:Geschlect_Id>/<Ort
>/<Land>', methods=['GET'])
def api_predict_cluster_by_customerparams(Company_Id, Geburtsdatum, Geschlect_Id, Ort,
Land):
    clustering_classifier = srzr.load('ml_models/s24_clust_km8_v-0.0.5.ml')
    cluster = clustering_classifier.predict(
    Ort, Geschlect_Id, Land, Geburtsdatum)
    response_model = clustering_classifier.get_products_by_cluster_id(cluster)
    JSON = json.dumps([ob.__dict__ for ob in response_model])
    return app.response_class(
    response=JSON,
    status=200,
    mimetype='application/json'
    )

if __name__ == '__main__':
    app.run()
```

Лістинг файлу CustomerGoodsSuggestionClassifier.py:

```
import vst.models as vst_models
from sklearn.cluster import KMeans
from sklearn.externals import joblib
import sklearn.preprocessing as preprocessing
import dateutil.parser as parser
import pandas as pd
import datetime as dt

dt_0 = dt.datetime(2018, 7, 3, 0, 0, 0, 0)

class CustomerGoodsSuggestionClassifier:
    _ortEncoder = preprocessing.LabelEncoder()
    _landEncoder = preprocessing.LabelEncoder()
    _classifier = KMeans()
```

```

_specData = pd.DataFrame()
_model_bday_terminator = dt_0

def init(self, model, articles_by_cluster, land_lb_encoder, ort_lb_encoder):
    self._classifier = model
    self._specData = articles_by_cluster
    self._landEncoder = land_lb_encoder
    self._ortEncoder = ort_lb_encoder

def get_products_by_cluster_id(self, clust):
    data = self._specData[clust[0]].sort_values(
        by='Anzahl', ascending=False)
    objList = list()

    for i in data.values[:300]:
        item = vst_models.ProductItem()
        item.Artikelnummer = i[0]
        item.Anzahl = i[1]
        objList.append(item)
    return objList

def predict(self, city, gender, land, birthday):
    c_land = land
    if(not land):
        c_land = 'unknown'

    if(str(land) == '' or len(str(land)) < 2 or str(land) == ' ' or not land):
        c_land = 'unknown'

    if(str(city) == '' or len(str(city)) < 3 or str(city) == ' ' or not city):
        city = 'unknown'

    bd_0 = parser.parse(birthday)
    if(not birthday):
        bd_0 = self._model_bday_terminator
    if(birthday == ''):
        bd_0 = self._model_bday_terminator

    c_Age = self._model_bday_terminator.year - bd_0.year - \
        ((self._model_bday_terminator.month,
         self._model_bday_terminator.day) < (bd_0.month, bd_0.day))
    if(c_Age > 94 or c_Age == 0):
        c_Age = 0
    c_land = self._landEncoder.transform([c_land.strip().lower()])[0]
    c_city = self._ortEncoder.transform([city.strip().lower()])[0]
    return self._classifier.predict([[c_Age, gender, c_city, c_land]])

```

Лістинг файлу ProductItem.py:

