

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ
ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ
ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

На правах рукопису
УДК 004.896

До захисту допущено
Завідувач кафедри ММСА

Оксана ТИМОЩУК

«__» _____ 2022 р.

Магістерська дисертація

на здобуття ступеня магістра за спеціальністю 124 Системний аналіз на
тему: «Система підтримки прийняття рішень для аналізу розвитку
фінансових процесів»

Виконала:

студентка II курсу, групи КА-13мп
Радіо Ольга Володимирівна _____

Керівник:

професор кафедри ММСА
д.т.н, проф. Бідюк Петро Іванович _____

Рецензент:

професор НТУУ “КПІ ім. Сікорського”
д.т.н, проф. Теленик С. Ф. _____

Засвідчую, що у цій магістерській дисертації
немає запозичень з праць інших авторів
без відповідних посилань

Студент _____

Київ
2022

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ
ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ
ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

Рівень вищої освіти — другий (магістерський)
Спеціальність (ОПП) — 124 «Системний аналіз» («Системний аналіз
фінансового ринку»)

ЗАТВЕРДЖУЮ
Завідувача кафедри ММСА
Оксана ТИМОЩУК
«__» _____ 2022 р.

ЗАВДАННЯ

на магістерську дисертацію студентці Радіо Ользі Володимирівні

1. Тема дисертації: «Система підтримки прийняття рішень для аналізу розвитку фінансових процесів», науковий керівник дисертації Бідюк Петро Іванович, д.т.н., професор, затверджені наказом по університету від «03» листопада 2022 р. № 4046-с.

2. Термін подання студентом дисертації: 12 грудня 2022 р.

3. Об'єкт дослідження: процес аналізу розвитку фінансових процесів фондового ринку.

4. Предмет дослідження: математичні моделі і методи опису гетероскедастичних процесів, методи прогнозування часових рядів, оцінювання та аналізу якості побудованих моделей та прогнозів, моделі та методи оцінювання ринкових ризиків, а також методи перевірки якості оцінок ризику.

5. Перелік завдань, які потрібно розробити:

1) дослідити сучасний стан та особливості застосування машинного навчання у вирішенні проблеми оцінювання ризиків акцій;

2) розробити математичну модель нейронної мережі для прогнозування волатильності фінансових процесів;

- 3) розібрати і підібрати моделі для прогнозування гетероскедастичних часових рядів;
- 4) на основі моделей створити програмний продукт для визначення ризику акцій фінансового ринку;
- 5) знайти дані для застосування в програмі;
- 6) розробити стартап-проект виведення на ринок результатів дослідження;
- 7) розробити концептуальні висновки за результатами наукового дослідження.

6. Орієнтовний перелік графічного (ілюстративного) матеріалу:

- 1) Огляд процесу ціноутворення (рис.1.1)
- 2) Алгоритм обраних математичних моделей для прогнозу волатильності (рис.2.1 – рис.2.7);
- 3) Робота створеного програмного продукту (рис.3.1 - рис.3.7);
- 4) Таблиці у розділі стартап-проекту.

7. Орієнтовний перелік публікацій:

1. Радіо О.В., Бідюк П.І., Система підтримки прийняття рішень для аналізу розвитку фінансових процесів, Перша Всеукраїнська науково-практична конференція «Системний аналіз та інформатика», м. Київ, 22-29 листопада, 2022 року, с. 198

8. Дата видачі завдання: 1 вересня 2022 року.

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1.	Концептуальний вступ дисертації. Формулювання об'єкта, предмета, цілі, завдань, новизни, практичної значущості результатів.	1.09.2022 – 5.09.2022	Виконано
2.	Огляд літературно-інформаційних джерел.	5.09.2022 – 10.09.2022	Виконано

3.	Перший розділ. Понятійно-категоріальний апарат. Характеристика об'єкта.	10.09.2022 25.09.2022	–	Виконано
4.	Другий розділ. Детальний розгляд систем для трейдингу і процесів на біржі.	25.09.2022 30.09.2022	–	Виконано
5.	Третій розділ. Імплементация отриманих результатів у програмний продукт. Тестування програми.	30.09.2022 25.10.2022	–	Виконано
6.	П'ятий розділ. Стартап-проект.	25.10.2022 20.11.2022	–	Виконано
7.	Концептуальні висновки. Перспективи розвитку отриманих рішень.	20.11.2022 9.12.2022	–	Виконано

Студент

Ольга РАДІО

Науковий керівник дисертації

Петро БІДЮК

РЕФЕРАТ

Магістерська дисертація: 118 с., 15 рис., 25 табл., 1 додаток, 58 джерел.
ЦІНОУТВОРЕННЯ НА ФОНДОВОМУ РИНКУ, ВОЛАТИЛЬНІСТЬ,
ПРОГНОЗУВАННЯ, УМОВНА АВТОРЕГРЕСІЙНА
ГЕТЕРОСКЕДАСТИЧНІСТЬ, РЕКУРЕНТНІ НЕЙРОННІ МЕРЕЖІ,
РИНКОВИЙ РИЗИК, VALUE-AT-RISK.

Об'єкт дослідження – процес аналізу розвитку фінансових процесів фондового ринку.

Предмет дослідження – математичні моделі і методи опису гетероскедастичних процесів, методи прогнозування часових рядів, оцінювання та аналізу якості побудованих моделей та прогнозів, моделі та методи оцінювання ринкових ризиків, а також методи перевірки якості оцінок ризику.

Методи дослідження – теорія моделювання і прогнозування, регресійний аналіз, статистичні методи.

Метою роботи є побудова системи підтримки прийняття рішень, яка включає в себе адекватну модель гетероскедастичного процесу для прогнозування волатильності та оцінювання ризику акцій фінансового ринку за її допомогою.

В роботі проведено огляд основних підходів до оцінювання ринкових ризиків, розглянуто та проаналізовано метод оцінки Value-at-Risk. Також проведений огляд моделей та їх особливостей для опису динаміки волатильності та її прогнозування. Було проаналізовано результати моделювання та оцінювання за-для обґрунтованого вибору найкращої моделі для оцінки ринкових ризиків. Моделювання процесів на базі умовно гетероскедастичних моделей та на базі рекурентних нейронних мереж для

оцінювання ризикової вартості за їх допомогою реалізовано на мові програмування Python.

ABSTRACT

Master`s thesis: 118 p., 15 fig., 25 tab., 1 appendix, 58 sources.

STOCK MARKET PRICING, VOLATILITY, FORECASTING, CONDITIONAL AUTOREGRESSIVE HETEROSCEDASTICITY, RECURRENT NEURAL NETWORKS, MARKET RISK, VALUE-AT-RISK.

Object of research - the process of analyzing the development of financial processes of the stock market.

Subject of research - mathematical models and methods of description of heteroscedastic processes, methods of forecasting time series, evaluation and analysis of the quality of built models and forecasts, models and methods of market risk assessment, as well as methods of checking the quality of risk assessments.

Research methods - theory of modeling and forecasting, regression analysis, statistical methods.

The aim is to build a decision support system that includes an adequate model of the heteroskedastic process for forecasting volatility and assessing the risk of financial market shares with its help.

In this paper, it is reviewed of the main approaches to market risk estimation, reviewed and analyzed the method for estimating Value-at—Risk and applied innovative methods for verifying the quality of these estimates. Also reviewed models and their features to describe the dynamics of volatility and its forecasting. Results of modeling, forecasting and evaluation were analyzed for selecting the best model for market risks estimation.

Modeling and forecasting of financial and economic processes based on autoregressive conditionally heteroscedastic models and recurrent neural networks estimating the risk with their help are implemented in the programming language Python.

ЗМІСТ

ВСТУП.....	10
РОЗДІЛ 1 АКТУАЛЬНІСТЬ ЗАСТОСУВАННЯ СИСТЕМ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ АНАЛІЗУ ФІНАНСОВИХ ПРОЦЕСІВ ФОНДОВОГО РИНКУ. ОСОБЛИВОСТІ ЦІНОУТВОРЕННЯ НА ФОНДОВИХ РИНКАХ.....	12
1.1 Поняття СППР. Види та переваги їх застосування	12
1.2 Поняття фондового ринку.....	16
1.3 Особливості ціноутворення на фондовому ринку та ринкові ризики	18
1.4 Поняття ринкового ризику.....	22
1.5 Методи прогнозування, які можна застосовувати для аналізу фінансових процесів фондового ринку.....	23
1.6 Висновки до розділу 1.....	31
РОЗДІЛ 2 ВИБІР ТА ОПИС МАТЕМАТИЧНИХ МОДЕЛЕЙ ДЛЯ ПРОГНОЗУВАННЯ ФІНАНСОВИХ ПРОЦЕСІВ ФОНДОВОГО РИНКУ	34
2.1 Огляд існуючих моделей та алгоритмів машинного навчання для прогнозування фінансових процесів на фондовому ринку та вибір оптимальної моделі.....	34
2.1.1 Визначення набору вхідних даних для прогнозування	34
2.1.3 Опис обраних моделей.....	48
2.2 Особливості перевірки на гетероскедастичність.....	51
2.4 Показники якості оцінок прогнозів можливих втрат	57
2.5 Висновки до розділу 2.....	58

РОЗДІЛ 3 ПОБУДОВА СППР ТА ВИКОНАННЯ ОБЧИСЛЮВАЛЬНИХ ЕКСПЕРИМЕНТІВ	60
3.1 Архітектура та функціональна схема СППР	60
3.2 Аналіз вхідних даних	61
3.3 Побудова математичних моделей фінансових процесів та дисперсії	65
3.4 Результати оцінювання ринкового ризику	68
3.5 Висновки до розділу 3.....	69
РОЗДІЛ 4 РОЗРОБКА СТАРТАП ПРОЕКТУ	70
4.1 Постановка задачі стартап-проекту.....	70
4.2 Карта проекту	71
4.3 Команда стартап- проекту	72
4.4 Розробка бізнес-моделі проекту	73
4.6 Розроблення ринкової стратегії проекту.....	84
4.7 Розробка маркетингової програми стартап-проекту	89
4.8 Висновки до розділу 4.....	93
ВИСНОВОК	94
ПЕРЕЛІК ПОСИЛАНЬ	96
ДОДАТОК А ЛІСТИНГ ПРОГРАМИ	102

ВСТУП

Актуальність теми. В нинішній час фондові ринки широко розповсюджені та в певній мірі впливають на макроекономічні показники будь-якої сфери чи в цілому країни та їх можна віднести до ефективних. Тобто фондовий ринок – це ринок, на якому активно конкурує велика кількість інвесторів, що прагнуть до максимізації прибутку компаній, кожна з яких намагається передбачити майбутню ринкову вартість окремих цінних паперів, і де важлива поточна інформація майже вільно доступна для всіх учасників. Іншими словами, на ефективному ринку будь-якої миті часу фактична ціна цінного паперу буде гарною оцінкою її внутрішньої вартості.

Денний оборот фондового ринку налічується мільярдами доларів та мільйонами торгових доручень. З кожним роком учасників ринку стає дедалі більше, серед них фінансові організації, хеджфонди, банки та приватні інвестори. Трейдери використовують різні стилі торгівлі та системи прийняття рішень, кожен стиль чи система має різну економічну ефективність та ступінь ризику.