```

class ProductItem:
    Artikelnummer = 0
    Anzahl = 0

```

ДОДАТОК Г

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Кузнєцова Н. В., Бідюк П. І. Інформаційна технологія аналізу фінансових даних на основі інтегрованого методу. *Системні дослідження та інформаційні технології*. 2011. № 1. С. 22–33. (Входить до наукометричних баз Index Copernicus, DOAJ, ВІНІТІ, РІНЦ).
2. Бідюк П. І., Кузнєцова Н. В., Терентьєв О. М. Система підтримки прийняття рішень для аналізу фінансових даних. *Наукові вісті НТУУ “КПІ”*. 2011. №1. С. 48–61. (Входить до наукометричних баз Index Copernicus, DOAJ, EBSCO, WorldCat, J-Gate та ін.).
3. Bidyuk P. I., Konovalyuk M. M., Kuznetsova N.V., Pudlo I. V. Adaptive Short-Term Forecasting of Selected Financial Processes. *Research bulletin of NTUU “KPI”*. 2014. N1. P.35–41. (Входить до наукометричних баз Index Copernicus, DOAJ, EBSCO, WorldCat, J-Gate та ін.).
4. Кузнєцова Н. В. Деякі аспекти мінімізації інформаційних ризиків у банківській діяльності. *Системні дослідження та інформаційні технології*. 2014. № 1. С. 7–19. (Входить до наукометричних баз Index Copernicus, DOAJ, ВІНІТІ, РІНЦ).
5. Бідюк П. І., Кузнєцова Н. В. Прогнозирование волатильности финансовых процессов с помощью моделей условной дисперсии. *Проблемы управления и информатики*. 2014. № 5. С. 47–54. (Входить до наукометричних баз **Scopus**, ISI Thomson Reuters, Mathematical Review, Applied Mechanics Reviews, ВІНІТІ, РІНЦ).
6. Кузнєцова Н. В. Практичні підходи до визначення та урахування невизначеностей, що формують фінансові ризики. *Тр. Одес. политехн. ун-та*. 2014. Вып.2 (44). С. 160–170. (Входить до наукометричних баз Index Copernicus, DOAJ, EBSCO, WorldCat, РІНЦ та ін.).
7. Кузнєцова Н.В., Бідюк П. І. Нейронні та мережі Байєса у задачі аналізу кредитних ризиків. *Реєстрація, зберігання і обробка даних*. 2015. Т.17, №2. С. 61–71. (Входить до наукометричних баз Index Copernicus, ICI World of Journals).
8. Kuznietsova N. V., Bidyuk P. I. Business Intelligence Techniques For Missing Data Imputation. *Research bulletin of NTUU “KPI”*. 2015. N5. P. 47–56. (Входить до наукометричних баз Index Copernicus, DOAJ, EBSCO, WorldCat, J-Gate та ін.).
9. Кузнєцова Н. В. Виявлення та оброблення невизначеностей у формі неповних даних методами інтелектуального аналізу. *Системні дослідження та інформаційні технології*. 2016. № 2. С. 104–115. (Входить до наукометричних баз Index Copernicus, DOAJ, ВІНІТІ, РІНЦ).
10. Kuznietsova N. V. Scoring technology for risk assessment of fraud in banking *CEUR Workshop Proceeding (ISSN 1613-0073)*. 2016. Vol. 1813. P.54-61 [Online]. Available: <http://ceur-ws.org/Vol-1813/paper8.pdf> (last accessed 25.09.2018). (Входить до наукометричних баз **Scopus**, DBLP).

11. Кузнєцова Н.В., Бідюк П. І. Моделювання фінансового ризику в телекомунікаційній сфері. *Наукові вісті НТУУ "КПІ"*. 2017. №5. С. 51–58. (Входить до наукометричних баз Index Copernicus, DOAJ, EBSCO, WorldCat, J-Gate та ін.).
12. Кузнєцова Н. В., Бідюк П. И. Моделирование кредитных рисков на основе теории выживания. *Проблемы управления и информатики*. 2017. №6. С. 33– 46. (Входить до наукометричних баз **Scopus**, ISI Thomson Reuters, Mathematical Review, Applied Mechanics Reviews, ВІНІТІ, РІНЦ та ін.).
13. Kuznietsova N. V. Information Technologies for Clients' Database Analysis and Behaviour Forecasting. *CEUR Workshop Proceeding (ISSN 1613-0073)*. 2017. Vol. 2067. P.56-62 [Online]. Available: <http://ceur-ws.org/Vol-2067/> (last accessed 25.09.2018). (Входить до наукометричних баз **Scopus**, DBLP).
14. Кузнєцова Н. В. Розробка скорингових карт для аналізу ризиків банківської діяльності. *Реєстрація, зберігання і обробка даних*. 2017. Т.19. №4.С.35–44. (Входить до наукометричних баз Index Copernicus, ICI World of Journals).