Таким чином, в нинішній час фондовий ринок стає популярним інвестиційним майданчиком як для інституційних, так і для індивідуальних інвесторів.

Волатильність ринку та інновації у фінансовій торгівлі зробили прийняття інвестиційних рішень все більш складним та ризикованим. Тим не менш, доступність надійних даних фінансового ринку та потужних програмних інструментів є найвищою. У таких умовах системи підтримки прийняття рішень (СППР) грають найважливішу роль і підвищують якість рішень, прийнятих як професійними, так і індивідуальними інвесторами.

Отже, користувачам, які прагнуть підвищити ефективність ухвалення рішень на фондовому ринку, необхідне застосування СППР, що й пояснює актуальність обраного дослідження.

Мета роботи: розробка системи підтримки прийняття рішень для аналізу фінансових процесів фондового ринку.

Задачі дослідження:

1. Показати актуальність застосування систем підтримки прийняття рішень для аналізу фінансових процесів фондового ринку. Розглянути особливості ціноутворення на фондових ринках.

2. Виконати вибір та навести опис математичних моделей для прогнозування фінансових процесів фондового ринку.

3. Побудувати СППР та виконати обчислювальні експерименти.

Об'єкт дослідження – процес аналізу розвитку фінансових процесів фондового ринку.

Предмет дослідження – система підтримки прийняття рішень для аналізу розвитку фінансових процесів фондового ринку.

Методи дослідження: аналіз, порівняння, статистики, дослідження, експерименту, математичного моделювання.

Наукова новизна отриманих результатів полягає у розробці ефективної системи підтримки прийняття рішень для аналізу розвитку фінансових процесів фондового ринку.

Практичне значення одержаних результатів: розроблену систему підтримки прийняття рішень можна застосовувати для аналізу розвитку фінансових процесів на фондовому ринку.

РОЗДІЛ 1 АКТУАЛЬНІСТЬ ЗАСТОСУВАННЯ СИСТЕМ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ АНАЛІЗУ ФІНАНСОВИХ ПРОЦЕСІВ ФОНДОВОГО РИНКУ. ОСОБЛИВОСТІ ЦІНОУТВОРЕННЯ НА ФОНДОВИХ РИНКАХ

1.1 Поняття СППР. Види та переваги їх застосування

В межах даного розділу магістерської дисертації покажемо актуальність застосування СППР для аналізу фінансових процесів фондового ринку, а також визначимо особливості ціноутворення на фондових ринках.

Система підтримки прийняття рішень, або СППР, являє собою комп'ютерну інформаційну систему, яка організовує, збирає й аналізує, які можуть використовуватися під час прийняття рішень для управління, експлуатації та планування [1].

Системи підтримки прийняття рішень призначені для того, щоб допомогти людям аналізувати великі обсяги даних для прийняття обґрунтованих рішень. Системи - це просто комп'ютерні системи, призначені для зберігання та аналізу даних, що мають відношення до прийняття. Іншими словами, коли потрібно ухвалити рішення, можна звернутися до нашої системи підтримки прийняття рішень і подивитися, що «кажуть» дані про це рішення або вибір, який необхідно зробити [2,3].

Системи підтримки прийняття рішень можуть створювати складні моделі або уявлення того, що відбувається з даними [4].

Приклади систем підтримки прийняття рішень включають ручні системи, гібридні системи, різноманітні типи аналітики, а також складне програмне забезпечення для підтримки прийняття рішень. Фактором, який відрізняє новіші комп'ютерні системи від ранніх систем підтримки ухвалення рішень, є їхня здатність аналізувати надзвичайно великі набори даних, надаючи засновані

на даних рекомендації, що виключають здогади під час ухвалення рішень.

Тоді як дехто заперечує проти того, щоб довіряти складним комп'ютерним програмним рішенням приймати рішення за них, більшість воліє використовувати згенеровану комп'ютером статистику для розуміння ключових тенденцій. До них належать такі аналітичні дані, як статистика продажів, гарантійні ставки та тенденції руху грошових коштів, які є важливими індикаторами, що допомагають користувачам визначати стан свого бізнесу і підказувати необхідність коригувальних дій.

Складність у тому, що цей рівень інформації не може визначити, яка з кількох можливостей максимізує віддачу при досягненні бажаного результату. Він також не може передбачити зовнішні зміни, які можуть вплинути на прибутковість, що є важливим чинником, оскільки більшість компаній працюють у невизначеному середовищі, яке визначається споживчими настроями, правовими нормами та жорсткою конкуренцією. Крім того, компанії вразливі до зовнішніх впливів, таких як політична невизначеність, великі погодні явища і торгові суперечки.

Три ключові елементи DSS включають в себе:

- організаційні дані: відповідна інформація і знання;
- модель: математичні та статистичні формули, які представляють бізнес і аналізують дані;
- призначений для користувача інтерфейс: інформаційні панелі або інші інтерфейси, що дають змогу користувачам взаємодіяти з результатами та переглядати їх.

Системи підтримки ухвалення рішень працюють на багатьох рівнях, і існує безліч прикладів їх повсякденного використання [2].

Як зазначається в [5], найефективнішими прикладами систем підтримки прийняття рішень є ті, які визначають найкраще рішення на основі певних критеріїв. Такі системи усувають суб'єктивність і упередженість у процесі прийняття рішень. Крім того, вони можуть оцінити безліч альтернативних

сценаріїв і визначити найкращий із них.

Традиційний підхід розробки СППР полягає в тому, щоб визначити математичну модель, подивитися, як система приймає рішення, і використовувати програмне забезпечення для оптимізації, щоб визначити результати різних сценаріїв. Цей метод ґрунтується на розпорядчій аналітиці і є надзвичайно ефективним. Хоча дехто вважає, що моделювати слід тільки процес ухвалення рішень, розробка повної моделі організації підвищує універсальність і точність з точки зору фінансових результатів [6].

Існує два підходи до оптимізації: підхід, заснований на правилах і моделі оптимізації. Моделі, засновані на правилах (евристика), добре працюють, коли можливі результати можуть бути значною мірою визначені наперед, наприклад, під час оцінки ризику. З іншого боку, моделі оптимізації більш адаптовані, можуть розв'язувати складніші задачі та мати справу з численними обмеженнями та компромісами.

СППР будь-якого типу містить [3]:

- підсистему введення та аналізу запитів користувача (ПВАЗ);
- підсистему оброблення запитів користувача та генерації результатів (ПОЗГР);
- базу знань і даних (БЗД);
- підсистему подання результатів (ППР) в зручній для користувача формі.

Як зазначається в [7], СППР можна розділити на категорії, кожна з яких ґрунтується на їхніх первинних джерелах інформації.

1. СППР на основі даних.

СППР, керована даними, - це комп'ютерна програма, яка приймає рішення на основі даних із внутрішніх або зовнішніх баз даних. Як правило, СППР, керована даними, використовує методи інтелектуального аналізу даних для виявлення тенденцій і закономірностей, що дає змогу прогнозувати майбутні події. Користувачі часто використовують СППР, керовані даними, для прийняття рішень різноманітні бізнес-процеси. Деякі з них використовуються

для допомоги в ухваленні рішень у державному секторі, наприклад, для прогнозування ймовірності злочинної поведінки в майбутньому.

2. СППР на основі моделі. Системи підтримки прийняття рішень, що базуються на базовій моделі прийняття рішень, налаштовуються відповідно до зумовленого набору вимог користувача, щоб допомогти аналізувати різні сценарії, що відповідають цим вимогам. Наприклад, СППР на основі моделі може допомогти в плануванні або розробленні фінансових звітів.

3. Комунікативні і групові СППР. У керованій комунікацією і груповій системі підтримки прийняття рішень використовуються різні засоби зв'язку, такі, як електронна пошта, обмін миттєвими повідомленнями або голосовий чат, щоб дати змогу кільком людям працювати над одним і тим самим завданням. Метою цього типу СППР є розширення співпраці між користувачами та системою і підвищення загальної ефективності та дієвості системи.

4. СППР, заснована на знаннях. У цьому типі системи підтримки прийняття рішень дані, які керують системою, містяться в базі знань, яку постійно оновлюють і підтримують системою управління знаннями. СППР, заснована на знаннях, надає користувачам інформацію, що відповідає бізнес-процесам.

5. СППР на основі документів. СППР, керована документами, - це тип системи управління інформацією, яка використовує документи для вилучення даних. Керовані документами СППР дають змогу користувачам виконувати пошук на веб-сторінках або в базах даних чи знаходити певні умови пошуку. Приклади документів, до яких звертаються DSS на основі документів, включають політики і процедури, протоколи зборів і корпоративні записи.

Переваги СППР [3]:

- підвищують ефективність та швидкість прийняття рішень;
- підвищують контроль, конкурентоспроможність та можливості прийняття рішень;

- заохочують навчатися або тренуватися;
- автоматизують монотонні процеси, отже, більше часу може бути витрачено безпосередньо для прийняття рішень.

На фондовому ринку успішні інвестори можуть отримати максимальний прибуток залежно від вибору акцій і відповідного часу для торгівлі. За допомогою СППР користувачі зможуть підвищити ефективність аналізу фінансових процесів фондового ринку, саме тому їх актуально застосовувати.

1.2 Поняття фондового ринку

Фондовий ринок відноситься до публічних ринків, які існують для випуску, купівлі та продажу акцій, які торгуються на фондовій біржі або позабіржовому ринку [8].

Акції являють собою пайове володіння компанією, а фондовий ринок - це місце, де інвестори можуть купувати і продавати право власності на такі інвестиційні активи. Ефективно функціонуючий фондовий ринок вважається критично важливим для економічного розвитку, оскільки він дає компаніям можливість швидкого доступу до капіталу від населення [9].

Фондовий ринок служить двом дуже важливим цілям. По-перше, надати капітал компаніям, які вони можуть використати для фінансування та розширення свого бізнесу. Наприклад, якщо компанія випускає один мільйон акцій, які спочатку продаються за ціною 10 доларів за акцію, то це дає компанії 10 мільйонів доларів капіталу, які вона може використати для розвитку свого бізнесу (за вирахуванням комісій, які компанія сплачує інвестиційному банку за управління акціями). Пропонуючи акції замість запозичення капіталу, необхідного для розширення, компанія уникає виникнення боргів і сплати відсотків за цим боргом [9].

Другорядна мета, якій служить фондовий ринок, полягає в тому, щоб дати інвесторам - тим, хто купує акції, - можливість брати участь у прибутках публічних компаній. Інвестори можуть отримати прибуток від купівлі акцій одним із двох способів. За деякими акціями виплачуються регулярні дивіденди (певна сума грошей на акцію, якою хтось володіє). Інший спосіб, яким інвестори можуть отримати прибуток від купівлі акцій, - це продати свої акції з прибутком, якщо ціна акцій збільшиться порівняно з їхньою купівельною ціною. Наприклад, якщо інвестор купує акції компанії за ціною 10 доларів за акцію, а ціна акцій згодом підвищується до 15 доларів за акцію, інвестор може отримати 50% прибутку від своїх інвестицій, продавши свої акції.

Є низка постійних учасників торгів на фондовому ринку [10].