15. Кузнєцова Н. В., Бідюк П. І. Динамічне моделювання фінансових ризиків. *Індуктивне моделювання складних систем*. 2017. Вип. 9. 2017. С. 122–137.
16. Кузнєцова Н. В. Інформаційні технології для аналізу фінансових зловживань на платформі ProZorro. *Наукові праці ВНТУ: електронне наукове фахове видання*. Електронні дані. Вінниця: Вінниц. Нац. тех. ун-т: 2018. №1. URL: <https://praci.vntu.edu.ua/index.php/praci/article/view/534/525>. (дата звернення 25.09.2018). (Входить до наукометричної бази РІНЦ).
17. Бідюк П. І., Кузнєцова Н. В. Ймовірно-статистичний метод оцінювання ризику фінансових втрат. *KPI Science News*. 2018. №2. С. 7–17. (Входить до наукометричних баз Index Copernicus, DOAJ, EBSCO, WorldCat, J-Gate та ін.).
18. Кузнєцова Н. В. Фінансовий ризик-менеджмент з урахуванням інформаційних ризиків. *Реєстрація, зберігання і обробка даних*. 2018. Т.20. №1. С. 30–39. (Входить до наукометричних баз Index Copernicus, ICI World of Journals).
19. Кузнєцова Н. В. Аналіз та прогнозування ризиків шахрайства з кредитними картками. *Інформатика та математичні методи в моделюванні*. 2018. Т.8, №1. С. 16–25. (Входить до наукометричних баз Index Copernicus, РІНЦ).
20. Кузнєцова Н.В., Бідюк П. І. Система підтримки прийняття рішень для аналізу інвестиційних ризиків фінансових ринків. *Вісник Національного університету «Львівська політехніка». Серія «Інформаційні системи та мережі»*. 2018. Вип. 887. С. 115–121.
21. Кузнєцова Н.В., Бідюк П.І. Структурно-параметрична адаптація ймовірно-статистичних моделей для оцінювання фінансових ризиків. *KPI Science News*. 2018. №3. С.23–34. (Входить до наукометричних баз Index Copernicus, DOAJ, EBSCO, WorldCat, J-Gate та ін.).

22. Кузнєцова Н. В., Бідюк П. І. Системний підхід до менеджменту фінансових ризиків. *Системні дослідження та інформаційні технології*. 2018. № 2. С. 124–140. (Входить до наукометричних баз Index Copernicus, DOAJ, ВІНІТІ, РІНЦ).
23. Кузнєцова Н. В., Бідюк П. І. Выявление рисков мошенничества в ходе тендерных закупок методами теории выживания. *Управляющие системы и машины*. 2018. № 2 (274). С.87–96. (Входить до наукометричної бази РІНЦ).
24. Кузнєцова Н. В., Бідюк П. І. Аналіз фінансових ризиків з використанням SAS-технологій обробки даних. *Електротехнічні і комп'ютерні системи*. 2016. № 22(98). С. 267–271. (Входить до наукометричних баз Index Copernicus, Ulrich's Periodicals Directory, РІНЦ, ВІНІТІ).
25. Кузнєцова Н. В. Практичне застосування методології розробки скорингових карт для аналізу ризиків автокредитування. *Електротехнічні та комп'ютерні системи*. 2017. № 24(100). С. 104 – 111. (Входить до наукометричних баз Index Copernicus, Ulrich's Periodicals Directory, РІНЦ, ВІНІТІ).
26. Кузнєцова Н. В. Динамічно-рівномірна дискретизація даних у гібридних мережах Байєса. *Вісник Хмельницького Національного Університету*. 2011. №5. Т.3(180). С. 247–251. (Входить до наукометричних баз Index Copernicus, РІНЦ, Polish Scholarly Bibliography).
27. Кузнєцова Н. В. Методика оцінювання ризику зниження фінансової стабільності за допомогою мереж Байєса. *Вісник Хмельницького Національного Університету*. 2013. №2. Т.3 (198). С. 187–190. (Входить до наукометричних баз Index Copernicus, РІНЦ, Polish Scholarly Bibliography).
28. Kuznietsova N. Requirements management. *Summer School IT-SEA 5-20 July'14. Report*. Odessa National Polytechnic University. Odessa, 2014. P. 98–121.
29. Kuznietsova N. Project management. Student's hackathon for mobile application *Summer School IT-SEA 5–20 July'14. Report*. Odessa National Polytechnic University. Odessa. 2014. P. 122–130.
30. Кузнєцова Н. В. Інформаційні технології обробки та аналізу даних у фінансовому ризик-менеджменті. *Інформаційні технології та спеціальна безпека*. 2015. №1. С. 86–98.
31. Кузнєцова Н. В. Скорингові карти для аналізу ризиків банківської діяльності. *Інформаційні технології та спеціальна безпека*. 2018. №1. С. 26–35.
32. Белевець М. С., Бідюк П. І., Кузнєцова Н. В. Розробка системного підходу до менеджменту ризиків. *Системні науки та кібернетика*. 2015. №1. С. 30 – 48. URL: http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc_1_2015.pdf.
33. Гуськова В. Г., Кузнєцова Н. В. Комбінування оцінок прогнозів, обчислених за різними методами для обраних країн світу. *Системні науки та кібернетика*. 2015. №1. С. 49 – 57. URL: http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc_1_2015.pdf.
34. Кузнєцова Н. В., Кінда В. В. Обробка і аналіз даних обрахунку тепловтрат приміщень у реальному часі. *Системні науки та кібернетика*. 2015.

- №1. С. 77 – 86. URL:
http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc_1_2015.pdf.
35. Фомін О. В., Кузнецова Н. В. Скорингові моделі поведінки клієнтів-власників кредитних карток для оцінки їх платоспроможності. *Системні науки та кібернетика*. 2016. № 5. С. 56 – 67. URL:
http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc_5_2016.pdf.
36. Кузнецова Н. В., Куца К.В., Штогрін С. Р. Застосування методології аналізу виживання для дослідження споживчих ризиків. *Системні науки та кібернетика*. 2017. №6. С. 126–135. URL:
http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc_6_2017.pdf.
37. Kuznyetsova N. Informational technologies for financial data analysis. *Proc. IV Int. Scientific Conf. "Intelligence, Integration, Reliability"*. Kyiv-Warsaw, April 21–22, 2011). К.: NTUU "KPI". 2011. Р. 32–33.
38. Кузнецова Н. В. Інтегровані моделі аналізу фінансових даних. *Системний аналіз та інформаційні технології (SAIT 2011)*: матеріали XIII міжнар. наук.-техн. конф. (Київ 23–28 травня, 2011 р.). К.: ННК "ІПСА" НТУУ "КПІ", 2011. С. 274.
39. Кузнецова Н. Аналіз даних клієнтів за допомогою мереж Байєса. *Комп'ютерні науки та інженерія*: мат. V міжнар. конф. молодих вчених CSE-2011, (Львів, 24–26 листоп. 2011 р.). Вид. Львівської політехніки, 2011. С.50–53.
40. Кузнецова Н. В. Аналіз і оцінювання інформаційних ризиків. *Системний аналіз та інформаційні технології (SAIT 2013)*: матеріали XV міжнар. наук.-техн. конф. (Київ, 27–31 травня, 2013 р.). К.: ННК "ІПСА" НТУУ "КПІ", 2013. С. 122–123.
41. Кузнецова Н. В. Застосування мереж Байєса до оцінювання інформаційних ризиків. *Информационные технологии и безопасность. Оценка состояния*: мат. междун. науч.-практ. конф. ИТБ-2013. (Киев, 18 июня 2013г.). Вып. 13. К.: НАН України. 2013. С. 95–104.
42. Кузнецова Н. В., Бідюк П. І. Інтегрований підхід до аналізу фінансових ризиків. *Сучасна інформатика: проблеми, досягнення та перспективи розвитку*: матеріали міжнар. наук. конф. (12–13 вересня 2013 року, Київ). Київ: Ін-т кібернетики ім. В.М.Глушкова НАН України, 2013. С. 153–155.
43. Kuznietsova N.V., Bidiyuk P.I. Systemic approach to estimation of financial risks. *Науково-технічна конференція «Інформатика, математика, автоматика. ІМА:2015* (Суми, 20–25 квітня 2015 року. Сумський державний університет). 2015. С. 46–47.
44. Кузнецова Н. В. Питання якості, конфіденційності та коректності даних в інформаційних технологіях аналізу фінансових ризиків. *Информационные технологии и безопасность*: мат. междун. науч.-практ. конф. ИТБ-2015 (Киев, 21 октября 2015 г.). Вып. 15. К.: ИПРИ НАН Украины, 2015. С. 131–136.
45. Bidiyuk P. I., Kuznietsova N.V. Decision support system for adaptive processes forecasting in various applied areas with alternative techniques.

Геоинформационные системы и компьютерные технологии эколого-экономического мониторинга: сб. докл. междунар. научно-техн. конф. (г. Днепропетровск, 13–15 апреля 2016 г.) / под ред. Л. В. Сарычевой: Электрон. данные. Днепропетровск: ГБУЗ «НГУ» МОН Украины, 2016. URL: http://gis.dp.ua/conf2016-publications/sections/iad/1_Bidyuk_Kuznetsova.pdf.

46. Кузнецова Н. В. Скорингові карти як інструмент аналізу ризиків. *Системний аналіз та інформаційні технології (SAIT 2016)*: матеріали 18-ї міжнар. наук.-техн. конф. (Київ, 25 – 29 травня 2016 р.). К.: ННК «ІПСА» НТУУ «КПІ», 2016. С. 106–107.