Інвестиційні банки проводять первинне публічне розміщення (IPO) акцій, яке відбувається, коли компанія вперше вирішує стати публічною компанією, пропонуючи акції.

Наприклад, компанія, що бажає стати публічною і запропонувати акції, звертається до інвестиційного банку з проханням виступити в якості «андерайтера» початкового розміщення акцій компанії. Інвестиційний банк, вивчивши загальну вартість компанії та взявши до уваги, від якої частки власності компанія бажає відмовитися у вигляді акцій, здійснює початковий випуск акцій на ринку за певну плату, гарантуючи при цьому, що компанії встановлена мінімальна ціна за акцію. Тому в інтересах інвестиційного банку стежити за тим, щоб усі пропоновані акції були продані і за максимально можливою ціною.

Акції, пропоновані в рамках IPO, найчастіше купуються великими інституційними інвесторами, такими як пенсійні фонди або компанії взаємних фондів.

Ринок IPO відомий як первинний або початковий ринок. Після того, як акції були випущені на первинному ринку, вся торгівля акціями після цього відбувається через фондові біржі на так званому вторинному ринку. Термін

"вторинний ринок" трохи вводять в оману, оскільки саме на цьому ринку щодня відбувається переважна більшість угод з акціями [10].

Біржові маклери, які можуть виступати або не виступати як фінансові консультанти, купують і продають акції для своїх клієнтів, які можуть бути або інституційними інвесторами, або індивідуальними роздрібними інвесторами.

Аналітики з дослідження акцій можуть бути найняті брокерськими фірмами, компаніями взаємних фондів, хедж-фондами або інвестиційними банками. Це люди, які досліджують компанії, що публічно торгуються, і намагаються передбачити, чи будуть акції компанії зростати або падати в ціні.

Управителі фондами або портфельні управителі, до яких відносяться управителі хедж-фондів, управителі взаємними фондами та управителі біржовими фондами (ETF), є важливими учасниками фондового ринку, оскільки вони купують і продають велику кількість акцій. Якщо популярний взаємний фонд вирішує вкласти значні кошти в певну акцію, цей попит на одну акцію часто буває досить значним, щоб помітно підняти ціну акції.

1.3 Особливості ціноутворення на фондовому ринку та ринкові ризики

Розглянемо особливості ціноутворення на фондовому ринку.

Як зазначається в роботі [11], комп'ютеризація фондових ринків і доступність детальних електронних записів потоку заявок і динаміки цін на фондових ринках за останнє десятиліття вивільнила терабайти високочастотних даних про транзакції, потік заявок і динаміку портфеля заявок на фондових ринках, які дають детальне уявлення про високочастотну динаміку попиту, пропозиції та цін на цих ринках. Ці дані можуть бути використані для дослідження природи механізму ціноутворення, який описує, як ринкові ціни реагують на коливання попиту та пропозиції. На високому рівні "механізм

ціноутворення" - це карта, яка відображає взаємозв'язок між ринковою ціною та такими змінними, як ціна, історія та потік замовлень [12]:

$$\begin{aligned} & Price(t + \Delta t) \\ & = F(Price\ history\ (0 \dots t),\ Order\ Flow\ (0 \dots t),\ Other\ Information) \\ & = F(X_t, \varepsilon_t), \end{aligned}$$

де X_t - набір змінних стану (наприклад, лагові значення ціни, волатильності та потоку заявок), наділених певною динамікою;

ε_t - випадковий "шум" або інноваційний член, що представляє надходження нової інформації та інші ефекти, які не повністю охоплюються змінними стану.

Емпіричні та теоретичні моделі мікроструктури ринку, стохастичні моделі та моделі прогнозування цін можуть розглядатися як різні способи представлення цієї карти F , з різною часовою роздільною здатністю Δt .

Одне з питань, яке присутнє в літературі, полягає в тому, наскільки ця карта F є універсальною (тобто незалежною від конкретного активу, що розглядається). Загальне, на відміну від специфічної для конкретного активу формулювання моделей мікроструктури ринку, здається, припускає таку універсальність. Емпіричні докази універсальності деяких стилізованих фактів [12] та літературні джерела [13, 14, 15], підтримують гіпотезу універсальності. Разом з тим, як зазначається в [11], практика статистичного моделювання фінансових часових рядів залишається специфічною для окремих активів.

Крім того, дані, що використовуються для оцінки, часто обмежуються нещодавнім періодом часу, що відображає переконання в тому, що фінансові дані можуть бути "нестационарними" і схильними до змін, які можуть зробити більш старі дані менш релевантними для прогнозування.

Через такі міркування моделі, що розглядаються у фінансовій економетриці, трейдингу та управлінні ризиками, є специфічними для

конкретних активів, а їх параметри переоцінюються з часом з використанням часового вікна останніх даних. Тобто, для активу i в момент часу t модель набуває вигляду:

$$Price_i(t + \Delta t) = F(X_{0:t}^i, \varepsilon_t | \theta_i(t)),$$

де параметр моделі $\theta_i(t)$ періодично оновлюється з використанням останніх даних про ціну та інші змінні стану, пов'язані з активом i .

В результаті, набори даних є фрагментованими за активами та часом i , навіть у високочастотній області, розмір наборів даних, що використовуються для оцінки та навчання моделі, на порядки менший, ніж в інших сферах, де успішно застосовується аналітика великих даних.

Це одна з причин, чому, за винятком кількох випадків, великомасштабні методи навчання, такі як глибоке навчання, не були до недавнього часу застосовані для кількісного моделювання у фінансах. Зокрема, значення аргументу, яке пов'язане з не стаціонарністю, іноді застерігає від їх використання.

З іншого боку, якби зв'язок між цими змінними був універсальним і стаціонарним, тобто якби параметр $\theta_i(t)$ не змінювався ні з активом i , ні з часом t , тоді можна було б об'єднати дані по різних активах і періодах часу і використовувати набагато багатший набір даних для оцінки/навчання моделі. Наприклад, дані про епізод раптового краху на ринку одного активу можуть дати уявлення про те, як ціна іншого активу відреагує на серйозні дисбаланси в потоці заявок, незалежно від того, чи траплявся такий епізод в історії цього ринку, чи ні.

В роботі [16] зазначається, що ціна на фондовому ринку формується на основі рівноваги пропозиції та попиту, а також на основі залежності кількості угод від ціни, як показано на рисунку 1.1.

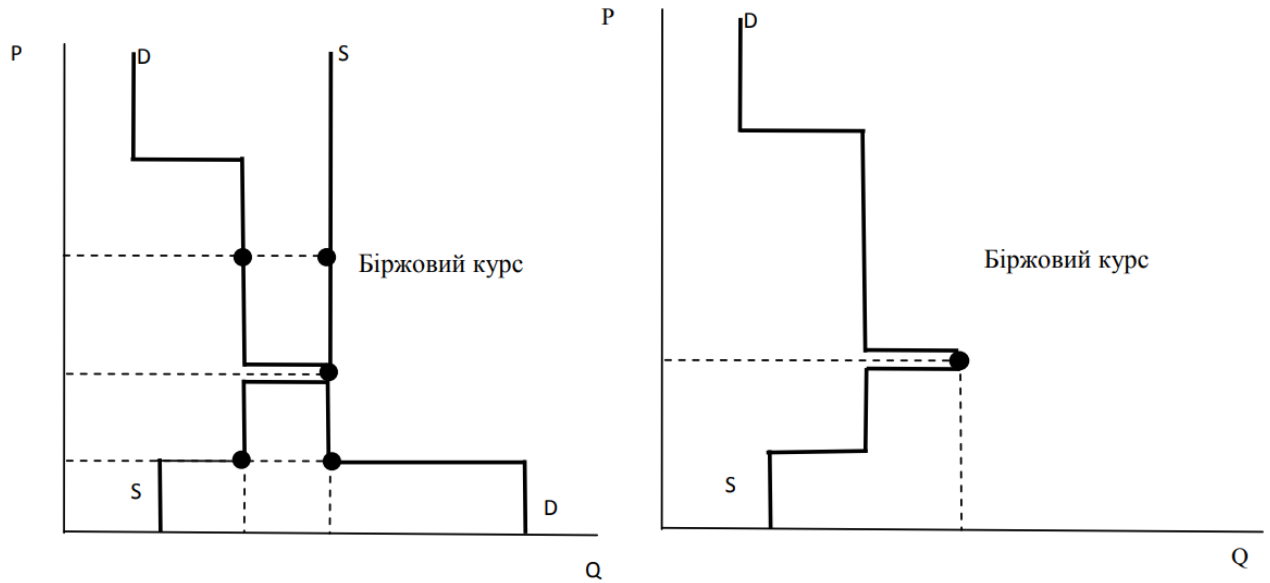


Рисунок 1.1 – Формування ціни на фондовому ринку: а) на основі рівноваги пропозиції та попиту; б) на основі залежності кількості угод від ціни

Також в роботі [16] зазначається, що також існують такі основні теорії ціноутворення фондового ринку:

- 1) теорія статичних цін (SPT) К. Ерроу та Дж. Дебре;
- 2) сучасна теорія портфелів Г. Марковіца й Модель капітального ціноутворення (CAPM) У. Шарпа;
- 3) теорія арбітражного ціноутворення Росса (APT);
- 4) модель Кокрейна.

Теорія статичних цін (SPT) Ерроу-Дебре – основоположна теорія ціноутворення, що дає можливість зрозуміти загальні його принципи на ринку цінних паперів. Модель капітального ціноутворення (CAPM) і теорія арбітражного ціноутворення (APT) побудовані на основі теорії статичних цін (SPT) із додаванням додаткових припущень щодо інвестиційних уподобань, розподілу доходів чи акцентування на умові відсутності арбітражу.

У основу моделі Кокрейна покладено принцип оцінки вартості майбутніх невизначених грошових потоків [16].

1.4 Поняття ринкового ризику

Під ринковим ризиком розуміють це ризик збитків за фінансовими вкладеннями, спричинений несприятливим рухом цін. Прикладами ринкового ризику є: зміни цін на акції або товари, зміни відсоткових ставок або коливання валютних курсів.

Волатильність цін часто виникає через непередбачені коливання факторів, які зазвичай впливають на весь фондовий ринок. Систематичний ризик не пов'язаний конкретно з компанією або галуззю, в яку він інвестує; натомість він залежить від продуктивності всього ринку. Таким чином, інвестору необхідно стежити за різними макрозмінними, пов'язаними з фінансовим ринком [17].

Оскільки ризик зачіпає весь ринок, його не можна диверсифікувати, щоб зменшити, але можна «геджувати» мінімальні ризики. В результаті інвестори можуть не отримати очікуваного прибутку, незважаючи на суворе застосування фундаментального та технічного аналізу конкретного варіанту інвестування.

Волатильність або абсолютна/відсоткова дисперсія цін часто вважається гарним заходом ринкового ризику. Професійні аналітики також схильні використовувати такі методи, як моделювання вартості під ризиком (VaR) для виявлення потенційних збитків за допомогою статистичного управління ризиками.

Метод VaR є стандартним методом оцінки ринкового ризику. Техніка VaR — це метод управління ризиками, який включає використання статистики, яка кількісно визначає ймовірні збитки за акціями або портфелями, а також ймовірність виникнення цих збитків. Детальніше про даний метод та особливості його застосування буде наведено у другому розділі магістерської дисертації.