47. Кузнецова Н. В. Скорингові технології оцінювання ризиків шахрайства в банківській діяльності. *Информационные технологии и безопасность*: мат. междунар. науч.-практ. конф. ИТБ-2016. (Киев, 1 декабря 2016 г.) К.: ИПРИ НАН Украины, 2016. С. 43 – 47.

48. Кузнецова Н. В., Бідюк П. І. Математичні моделі виживання для прогнозування фінансових ризиків. *Моделирование та прогнозування економічних процесів*: матеріали XI науково-практичної конференції. (Київ, 6–8 грудня 2017р.). К.: КПІ ім. Ігоря Сікорського, 2017. С. 47–48.

49. Кузнецова Н. В. Інформаційні технології аналізу клієнтської бази абонентів та прогнозування їх поведінки. *Информационные технологии и безопасность*: мат. междунар. науч.-практ. конф. ИТБ-2017 (Киев, 30 ноября, 2017 г.). К.: ИПРИ НАН Украины, 2017. С. 114 – 120.

50. Кузнецова Н. В., Фомін О. В. Прогнозування ризику втрати користувачів онлайн-платформи. *Системний аналіз та інформаційні технології (SAIT 2018)*: матеріали 20-ї Міжн. науково-техн. конф., (Київ 21–24 травня, 2018 р.). К.: ННК «ІПСА» НТУУ «КПІ», 2018. №20. С. 139–140.

51. Kuznetsova N., Bidyuk P. Forecasting of Financial Risk Users' Outflow. *IEEE First International Conference on System Analysis & Intelligent Computing (SAIC)*, (Kyiv: 08–12 October). 2018. P.250–255. DOI: <https://ieeexplore.ieee.org/abstract/document/8516782>. (Входить до бази **Scopus**).

52. Kuznetsova N., Seebauer M., Zabielin S. Some Methods for Estimating Financial Risks in Banking. *IEEE First International Conference on System Analysis & Intelligent Computing (SAIC)*, (Kyiv: 08–12 October, 2018) 2018. P.271–274. DOI: <https://ieeexplore.ieee.org/document/8516873>. (Входить до бази **Scopus**).

53. Бідюк П. І., Коршевнюк Л. О., Кузнецова Н. В. Моделі і методи прикладної статистики: **навч. посіб.** з грифом МОН України. Київ: НТУУ «КПІ», 2014. 722 с.

ДОДАТОК Д АКТИ ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОЇ РОБОТИ

Результати дисертаційної роботи використані та впроваджені у:

- Національному Банку України при розробці методології оцінювання та аудиту комерційних банків;
- компанії ТОВ «САС Інстїтьют ЕЛ.ЕЛ.СІ.» при розробці нових інформаційних технологій та інструментальних рішень для банківської сфери з можливістю біхевіоріального скорингу та динамічного прогнозування фінансових ризиків;
- компанії ТОВ «ЕЛСІКО» при аналізі тендерних закупівель та поведінки учасників торгів на онлайн-платформі ProZorro;
- телекомунікаційній компанії ТОВ «Лайфселл» при прогнозуванні відтоку абонентів та можливих втрат, а також прогнозуванні витрат абонентів при користування мобільними послугами за кордоном, що дало можливість зменшити втрати компанії на 12%;
- іноземній компанії Artcom Venture GmbH при аналізі поведінки клієнтів, що дозволило на 7% підвищити якість прогнозів та збільшити обсяги продажів;
- у навчальному процесі кафедри математичних методів системного аналізу ІПСА НТУУ «КПІ ім. Ігоря Сікорського» при викладанні навчальних курсів «Системний аналіз та прогнозування ризиків», «Прикладні методи прогнозування», «Системний аналіз фінансових ризиків методами штучного інтелекту», «Управління ризиками на фінансових ринках», «Ризик-менеджмент методами інтелектуального аналізу даних», «Системи і методи підтримки прийняття рішень»,

що підтверджується відповідними актами, відгуками та довідками про впровадження.

Відгук
про застосування результатів
докторської дисертаційної роботи
Кузнєцової Наталії Володимирівни

Відповідно до положень Конституції України і законів України "Про Національний банк України" і "Про банки і банківську діяльність" Національний банк України (НБУ) здійснює регулювання та банківський нагляд для забезпечення стабільності банківської системи і захисту інтересів вкладників і кредиторів банків. Під час оцінювання фінансової діяльності банків враховуються як інформація про їх діяльність так і результати перевірки та звіти співробітників НБУ щодо обсягів капіталів, стабільності портфелю вкладників та кредиторів банку.