1.5 Методи прогнозування, які можна застосовувати для аналізу фінансових процесів фондового ринку

Метою аналізу часового ряду є побудова прогнозу його значень на майбутній період, а задачами – виявлення компонент, під впливом яких формуються значення, і побудова математичної моделі кожної компоненти чи їх сукупності [18].

Метод прогнозування – певний набір дій, у результаті якого, визначається модель прогнозування конкретного часового ряду.

Модель прогнозування – функціональне уявлення, адекватно описує досліджуваний процес і є основою отримання його майбутніх значень.

Прийом прогнозування – математична чи логічна операція, яка дозволяє отримати конкретні результати у процесі розробки прогнозів.

Методи прогнозування умовно можна розділити на дві групи: інтуїтивні та формалізовані.

Інтуїтивні методи прогнозування – це судження та оцінка експертів. Якщо об'єкт занадто простий або навпаки занадто складний, такий, що аналітично неможливо врахувати вплив багатьох чинників, застосовується інтуїтивне прогнозування. Тобто опитують експертів та отримані експертні оцінки як індивідуальні, так і колективні використовуються як кінцеві прогнози або як вихідні дані в комплексних системах прогнозування. Цей метод активно застосовується у маркетингу, економіці, політиці, проте не може застосовуватися для прогнозування ціни на фондових ринках.

Формалізовані методи прогнозування – це методи прогнозування, у яких будують моделі прогнозування, тобто визначають таку математичну залежність, яка дозволяє зробити прогноз, тобто обчислити майбутнє значення процесу.

Моделі для прогнозування часових рядів поділяють на статистичні та структурні.

Функціональна залежність у статистичних моделях, між фактичними та майбутніми значеннями часового ряду, а також зовнішніми факторами задана аналітично.

До статистичних моделей належать, але не вичерпують [19]:

- регресійні моделі;
- авторегресійні моделі;
- моделі експонентного згладжування.

До структурних моделей відносять:

- нейромережеві моделі;
- моделі на базі ланцюгів Маркова;
- моделі з урахуванням класифікаційно-регресійних дерев.

Прогнозування часових рядів цін на фондових ринків - одне із найскладніших завдань у галузі часових рядів. Моделі прогнозування залежать від вибору завдання, тобто від того, чи хочуть вони спрогнозувати тенденцію (вгору чи вниз) на фондовому ринку, яка моделюється як завдання класифікації, або прогноз числового значення, тобто ціни акцій чи фондових індексів, які моделюються як завдання регресії.

В останні кілька років було запропоновано низку підходів, які можна застосувати для прогнозування часових рядів фондового ринку.

Щоб врахувати різні за своєю природою і навіть структурою, аналізовані показники підпорядковуються наступним операціям (у відповідному порядку):

- нормалізація (зазвичай усі показники приводяться до безрозмірного виду);
- зважування (зазвичай шляхом простого множення вагового коефіцієнта значення показника);
- згортки (об'єднання на єдиний підсумковий результат всіх окремих нормованих і зважених показників).

Ці операції можна виконувати різними способами, серед яких найчастіше використовуються такі три комбінації, які представлені формулами:

- мультиплікативна формула:

$$x_i = \frac{\prod_{k=1}^{l_1} (u_{ik} \cdot \rho_{ik} \cdot w_{ik})}{\prod_{k=l_1+1}^{l_1+l_2} (u_{ik} \cdot \rho_{ik} \cdot w_{ik}^{-1})},$$

де l_1 - кількість вхідних індикаторів, збільшення яких викликає збільшення значення x_i , такі індикатори, згідно з термінологією [19], ми називатимемо стимулюючими, підсилюючими (індикатори-стимулятори);

ρ_{ik} - коефіцієнт приведення k -го показника до єдиної міри (розмірності), нормуючий коефіцієнт, який часто розраховується як $1/u_{ik \max}$;

ω_{ik} – ваги відповідних показників.

У цьому підході ми також повинні враховувати константу l_2 - кількість вхідних індикаторів, збільшення яких викликає зменшення прогнозованого значення x_i (назвемо їх дестимульованими, що знижуються). Тоді справедливий такий вираз:

$$l_1 + l_2 = l.$$

- Формула, заснована на теорії адаптивної корисності:

$$x_i = \sum_{k=1}^{l_1} (u_{ik} \cdot \rho_{ik} \cdot w_{ik}) + \sum_{k=l_1+1}^{l_1+l_2} \left(\frac{w_{ik}}{u_{ik} \cdot \rho_{ik}} \right).$$

Остання формула взята за основу в [19] та ряді аналогічних робіт (з невеликими варіаціями). Зокрема, метод прогнозування складається із трьох етапів:

- розрахунок відносних значень кожного показника (тобто приведення їх до єдиної міри), тобто аналог розрахунку $u_{ik} \cdot \rho_{ik}$;

- розрахунок індексу загального зростання x_i шляхом множення кожного нормалізованого параметра на його ваговий коефіцієнт та його підсумовування відповідно до формули (1.4);

- прогнозування нового значення ціни на фондовому ринку (наприклад, акції), шляхом множення його існуючого значення на коефіцієнт поправки, який також нормалізується за формулою (1.1):

$$C_i = C_{i-1} \cdot \left(1 + \frac{x_i}{x_m} \right), \quad (1.1)$$

де x_m - максимальне значення поправочного коефіцієнта, яке досягається, коли всі значення $u_{ik} \cdot \rho_{ik}$ досягають своїх крайніх значень і дорівнюють простій сумі всіх ваг w_{ik} .

Це підхід, заснований на обліку поточних значень показників, які, на думку експертів, є важливими для зміни ціни на акції в майбутньому щодо її існуючого значення.

Розглянемо також існуючий підхід до побудови тимчасового тренду (тимчасового ряду), який ефективно використовується у розділі математичної статистики, як регресійний аналіз.

У даному випадку ціна на акції змінюється з часом. Строго математичною мовою одне значення C (ціна) залежить від іншого значення t (часу). Будемо тимчасово використовувати позначення x для аргументу, а y - для функції. Встановивши остаточний вид формули, яка буде розроблена, повернемося до позначень t , C .

Одне контрольоване значення на виході може бути функцією багатьох параметрів:

$$y = f(x_1, x_2, \dots, x_m).$$

Метою багатовимірного регресійного аналізу є визначення коефіцієнтів a_i на основі наявних експериментальних даних [19].

Для виконання певних математичних операцій необхідно виконати обробку трьох наведених нижче матриць.

Перша – це матриця регресії:

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix}.$$

Матриця-стовпець контрольованих значень:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}.$$

Останній - матриця-стовпець шуканих коефіцієнтів:

$$Y = \begin{pmatrix} a_0 \\ a_1 \\ \dots \\ a_m \end{pmatrix}.$$

Використовуючи метод найменших квадратів, операції диференціювання за компонентами вектора та виконання необхідних перетворень, отримуємо наступну формулу для знаходження невідомих коефіцієнтів рівняння множинної лінійної регресії:

$$A = (X^T X)^{-1} X^T Y \quad (1.2)$$

де буква T означає операцію транспонування (заміну рядків стовпцями чи навпаки); степінь мінус 1 означає перевернуту матрицю.

Як бачимо, залежність (1.2) складна, оскільки потребує множення матриць і, найголовніше, знаходження інвертованої матриці. Ці операції потребують використання спеціалізованого математичного програмного забезпечення. Справді, немає сенсу використовувати складні алгоритмічні методи у мовах програмування загального призначення (таких як РНР чи JavaScript), хоча існують спеціалізовані програмні продукти, у яких операції виконуються однією командою. Формула для ймовірного визначення ціни на акцію за допомогою такого підходу на фондовому ринку буде такою:

$$C_{i+1} = \left(\frac{\frac{\sum t_i C_i}{n} - t_{av} C_{av}}{\frac{\sum t_i^2}{n} - (t_{av})^2} t + \left(C_{av} - \frac{\frac{\sum t_i C_i}{n} - t_{av} C_{av}}{\frac{\sum t_i^2}{n} - (t_{av})^2} t_{av} \right) \right) \times$$

$$\times \left(1 + \frac{\sum_{k=1}^{l_1} (u_{ik} \cdot \rho_{ik} \cdot w_{ik}) + \sum_{k=l_1+1}^{l_1+l_2} \left(\frac{w_{ik}}{u_{ik} \cdot \rho_{ik}} \right)}{x_m} \right).$$

Виконуючи нескладні арифметичні перетворення та підстановки, отримуємо ймовірну формулу для розрахунку ціни на фондовому ринку акції:

$$C_{i+1} = \left(\frac{\frac{\sum t_i C_i}{n} - t_{av} C_{av}}{\frac{\sum t_i^2}{n} - (t_{av})^2} (t - t_{av}) + C_{av} \right) \times$$

$$\times \left(1 + \frac{\sum_{k=1}^{l_1} (u_{ik} \cdot \rho_{ik} \cdot w_{ik}) + \sum_{k=l_1+1}^{l_1+l_2} \left(\frac{w_{ik}}{u_{ik} \cdot \rho_{ik}} \right)}{\sum_{k=1}^{l_1+l_2} w_{ik}} \right).$$

Одним з ефективних методів є застосування метричних моделей для подальшого прогнозування і побудова анаморфоз.

Метричні моделі - алгоритм класифікації, заснований на обчисленні оцінок подібності між об'єктами. Найпростішим метричним класифікатором є метод найближчих сусідів, в якому об'єкт відноситься до того класу, якому належить більшість схожих з ним об'єктів.

Логічні моделі - алгоритми класифікації, засновані на припущенні про те, що в даних є деякі «логічні закономірності». Для вибірки будується комбінація логічних предикатів, що однозначно дозволяє класифікувати довільні об'єкти.

Типовим методом вирішення логічного класифікатора є вирішальне дерево. Перевагою логічних методів класифікації є висока інтерпретованість і простота, але з іншого боку, для них властиво перенавчання і висока чутливість до завад.

Байєсова теорія класифікації - широкий клас алгоритмів класифікації, заснований на принципі максимуму апостеріорної ймовірності. Для об'єкта обчислюються функції правдоподібності кожного з класів, по ним обчислюються апостеріорні ймовірності класів. Об'єкт відноситься до того класу, для якого ймовірність максимальна.

Метод логістичної регресії лежить на стику ідей лінійних і байєсових класифікаторів - він дозволяє будувати роздільну площину між класами і оцінювати апостеріорні ймовірності приналежності об'єктів класів.

Анаморфозою називається таке перетворення системи координат, при якому дані, які підлягають обробці, повинні вишикуватися в лінійну залежність. З даної лінійної залежності в результаті визначаються безпосередньо параметри розподілу [19].

Важливо відзначити, що такі величини, як дисперсія і математичне очікування повинні обчислюватися в тому припущенні, що вибірка належить безпосередньо одній генеральній сукупності.

Побудова анаморфоз для різного роду розподілів вказує на те, що існують якісь «хвости», які можуть значно відрізнитися від основної частини певної вибірки.

Таким чином, побудова таких анаморфоз дає можливість уточнення характеристик параметрів (математичного очікування, дисперсії).