Впроваджуючи міжнародні практики Базеля III з урахуванням українських реалій, НБУ напрацьовує методичні рекомендації щодо організації та функціонування систем ризик-менеджменту в Україні. Запропонована Кузнєцовою Н.В. методологія системного аналізу фінансових ризиків в умовах наявності невизначеностей, характерних для процесів моделювання, прогнозування і оцінювання можливих втрат, з урахуванням динамічного оцінювання та прогнозування фінансових ризиків, яке відрізняється урахуванням часу та оцінюванням моменту настання переходу на вищий ступінь ризику і забезпечує можливість прогнозування критичного часу є оригінальною розробкою, що може бути використана для більш ефективної організації управлінської діяльності комерційних банків та роботи департаменту банківського нагляду НБУ.

Урахування інформаційної складової в моделях оцінювання та менеджменту фінансового ризику, запропоноване у дисертаційній роботі Кузнєцовою Н.В., дозволяє отримати підвищення якості оцінок можливих втрат, є корисним для практичного застосування під час перевірки і моніторингу банку та завчасного введення в комерційні банки тимчасового адміністратора НБУ, налагодження роботи банку, а отже, зменшення ризику неплатоспроможності банку.

«21» 05 2018р.

Заступник начальника відділу аналізу
 Департаменту банківського нагляду
 Національного банку України

В.В. Шульженко
 Начальник управління із
 запровадження та адміністрування
 персоналу Департаменту персоналу

І.В. Волошин

Сідрис Валентина
засвідчує



01601, Київ,
вул. Шовковична, 42-44
Тел.: +380 (44) 459-03-55
Факс: +380 (50) 385-52-65

Акт

впровадження результатів дисертаційного дослідження Кузнєцової Наталії Володимирівни

У дисертаційному дослідженні Кузнєцової Наталії Володимирівни представлена розробка базових модулів комплексної інформаційної системи підтримки прийняття рішень (ІСППР) на основі запропонованих нових динамічних моделей і методів, що дозволяє підвищити ефективність обробки фінансових даних різних типів, провести аналіз фінансових та інформаційних ризиків, зокрема ризиків кредитування, інвестування, операційних ризиків. Розроблена Кузнєцовою Н.В. ІСППР може використовуватись як окремими модулями для підтримки роботи депозитного, кредитного та аналітичного відділів у відділенні банку, так і в цілому в якості централізованої аналітичної системи фінансової установи або компанії; запропонована ІСППР також може бути використана для розв'язання завдань побудови швидкої аналітичної оцінки нових видів ризиків.

Результати дисертаційного дослідження використовувались в учбовому процесі глобальної академічної програми SAS в Україні в рамках циклу курсів SAS Risk Management Group «Менеджмент ризиків з використанням SAS технологій».

Основні впроваджені результати:

- розроблені власні коди програм на SAS Base;
- розроблені нові моделі обробки неповних даних на основі власного комбінованого ймовірнісно-регресійного методу для Impute-вузла в SAS Enterprise Miner;
- запропоновано динамічний підхід оцінювання ризиків різних типів, який легко інтегрується в існуючу систему;
- розроблено метод оцінювання втрат на основі динамічного підходу та моделей аналізу виживання.

Даний акт не є документом для фінансових розрахунків.

Директор ТОВ «САС Інстїтют Е.Е.Е.СІ.»



Баркалов

ТОВ «ЕЛСІКО»

Вих. № 1506
Від «13» червня 2018р.

Акт

впровадження результатів докторської дисертаційної роботи Кузнєцової Наталії Володимирівни