Також сучасними методами прогнозування часових рядів є: нейронна мережа, генетичний алгоритм, нечітка логіка та інші підходи до еволюційних обчислень.

Штучна нейронна мережа – один із ефективних методів прогнозування фондового ринку, оскільки він не містить складних формул порівняно з традиційними лінійними та нелінійними моделями. ШНМ стали популярним методом, який використовується для прогнозування фондового ринку з останнього десятиліття. Широке використання ШНМ для прогнозування часових рядів фондового ринку обумовлено здатністю ШНМ обробляти дані, що характеризуються нелінійністю, високочастотними поліноміальними компонентами та перервністю [20].

ІНС – це керовані даними методи, що самоадаптуються, які можуть приймати нелінійність часових рядів без урахування будь-яких статистичних припущень про дані [20].

ШНМ – модель, що широко використовується для завдання прогнозування часових рядів у контексті її універсальних можливостей апроксимації. Звичайні ШНМ з дрібною архітектурою важко навчити, якщо вони стають занадто складними, наприклад, коли мережа включає багато рівнів і, отже, безліч параметрів.

На відміну від архітектур дрібної ШНМ широко продемонстровано, що архітектура глибокої ШНМ, яка називається глибокою нейронною мережею (ГНМ), перевершує традиційну архітектуру дрібної ШНМ [20].

Останнім часом глибоке навчання набуло значної популярності у співтоваристві машинного навчання, оскільки воно вважається загальною структурою, яка полегшує навчання глибоких нейронних мереж з багатьма прихованими шарами.

Найважливішою перевагою глибокого навчання є те, що для вивчення не потрібні будь-які створені функції вручну, і він може легко вивчити ієрархічне представлення функцій безпосередньо з необроблених даних.

Існує безліч моделей нейронних мереж, які широко використовуються для вирішення кількох типів задач прогнозування часових рядів. Серед цих моделей велика увага приділяється й рекурентним нейронним мережам (РНН, RNN). Причина полягає в тому, що РНН є класом моделей ШНМ, які мають внутрішній стан або короткострокову пам'ять через повторювані з'єднання зворотного зв'язку, що робить RNN придатними для моделювання послідовних даних або даних тимчасових рядів.

При такому моделюванні RNN підтримує вектор параметрів активації кожного тимчасового кроку, особливо коли у вхідні дані включені короткострокові залежності. Однак, якщо вони навчені зі стохастичним градієнтним спуском, їм буде складно вивчити довгострокові залежності, які закодовані у вхідних послідовностях через проблему градієнта, що зникає [20].

Ця проблема була вирішена шляхом використання спеціалізованої нейронної або клітинної структури в мережі довгострокової короткострокової пам'яті (LSTM), яка підтримує постійний зворотний потік сигналу помилки, який дозволяє LSTM вивчати довгострокові залежності.

Детальний огляд існуючих моделей та алгоритмів машинного навчання буде наведено у наступному розділі роботи.

1.6 Висновки до розділу 1

Виходячи з огляду літератури з теми дослідження, який наведений у розділі 1, можна встановити, що системи підтримки прийняття рішень - це

інтерактивні програмні системи, призначені для допомоги у прийнятті рішень шляхом доступу до великих обсягів інформації, що генерується з різних пов'язаних інформаційних систем. Доведено, що переваги СППР дозволяють застосовувати їх при аналізі фінансових процесів на фондовому ринку.

Ціноутворення на ринку цінних паперів можна описати за допомогою моделей, які відрізняються переважно своїми припущеннями, тобто єдина універсальна модель ціноутворення відсутня.

Фондові ринки динамічні та демонструють широкі варіації, тому прогнозування фондового ринку стає дуже складним завданням через вкрай нелінійну природу та складну розмірність. Прогнозування фінансового індексу характеризується інтенсивністю даних, шумом, нестационарністю, неструктурованим характером, високим ступенем невизначеності та прихованими залежностями.

Виконавши огляд існуючих методів прогнозування числових рядів встановлено, що застосування нейронних мереж для аналізу фінансових процесів є оптимальним.

Нейронні мережі для прогнозування фондового ринку мають низку таких переваг у порівнянні з розглянутими методами:

- відносна простота у використанні, так як нейронні мережі навчаються на прикладах. Користувач нейронної мережі підбирає представницькі дані, а потім запускає алгоритм навчання, який автоматично сприймає структуру даних.

- нейронні мережі привабливі з інтуїтивної точки зору, оскільки вони ґрунтуються на примітивній біологічній моделі нервових систем.

В межах написання магістерської дисертації пропонується розробка СППР, яка буде направлена на аналіз фінансових процесів фондового ринку. Для досягнення мети в межах наступних розділів необхідно виконати наступне:

1. Обрати оптимальну модель машинного навчання для прогнозування фінансових процесів на фондовому ринку.

2. Розглянути особливості перевірки на гетероскедастичність.

3. Розглянути показники якості оцінок прогнозів можливих втрат.
4. Розробити архітектуру та функціональну схему СППР.
5. Побудувати математичну модель фінансових процесів та дисперсію.
6. Виконати обчислювальні експерименти та зробити аналіз отриманих результатів.

РОЗДІЛ 2 ВИБІР ТА ОПИС МАТЕМАТИЧНИХ МОДЕЛЕЙ ДЛЯ ПРОГНОЗУВАННЯ ФІНАНСОВИХ ПРОЦЕСІВ ФОНДОВОГО РИНКУ

2.1 Огляд існуючих моделей та алгоритмів машинного навчання для прогнозування фінансових процесів на фондовому ринку та вибір оптимальної моделі

2.1.1 Визначення набору вхідних даних для прогнозування

Як вже зазначалося у першому розділі роботи, при прогнозуванні на фондових ринках виникає завдання у прогнозуванні часових рядів.

Основна мета методів прогнозування часових рядів у тому, щоб спрогнозувати майбутні значення ряду з урахуванням регулярного патерну, що присутній у минулих спостереженнях самого ряду.

Розв'язання задачі прогнозування часових рядів на фондових ринках складається з наступних етапів:

- вибір вхідних змінних;
- попередня обробка даних;
- вибір та вилучення ознак;
- навчання (вирішення задачі) з використанням моделі прогнозування;
- оцінка продуктивності запропонованої моделі.

Перший крок у процесі прогнозування фондового ринку складається з вибору вхідних характеристик, які мають бути змодельовані методами прогнозування, та вихідних даних, що підлягають прогнозуванню. У контексті прогнозування фондового ринку доступні різні фундаментальні та технічні вхідні змінні.

Вибір вхідних змінних є основним питанням при прогнозуванні фондового ринку, і рішення про те, які вхідні змінні слід використовувати, – непросте завдання.

Наступним кроком у процесі є попередня обробка даних, обраних на першому кроці. Дані попередньо обробляються, щоб підвищити передбачувальну здатність моделей.

Механізм попередньої обробки може використовуватися для видалення шуму, виявлення викидів, обробки пропущених значень та нормалізації даних.

На третьому етапі застосовуються різні методи вибору або отримання ознак, щоб отримати найкращі репрезентативні змінні вхідних даних для того, щоб зменшити розмірність вхідних даних і зменшити обчислювальну складність моделі.

На наступному етапі визначається методи прогнозування, які будуть використовуватися безпосередньо для вирішення задачі та безпосередньо розв'язання задачі.

Оцінка продуктивності останнього етапу включає вибір відповідних показників продуктивності і вимірювання точності моделі і прийняття правильного рішення.

Напрямок індексу фондового ринку відноситься до руху індексу цін або тенденції коливань індексу фондового ринку в майбутньому. Передбачення напряму - це практичне питання, яке сильно впливає на рішення фінансового трейдера про купівлю або продаж інструменту. Точний прогноз динаміки фондового індексу може допомогти інвесторам отримати можливості для прибутку на біржі. Отже, точне прогнозування динаміки індексу цін на фондових ринках може бути надзвичайно вигідним [21,22].

Перший етап прогнозування на фондовому ринку – це вибір вхідних змінних. У літературі різні автори використовували різну кількість вхідних змінних. Два найбільш поширені типи функцій, які широко використовуються для прогнозування фондового ринку, - це фундаментальні індикатори та

технічні індикатори. Наприклад, у роботі [23] використовується фундаментальний аналіз і згадується, що користувачі повинні проаналізувати різні фундаментальні фактори.

Технічний аналіз залежить від внутрішніх та зовнішніх фундаментальних атрибутів у сфері прогнозування ринку. На думку технічного аналітика, більшість фундаментальних факторів, що впливають на фондовий ринок, відображаються у самій ціні акцій. Технічний аналітик моделює ціни на акції як тимчасові ряди та намагається визначити майбутні моделі на основі минулих значень тимчасових рядів.

У роботі [24] проводиться дослідження технічних індикаторів, які визначаються шляхом застосування математичних формул до цін акцій, таких як ціна відкриття, ціна закриття, максимальна ціна та мінімальна ціна, та намагалися знайти майбутні ціни акцій.

У роботі [25] автори використовували різні технічні індикатори, а саме ковзну середню, індекс відносної сили, імпульс, швидкість зміни, балансовий обсяг, індикатори спрямованого руху серед інших факторів для прогнозування фондового ринку.

У роботі [26] автори використовували квартальні дані за 40 років з квітня 1972 р. по липень 2012 р., що містять 163 випадки для прогнозування, проте не отримали очікуваного результату. Також використовували денні дані за період з 22 січня 2003 року по 27 лютого 2006 року для індексу ТАІЕХ, що включає 781 екземпляр, та з 22 лютого 2004 року по 29 лютого 2008 року для індексу NIKKEI, що містить 1000 екземплярів.

У роботах [27] і [28] автори вибрали ціни відкриття, закриття, максимуму і мінімуму, а також обсяг акцій, що торгуються, як вхідні дані для моделі.

У роботі [29] автори використовували тренд Google з ціною відкриття, високою ціною, ціною закриття і мінімальною ціною і обсягом акцій, що торгуються для розробки моделі прогнозування.

Автори в роботі [30] застосували чотири фундаментальні показники, а саме коефіцієнт цінового прибутку (P/E), коефіцієнт виплати дивідендів, рентабельність власного капіталу (ROE) та балансову вартість для прогнозування австралійського фондового ринку.

Автори роботи [31] використовували 6 макроекономічних факторів, таких як індекс споживчих цін, процентні ставки за депозитами, обмінний курс долара США, індекс промислового виробництва, республіканська ціна продажу золота, процентні ставки за казначейськими векселями та ціна закриття 4 індексів, а саме DJI.

Автори роботи [32] використовували 27 економічних та фінансових факторів як вихідні дані і вибрали 20 факторів, які мають значний вплив на рух індексу S&P 500.

У роботі [33] використовується 7 фундаментальних змінних, що належать біржовому ринку, таких як загальний борг до загальних активів (TDTA), грошовий потік, поділений на загальні активи (CFTA), поточні зобов'язання, поділені на загальні активи (CLTA), оборотний капітал, поділений на загальні активи (WCTA), прибуток до сплати відсотків і податків, поділений на загальні активи (EBTA), і чистий прибуток, поділений на загальні активи (NITA), поточні активи, поділені на загальні активи (CATA), як вхідні дані для запропонованої моделі прогнозування шість років поспіль.

Автори роботи [34] використали 71 фінансовий та макроекономічний індикатор для розробки моделі прогнозування індексу Nikkei 225. Повний набір даних включає 237 щомісячних спостережень за період з листопада 1993 по липень 2013 року.

У роботі [35] автором використано такі технічні індикатори: проста 10-денна ковзна середня, зважена 10-денна ковзна середня, стохастичний K%, індекс відносної сили (RSI), стохастичний D%, дивергенція ковзної середньої (MA), осцилятор A/D (накопичення/розподіл).

Таким чином, існує безліч підходів до вибору оптимального набору змінних вхідних.

Набір вхідних змінних, які найчастіше використовувалися в досліджуваних роботах, наведено у вигляді рисунка 2.1.

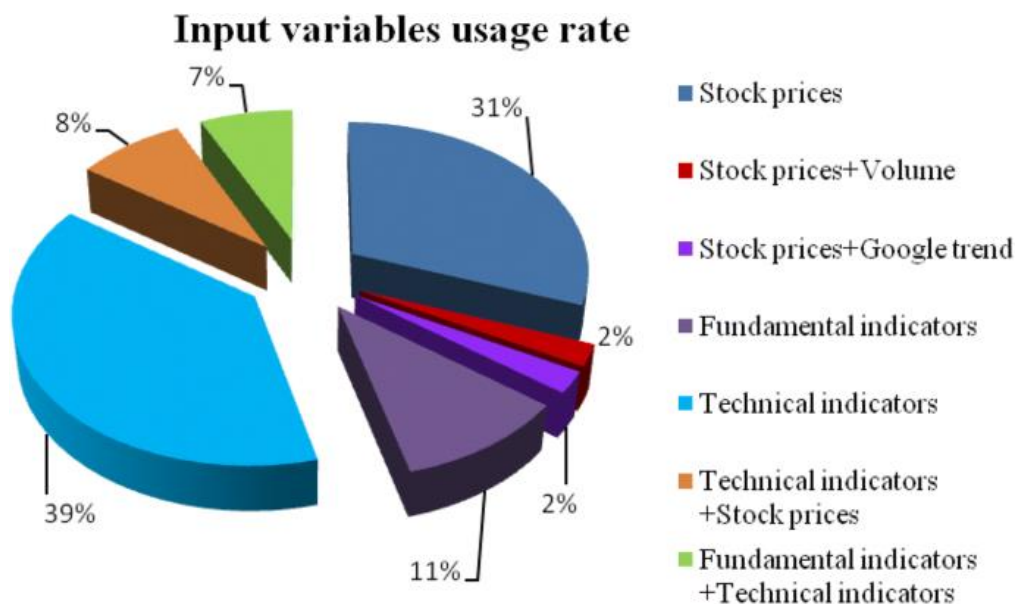


Рисунок 2.1 – Найуживаніший набір вхідних змінних для прогнозування фінансових показників на фондовому ринку

Другий етап – попередня обробка даних. На фондовому ринку якість прогнозування даних є основним фактором, оскільки точність та надійність моделі прогнозування залежать від якості даних.

Будь-які небажані аномалії у наборі даних відомі як шум. Викиди - це набір спостережень, який підпорядковується загальній поведінці набору даних

Наявність шуму та викидів може призвести до поганої точності прогнозів моделей прогнозування. Дані повинні бути підготовлені таким чином, щоб охопити діапазон входів, для яких буде використовуватись мережа. Методи попередньої обробки даних намагаються зменшити помилки та видалити викиди, тим самим підвищуючи точність моделей прогнозування. Одним із методів попередньої обробки даних є перетворення даних. Перетворення даних

із однієї шкали на іншу корисно у більшості евристичних підходів, особливо під час вирішення завдань прогнозування.

Третій етап - вибір та вилучення ознак.

2.1.2 Огляд моделей, методів і алгоритмів прогнозування, які застосовувались для прогнозування на фондових ринках

В першому розділі роботи було визначено, що застосування нейронних мереж є актуальним.

Розглянемо алгоритми прогнозування з застосуванням нейронних мереж.

У роботі [36] використовується модель Long short-term memory, яка підтримує постійний зворотний потік сигналу помилки, який, у свою чергу, дозволяє LSTM вивчати довгострокові залежності.

Як нещодавній засіб захисту в роботі [37] представлена архітектура LSTM, названа DLSTM, яка прагнула обійти обмеження традиційної моделі LSTM.

DLSTM включає складові шари LSTM один над іншим.

Підходи нечіткої логіки застосовувалися багатьма авторами із відносно високим ступенем успіху для моделювання та прогнозування часових рядів цін на акції.

У роботі [38] автори запропонували комбінований підхід шляхом інтеграції генетичної нечіткої системи (GFS) та штучної нейронної мережі (ШНМ) для створення інтелектуальної системи прогнозування цін на акції. Вони застосували покроковий регресійний аналіз для визначення ключових характеристик, які мають найбільший вплив на ціну акцій, і вибрані функції були розділені на кілька кластерів з використанням картки (SOM), що

самоорганізується. Нарешті, всі кластери використовуються як вхідні дані для GFS для отримання бази правил.

Результати показали, що запропонована модель показує кращу продуктивність порівняно з іншими підходами, такими як ARIMA та ANN.

Автори [39] розробили підхід, заснований на нечітких правилах типу Такагі - Сугено - Канга (TSK) для прогнозування фондового ринку. Вони застосували імітацію відпалу (SA) визначення набору найкращих параметрів нечіткої системи. Щоб виправдати точність прогнозування, вони показали, що система нечітких правил TSK працює краще, ніж нейронна мережа зворотного поширення (BPNN) та метод множинної регресії.

Автори роботи [40] впровадили систему нейро-нечіткого виводу, адаптувавши систему нечітких правил типу TSK для прогнозування фондового ринку, та застосували кластеризацію нечіткого C-середнього для визначення кількості правил.

Автори роботи [41] досліджували можливості адаптивної мережевої системи нечіткого виведення (ANFIS) для прогнозування доходності фондового ринку. Результати експериментів показують, що ANFIS може бути ефективно використаний на вирішення завдань прогнозування ринку.

У роботі [42] автори створили нечітку базу правил прогнозування рівня довіри до акцій (високий чи низький). Пропонована модель реалізована у трьох модулях. У першому модулі вони використовували алгоритми машинного навчання прогнозування ціни закриття акцій. У другому модулі вони набувають значення тональності заголовків останніх новин з кожної акції.

У третьому модулі виходи обох модулів використовуються як вхідні дані для модуля нечіткої логіки для створення бази нечітких правил.

Автор у роботі [43] застосував нечітку систему 2-го типу для прогнозування цін на акції, використовуючи як фундаментальні, так і технічні індикатори як вхідні дані для запропонованої моделі. Всі дослідженні параметри функції власності уточнюються генетичним алгоритмом.

У роботі [28], яка вже згадувалася раніше, автори досліджували переваги нейронечіткої системи 2-го типу для моделювання проблеми прогнозування цін на акції. Набір даних сегментується на кластери з використанням методу самоналаштування кластеризації, а потім до кожного кластера генерується база правил правила TSK типу 2. Параметри, пов'язані з нейронечіткою моделлю типу 2, налаштовуються за допомогою оптимізації рою частинок і оцінки методом найменших квадратів.

Автор [44] представив чотирирівневу нечітку багатоагентну систему (FMAS) для створення гібридної інтелектуальної системи, яка поєднує кілька інтелектуальних агентів для прогнозування ціни акцій наступного дня. Перший рівень використовується для збирання відповідної інформації про проблемну сферу з використанням експертних знань. Другий рівень присвячений добору ознак та формуванню кластерів. Роль третього рівня полягає у побудові моделі для всіх кластерів з використанням генетичної нечіткої системи та оптимізації побудованої моделі для вибору кращої нечіткої системи для кожного кластера. Метою четвертого рівня є аналіз моделі та уявлення знань.

У роботі [45] створено модель прогнозування, в якій використовується адаптивна нейронечітка система виведення (ANFIS) для прогнозування ціни закриття фондової біржі за 5 днів наперед індивідуально для кожного дня.

Крім цього, автор запропонував систему рекурентного нечіткого виводу (SERFIS), що саморозвивається, для підвищення здатності прогнозування традиційної нейро-нечіткої системи. Модель створюється з використанням нечіткої системи першого порядку типу Такагі - Сугено - Канга (TSK) з двома варіантами підключення зі зворотним зв'язком, тобто шляхом забезпечення сили спрацьовування нечіткого правила та введення петель зворотного зв'язку з тимчасовою затримкою у вихідному шарі.

У роботі [46] автори запропонували модифікований пошук диференціальної гармонії (MDHS) для оптимізації циклу зворотного зв'язку, попередніх та наступних параметрів запропонованого підходу. Щоб порівняти

результат запропонованої моделі, у статті також представлено іншу мережу, а саме штучну нейронну мережу з рекурентними функціональними зв'язками (RCEFLANN).

Автор також представив нову модель прогнозування цін на акції, використовуючи адаптивну систему нейронечіткого виведення (ANFIS) та методи інтегрованого нелінійного вибору ознак (INFS). У цьому дослідженні використовувався метод INFS для вибору ключових технічних індикаторів, які використовуються як вхідні дані для моделі прогнозування ANFIS для отримання первинного прогнозованого значення, і, нарешті, модель адаптивних очікувань, яка використовується для подальшого підвищення продуктивності моделі прогнозування.

У роботі [47] автори спочатку використовували логіку диференціальних нечітких тимчасових рядів для прогнозування тенденцій даних фондового ринку, а потім представили новий еволюційний метод, а саме імперіалістичний конкурентний алгоритм (ICA), для подальшого підвищення точності прогнозів вихідної моделі шляхом оптимізації параметрів моделі.

Крім цього, автор представив нечітку модель короткострокової торгівлі з використанням нечіткої системи Мамдані та суміші технічних індикаторів, що часто використовуються і рідко використовуються для генерації торгових сигналів на фондовому ринку.

У роботі [34], яка раніше вже згадувалася, автор представив інтервальну систему нечіткої логіки типу 2 (IT2FLS) для прогнозування фондових індексів на основі нечітких часових рядів та нечіткої логічної карти відносин (FLRM). Запропонована модель працює у п'ять етапів. Перший крок включає пошук варіації у часовому ряду. Другий крок включає фазифікацію часових рядів та визначення нечіткої множини. На третьому кроці визначаються вхідні інтервальні нечіткі множини типу 2 (IT2FS) та вихідні інтервали інтервальної нечіткої логічної системи типу 2 (IT2FLS). Четвертий крок включає створення

нечітких логічних відносин (FLR) та нечіткої логічної карти відносин (FLRM) і, нарешті, передбачення майбутніх значень з застосуванням ШНМ.

У роботі [38] автор запропонував новий підхід, заснований на нечіткій логіці та комбінації графа видимості та передбачення зв'язків для підвищення точності завдання передбачення часових рядів з застосуванням ШНМ. У запропонованому методі часові ряди спочатку перетворюються на граф видимості, потім застосовується метод прогнозування зв'язку для отримання початкового прогнозу і, нарешті, використовується нечітка логіка для подальшого підвищення точності прогнозування шляхом визначення нечітких правил на основі взаємозв'язку між історичними даними. На основі порівняння попередніх досліджень автори показали, що запропонований метод має кращу передбачуваність.