Результати дисертаційної роботи Кузнєцової Н.В. у вигляді біхевіоріальних непараметричних моделей, що характеризують поведінку компаній-учасників торгів, були використані компанією «ЕЛСІКО» для проведення порівняльного аналізу дій конкурентів при їх участі в торгах на онлайн-платформі Prozorro. Результати аналізу були використані для розробки власної стратегії поведінки під час участі в торгах.

Ще однією сферою впровадження результатів роботи став аналіз фінансової діяльності компанії «ЕЛСІКО» на основі динамічного оцінювання та прогнозування фінансових ризиків та методу структурно-параметричної адаптації ймовірно-статистичних моделей оцінювання ризиків з метою розробки пакету пропозицій для аналізу інвестиційних ризиків роздрібною торгівлі (за групами товарів) на українському ринку і залучення нових іноземних інвесторів для розширення власного бізнесу.

Даний акт не є документом для фінансових розрахунків.

Комерційний директор ТОВ «ЕЛСІКО»

Денисюк Д.М.





Товариство з обмеженою відповідальністю «лайфселл»
 вул. Солом'янська, 11, літера «А», м. Київ, 03110, Україна
 тел.: +38(044) 233-31-31, факс: +38 (044) 594-40-90
 e-mail: reception@lifecell.com.ua, web: www.lifecell.ua
 Код ЄДРПОУ: 22859846

Від «10» 07 2018р.

Акт

впровадження результатів докторської дисертаційної роботи Кузнецової Наталії Володимирівни

Телекомунікаційна галузь є однією з найтехнологічніших галузей в світі, з постійним розвитком технологій зв'язку, нових стандартів, пристроїв та різноманітних шляхів комунікацій. Особливість роботи операторів стільникового зв'язку пов'язана з необхідністю постійного оновлення інформаційних технологій, розробкою та наданням нових послуг на основі новітніх технологій та, відповідно, необхідністю передбачення не лише уподобань клієнтів та їх подальших дій, а й пропонуванням нових послуг та вигідних умов для користування послугами як в Україні, так і у роумінгу.

Розроблені Кузнецовою Н.В. математичні моделі для виявлення клієнтів, схильних до відтоку, показали високу точність прогнозування і дозволили виявити клієнтів, які лише планують змінювати оператора зв'язку. Було вперше запропоновано метод динамічного оцінювання та прогнозування фінансових ризиків телекомунікаційної компанії на основі поведінкових моделей. Розроблені динамічні моделі для оцінювання можливого критичного моменту часу, коли починається перехід клієнта з допустимого рівня до критичного і клієнт перестає користуватись послугами, а компанія несе фінансові втрати. Це дозволило завчасно виявити групи клієнтів, у яких висока ймовірність відтоку, та оцінити можливі втрати компанії, з урахуванням моменту їх виникнення, завдяки моделям на основі пропорційних ризиків. Для оцінювання доцільності утримання клієнтів Кузнецова Н.В. запропонувала застосування принципу адаптивного менеджменту ризиків із структурно-параметричним налаштуванням ймовірно-статистичних моделей оцінювання фінансових втрат, як функцій часу, для конкретних груп клієнтів. Такий підхід дозволяє комплексно врахувати витрати, пов'язані з відтоком клієнтів, порівняно з витратами на розробку нових пропозицій та надання додаткових безкоштовних послуг.

Кузнецовою Н.В. було запропоновано принцип інтегрованого динамічного урахування факторів часу, ступеня та рівня ризику, що на практиці дозволило оцінювати клієнтів, які потенційно збираються за кордон, прогнозувати їх витрати на мобільні послуги за кордоном, кількість днів, які вони будуть користуватись послугами у роумінгу, обсяг витрачених коштів, та, відповідно, витрати компанії на забезпечення клієнтів такими послугами у роумінгу. Це дозволило зменшити втрати компанії на 12% за рахунок розробки спеціальних пакетів послуг у роумінгу, що є вигідними для компанії за рахунок обсягів використаних послуг зв'язку клієнтами, що перебувають за кордоном.

Начальник відділу
 моніторингу бізнес діяльності
 та гарантування прибутків



Хлівна І.С.

Адреси для листування:

т.: 0 800 20 54 33,
 т.: (044) 233 31 31,
 ф.: (044) 594 40 90,
 www.lifecell.ua

м. Київ, 03110,
 вул. Солом'янська, 11,
 БЦ Ілєєв

м. Харків, 61003,
 вул. Кооперативна,
 6/8

м. Дніпро, 49044,
 вул. Виконкомівська, 7

м. Львів, 79010,
 вул. Ленчківська, 72

м. Одеса, 65044,
 пр-т. Шевченка, 4 Д



Act of Implementation of doctoral work
by Nataliia Kuznietsova

The analysis, assessment and forecasting of individual customer needs is an important aspect of successful product and service management, as it can help increase market share through re-purchases. Applying the methods of data mining intelligence, Nataliia Kuznietsova carried out an extensive analysis of the data of one of our clients, taking into account perspective and stochastic uncertainties related with the need of prediction consumer preferences and actions, behavior of clients, and also developed an algorithm of machine learning that envisaged forecasting and product recommendations, based on customer preferences of previous behavior and customer purchases. Implementing the information system for Shop24Direct to support decision-making has allowed 7% more accurate forecasts for target customers in order to increase customer value and corporate profitability by increasing the order of related and new products from customers.

Акт впровадження результатів докторської дисертаційної роботи
Кузнецової Наталії Володимирівни

Аналіз, оцінювання та прогнозування індивідуальних потреб клієнтів є важливим аспектом успішного управління продуктами та послугами, оскільки це може сприяти збільшенню частки ринку за рахунок повторних покупок. Застосовуючи методи інтелектуального аналізу даних, Кузнецова Наталія здійснила розширений аналіз даних діяльності одного з наших клієнтів, з урахуванням перспективної та стохастичної невизначеностей, пов'язаних з необхідністю прогнозування дій та споживчих уподобань та поведінки клієнтів, а також розробила алгоритм машинного навчання, який передбачав прогнозування і рекомендації товарів, з урахуванням уподобання клієнтів на основі попередньої поведінки і покупок. Впровадження інформаційної системи для компанії Shop24Direct для підтримки прийняття рішень дозволило отримати на 7% точніші прогнози для цільових аудиторій клієнтів з метою підвищення обсягів продажу та корпоративної прибутковості за рахунок збільшення замовлення супутніх та нових товарів від клієнтів.





УКРАЇНА

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

03056, м. Київ, пр-т Перемоги, 37; тел. (+38 044) 204-82-82 тел./факс (+38 044) 204-97-88
<http://www.kpi.ua> e-mail: mail@kpi.ua СДРПОУ 02070921

25.06.2018 р. № 9/4040

Д о в і д к а
про впровадження результатів
докторської дисертаційної роботи
Кузнєцової Наталії Володимирівни

Довідка видана Кузнєцовій Наталії Володимирівні у тому, що результати виконання її дисертаційної роботи, зокрема:

- метод динамічного оцінювання та прогнозування фінансових ризиків;
- метод структурно-параметричної адаптації ймовірно-статистичних моделей;
- методологія системного аналізу фінансових ризиків в умовах наявності невизначеностей, характерних для процесів моделювання, прогнозування і оцінювання можливих втрат;
- принцип інтегрованого динамічного урахування факторів часу, ступеня та рівня ризику у фінансовому ризику-менеджменті;
- інформаційна система аналізу фінансових ризиків на основі оригінальних моделей динамічного оцінювання фінансових ризиків

використовуються на кафедрі математичних методів системного аналізу при викладанні Кузнєцовою Н. В. таких курсів:

1. Системний аналіз та прогнозування ризиків.
2. Прикладні методи прогнозування.
3. Системний аналіз фінансових ризиків методами штучного інтелекту.
4. Управління ризиками на фінансових ринках.
5. Ризик-менеджмент методами інтелектуального аналізу даних.

Заступник директора ІПСА НТУУ «КПІ імені І.Сікорського»

д.т.н., професор

Професор ІПСА НТУУ «КПІ імені І.Сікорського».

д.т.н.



В. Д. Романенко

П. І. Бідюк