Автор також запропонував використати модель нечітких часових рядів та індукцію грубого набору правил для прогнозування фондових індексів.

У роботі [39] автори застосували алгоритм індукції грубого набору правил, саме LEM2 (Learning from Example Module version 2), щоб витягти правила прогнозування продажів чи покупок з часових рядів для одержання початкового прогнозу. Після цього модель адаптивного очікування використовується для підвищення ефективності початкового прогнозу. Модель адаптивного очікування – це розумна модель прогнозу під час прогнозування часових рядів. При прогнозуванні часових рядів цін на акції з використанням моделі адаптивного очікування майбутні ціни на акції генеруються на основі минулого періоду ціни акцій та помилки прогнозування для одного останнього періоду.

Автори роботи [48] запропонували парадигму, що ґрунтується на генетичному алгоритмі (ГА), для побудови моделі класифікації, яка може знаходити торгові правила на основі технічних індикаторів. Вони використовували ГА для оптимізації початкових ваг нейронних мереж та алгоритм зворотного розповсюдження помилки Левенберга - Марквардта

(LMBP) для навчання нейронної мережі з прямим зв'язком. Результат показує, що запропонована модель прогнозування здатна справлятися з коливаннями ринку і дає хороші результати прогнозування.

Автори роботи [49] побудували асоціативний класифікатор, запропонувавши підхід на основі генетичного алгоритму (ГА) для генерації сигналів продажу та купівлі. Запропонована модель визначає правила торгівлі за технічними індикаторами.

Крім цього, автори розробили гібридну модель прогнозування, об'єднавши інтегровану авторегресійну ковзну середню (ARIMA) з генетичним програмуванням (GP) для прогнозування нелінійних часових рядів. ARIMA використовується для обробки лінійної частини часового ряду, а GP використовується для обробки нелінійної частини часового ряду для підвищення точності. Щоб перевірити точність запропонованої гібридної моделі, автори використовували три часові ряди і показали, що гібридна модель може найкраще використовуватися для завдань прогнозування часових рядів.

Автори роботи [33] розробили нову модель, використовуючи генетичний алгоритм налаштування ваги з'єднання частково пов'язаної нейронної мережі для прогнозування тенденції на фондовому ринку. Вони розробили парадигму для проблеми вибору акцій із використанням регресії опорних векторів (SVR) та генетичного алгоритму (GA). У цьому дослідженні SVR використовується для прогнозування прибутковості збору акцій, а створення портфеля вибираються найбільш прибуткові акції. GA використовується для отримання найкращих параметрів SVR.

У гібридних методах прогнозування об'єднання основних підходів CI стало корисною процедурою підвищення точності моделі прогнозування шляхом об'єднання переваг окремих підходів і усунення недоліків певних окремих методів.

Автор роботи [48] говорив про те, що гібридні алгоритми є більш ефективними у підвищенні точності.

Розглянемо різні статті, що базуються на гібридизації штучної нейронної мережі з іншими методами.

У роботі [50] автор запропонував гібридну мультиагентну систему нейронної мережі кажанів (BNNMAS) для прогнозування курсу акцій. Запропонована модель працює у чотирирівневому мультиагентному сценарії. У першому шарі виконується завдання збору та попередньої обробки даних. У другому шарі метод взаємної кореляції та вибір лага використовувалися для вибору відповідної функції та кращого тимчасового лага. З 20 фундаментальних показників та технічних індикаторів для завдання прогнозування вибрано 13 важливих характеристик. У третьому рівні алгоритм кажана (BA) використовувався як алгоритм навчання трирівневої нейронної мережі, а радіальна базова функція (RBF) - як функція активації для прихованого шару, а четвертий агент використовується для перевірконого тестування. Продуктивність BNNMAS порівнювалася з іншими методами,

Статистика середньої абсолютної процентної помилки (MAPE) показує перевагу BNNMAS над іншими методами.

Автор роботи [51] запропонував підходи двоетапного синтезу. На початковому етапі, підтримка вектора регресії (CBP) використовуються для прогнозування майбутнього значення технічних індикаторів.

Штучна нейронна мережа (ІНС), SVR та випадковий ліс (RF) використовуються на останньому етапі об'єднання, внаслідок чого створюються моделі прогнозування об'єднання SVR – ANN, SVR – SVR та SVR – RF для прогнозування майбутнього значення ціни закриття. Точність прогнозування запропонованих гібридних моделей порівнювалася з одноетапними сценаріями, в яких ІНС, SVR та RF використовуються окремо. Результати експериментів показують, що моделі двоетапного прогнозування перевершують моделі одноетапного прогнозування.

Крім цього автор роботи [51] представив динамічну нейронну мережу на основі нескінченного імпульсного відгуку (IIR) для прогнозування індексу цін

на акції від 1 до 30 днів вперед для чотирьох різних акцій, а саме Бомбейської фондової біржі (BSE), акцій IBM, акцій RIL та акцій корпорації Oracle. Для навчання DNN використовувалися чотири різні алгоритми навчання, такі як алгоритм рекурентного навчання в реальному часі (RTRL), диференціальна оцінка (DE), фільтр Калмана (UKF) та гібридний підхід DEUKF.

За результатами зроблено висновок, що DEUKF показує високі показники за всіма індексами фондового ринку порівняно з DE, UKF та RTRL. Ефективність DNN та DEUKF порівнювалася з іншими модифікованими методами нейронних мереж, такими як локальна лінійна нейронна мережа вейвлетів (LLWNN), локальна лінійна RBFNN (LLRBFNN) та Laguerre FLANN.

Потім автори застосували функціональну нейронну мережу з радіальним базисом (RBFNN) для навчання історичних даних фондового ринку Шанхайської фондової біржі та використали алгоритм штучного рою риб (AFSA) для оптимізації параметрів RBFNN. Для підвищення точності передбачення вони використовували AFSA для оптимізації ширини та ваги центру алгоритму кластеризації K-середніх під час навчання RBFNN. Для того, щоб продемонструвати ефективність запропонованої моделі, автори порівняли результати прогнозування з RBFNN та алгоритмом ARIMA. Результати показують, що запропонована модель перевершує інші моделі.

Автори роботи [52] розробили гібридний метод, що складається з оптимізації рою частинок (PSO) та ансамблевої нейронної мережі з нечіткою агрегацією для прогнозування складних часових рядів. PSO використовується для налаштування параметрів нейронних мереж та визначення оптимальної кількості окремих нейронних мереж, що використовуються для створення ансамблевої нейронної мережі.

Автори використовували нечіткий тип 2 для об'єднання вихідних даних окремих нейронних мереж, що утворюють ансамбль.

Потім автори виконали порівняння алгоритму диференціальної еволюції (DE) та оптимізації рою частинок (PSO) за оптимізації параметрів нейронної

мережі прямого поширення (FFNN), яка використовується для прогнозування щоденних цін на акції. Порівняння двох методів оптимізації було проведено на основі точності прогнозування, швидкості збіжності та здатності узагальнення. Результат цього дослідження показує, що обидва методи оптимізації дозволяють уникнути проблеми локальних мінімумів.

Метод DE сходиться до глобальних мінімумів швидше, ніж PSO, і DE працює краще, ніж PSO.

Автор роботи [52] також інтегрував нейронну мережу зворотного поширення (BPNN) та покращену оптимізацію бактеріального хемотаксису (IBCO), щоб створити модель прогнозування, яка може ефективно передбачати фондовий індекс індексу S&P 500. Запропонована модель використовувалася як для короткострокового (наступного дня), так і та для довгострокового (15 днів) прогнозування фондових індексів.

Середньоквадратична помилка, обчислена в результаті експерименту, показує, що модель BP-IBCO перевершує традиційну модель BPNN.

Традиційні методи статистичного прогнозування, такі як авторегресійне ковзне середнє (ARMA), експоненційне згладжування (ES), авторегресійне інтегроване ковзне середнє (ARIMA), авторегресійна умовна гетероскедастичність (ARCH) та узагальнена авторегресійна (GARCH) також є ефективними [53]. Ці моделі засновані на припущенні, що фінансові тимчасові ряди, що досліджуються, генеруються лінійним процесом, і намагаються змодельовати процес тимчасових рядів, щоб спрогнозувати майбутню цінність ряду.

Виходячи з проведеного огляду, остаточно прийнято рішення зупинитися на нейронній мережі LSTM для прогнозування ціни на фондовому ринку, оскільки вона показала найвищу ефективність при прогнозуванні цін на фондовому ринку. В якості методів моделювання дисперсії будемо використовувати ARCH/GARCH методологію, яка також є найбільш ефективною серед аналогів, виходячи з роботи [53, 54].

2.1.3 Опис обраних моделей

Як зазначається в [54], загалом, ARCH/GARCH методологію можна охарактеризувати як методологію моделювання дисперсії досліджуваного показника.

Оскільки дисперсія – момент другого порядку, модель дисперсії нелінійна, отже, її не можна оцінювати методами, розробленими для лінійних моделей, зокрема за допомогою ARIMA моделей.

Абревіатура ARCH перекладається як авторегресійна умовна гетероскедастичність, тобто зміна в часі дисперсії випадкових величин (збурень), які формалізовано можна представити таким чином:

$$a_{t+1} = y_{t+1} - \mu_y(y_1, y_2, y_3, \dots, y_t),$$

де a_{t+1} – збурення (випадкова величина) в (t+1) період часу;

y_{t+1} – значення показника в (t+1) період часу;

$\mu_y(y_1, y_2, y_3, \dots, y_t)$ – середнє значення показника, оцінене на основі даних, які передують моменту (t + 1).

Також в [54] зазначається, що абревіатура GARCH перекладається як узагальнена авторегресійна модель з умовною гетероскедастичністю.

У GARCH моделях на відміну від ARCH моделей у рівнянні дисперсії враховано, крім лагових змінних випадкових величин, ще й лагові змінні умовної дисперсії.

Оскільки ARCH/GARCH вважаються моделями дисперсії, середнє значення показника можна моделювати на основі різних підходів, зокрема, простим середнім, лінійною регресійною моделлю, ARIMA, SARIMA моделями тощо.

Власне загальну ARCH/ GARCH модель можемо представити як послідовність фільтрів, яка представлена на рисунку 2.2 [54].

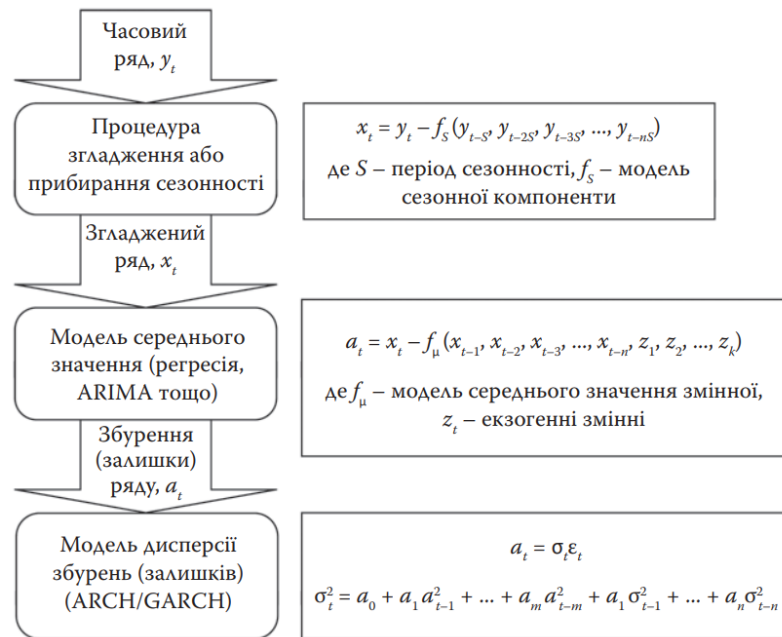


Рисунок 2.2 – Узагальнений алгоритм послідовної реалізації ARCH/GARCH моделі

Загалом, ARCH/GARCH модель можна вважати «доповненням» як до одно-, багатofакторної лінійної регресії, так і до ARIMA моделей. Однак, оцінювання ARCH/GARCH моделей неможливо здійснити методом найменших квадратів (МНК). Для їх оцінювання розроблено спеціальні процедури, зокрема метод максимальної вірогідності [54].

Мережа LSTM - це тип рекурентної нейронної мережі (RNN), яка може вивчати довгострокові залежності між тимчасовими кроками даних послідовності [55].

Основними компонентами мережі LSTM є рівень введення послідовності та рівень LSTM. Шар введення послідовності вводить дані послідовності або часового ряду в мережу. Шар LSTM вивчає довгострокові залежності між тимчасовими кроками даних послідовності.

Рисунки 2.3, 2.4 ілюструють архітектуру простої LSTM мережі для класифікації. Мережа починається з шару введення послідовності, за яким слідує шар LSTM. Щоб передбачити мітки класів, мережа закінчується повним шаром, шаром softmax і вихідним шаром класифікації.



Рисунок 2.3 – Архітектура LSTM мережі для класифікації [55]

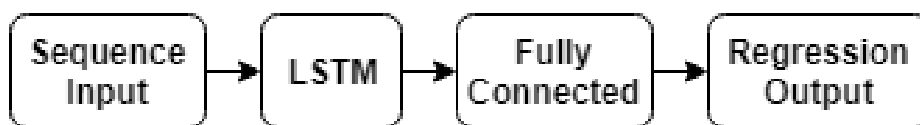


Рисунок 2.4 – Архітектура простої LSTM мережі для регресії

Функція активації:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f).$$

Мережа починається з шару введення послідовності, за яким слідує шар LSTM. Мережа закінчується повністю підключеним шаром та вихідним шаром регресії.

2.2 Особливості перевірки на гетероскедастичність

Поняття гетероскедастичності, протилежне гомоскедастичності, використовується в статистиці, особливо в контексті лінійної регресії або для аналізу часових рядів, для опису випадку, коли дисперсія помилок моделі не однакова для всіх спостережень, хоча часто одне з основних припущень у моделюванні полягає в тому, що дисперсії і що помилки моделі однаково розподілені.

Рисунок 2.5 відображає одну з типових ситуацій на фондовому ринку, а саме кластеризацію волатильності, що свідчить про гетероскедастичність умовної дисперсії на фондовому ринку [55].

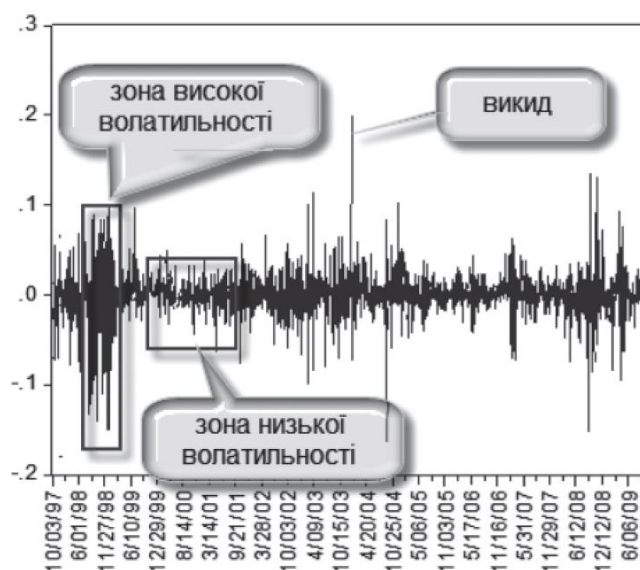


Рисунок 2.5 – Свідчення гетероскедастичності умовної дисперсії на фондовому ринку [55]

У лінійному регресійному аналізі той факт, що помилки моделі (також звані залишками) не є гомоскедастичними, призводить до того, що коефіцієнти моделі, що оцінюються з використанням звичайного методу найменших квадратів, не є ні об'єктивними, ні коефіцієнтами з мінімальною дисперсією. Оцінка їхньої дисперсії ненадійна.

Якщо передбачається, що дисперсії неоднорідні (подання залишків по відношенню до незалежних змінних може виявити гетероскедастичність), тоді необхідно виконати тест на гетероскедастичність.

Було розроблено кілька тестів з наступними нульовими та альтернативними гіпотезами [56]:

H_0 : залишки гомоскедастичні;

H_a : залишки гетероскедастичні.

Розглянемо декілька тестів, які широко розповсюджені в нинішній час для перевірки на гетероскедастичність.

1. Тест Бреуша-Пагана.

Цей тест на гетероскедастичність був розроблений Бреушем і Паганом, а потім покращений Кенкером – тому цей тест іноді називають тестом Бреуша-Пагана та Кенкера – для виявлення випадків гетероскедастичності, які роблять класичну оцінку параметрів лінійної регресії ненадійною. Якщо e є вектором помилок моделі, нульова гіпотеза H_0 , нульова гіпотеза може бути записана:

$$H_0: E(e^2) = \sigma^2.$$

Щоб переконатися, що квадратичні помилки не залежать від незалежних змінних, які можуть бути перетворені на багато функціональних форм, найпростішим є регресія квадратів помилок щодо незалежних змінних. Якщо дані гомоскедастичні, то коефіцієнт детермінації R^2 не повинен дорівнювати 0.

Якщо H_0 не відкидається, ми можемо зробити висновок, що гетероскедастичність, якщо вона існує, не набуває функціональної форми, що використовується.

Практика показує, що гетероскедастичність не є проблемою, якщо H_0 не приймається. Якщо H_0 відкидається, ймовірно, має місце гетероскедастичність і вона набуває описаної вище функціональної форми.

Формула, що використовується для тесту, яка запропонована Кенкером:

$$LM = nR,$$

де LM - множник Лагранжа.

Перевага цієї формули полягає в тому, що вона асимптотично слідує розподілу χ^2 -квадрат з p ступенями свободи, де p - кількість змінних.

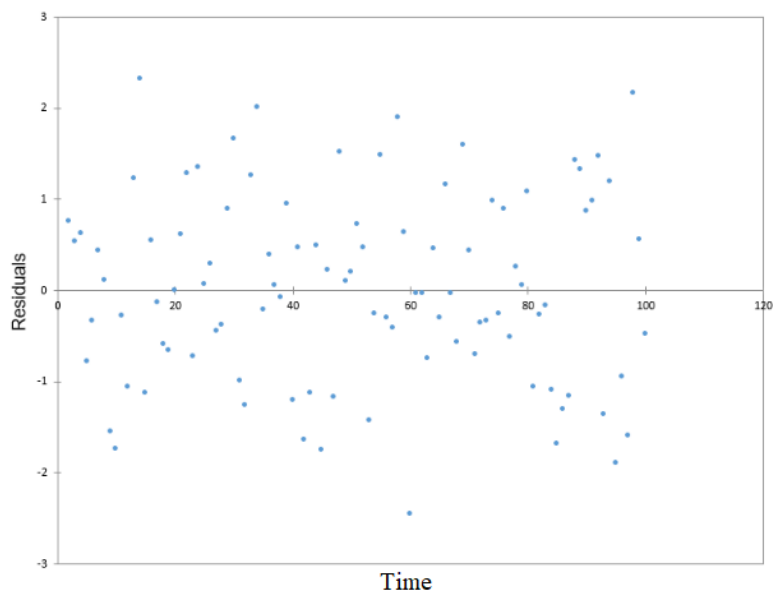
Якщо нульова гіпотеза відкинута, необхідно буде перетворити дані перед виконанням регресії або використовувати методи моделювання, щоб врахувати мінливість дисперсії [56].

2. Тест Уайта та модифікований тест Уайта (Вулдрідж). Цей тест було розроблено Уайтом для виявлення випадків гетероскедастичності, що роблять класичні оцінки параметрів лінійної регресії ненадійними. Ідея аналогічна ідеї Бреуша та Пегана, але вона спирається на більш слабкі припущення щодо форми, яку набуває гетероскедастичність. Це призводить до регресії квадратичних помилок щодо незалежних змінних, а також квадратів і перехресних творів останніх. Статистика, що використовується, така сама, як і в тесті Бреуша-Пагана, але через присутність набагато більшої кількості регресорів тут застосовуються $2p + p \cdot (p-1)/2$ ступеня свободи для розподілу χ^2 -квадрат.

Щоб не втратити надто багато ступенів свободи, Вулдрідж запропонував регресувати квадрати помилок за прогнозами моделі та їх квадратом. Це зменшує кількість ступенів свободи χ^2 -квадрату до 2.

Наразі існують інструменти, за допомогою яких можна перевірити гетероскедастичність. Одним із таких інструментів є XLSTAT.

Приклад результату тесту гетероскедастичності представлений на рисунку 2.6 [57].



Breusch-Pagan test:

LM (Observed value)	0.980
LM (Critical value)	3.841
DF	1
p-value (Two-tailed)	0.322
alpha	0.05

Test interpretation:

H0: Residuals are homoscedastic

Ha: Residuals are heteroscedastic

As the computed p-value is greater than the significance level $\alpha=0.05$, one cannot reject the null hypothesis H0.

Рисунок 2.6 - Приклад результату тесту гетероскедастичності за допомогою інструменту XLSTAT [57]

Оцінка ринкового ризику за методикою VaR буде наведена далі.

2.3 Оцінка ринкового ризику за методикою VaR

Однією з найбільш традиційних моделей ефективності, що використовуються при роботі з фондовими ринками, є вартість під ризиком (VaR) - економетричний інструмент, що визначається як значення, при якому існує можливість максимальних збитків протягом наступних днів [58].

З моменту його створення було створено декілька моделей з метою покращення прогностичної здатності даного методу. Основне розходження між методами VaR полягає у висновку про розподіл прибутку. Параметричні моделі припускають, що функція густини факторів ризику прибутковості активів повинна відповідати нормальному розподілу. Непараметричні моделі не вимагають будь-яких статистичних припущень, крім стаціонарного розподілу доходності, оскільки причина нормальності не відбиває ринкову реальність [58].

VaR відрізняється від інших показників тим, що його метою є уявлення ймовірності можливих змін вартості портфеля. Це сукупна міра ризику за всіма факторами ризику, що дає гарне уявлення про «апетит до ризику» інвесторів фондового ринку, враховуючи, що він оцінює найбільшу втрату за певного рівня впевненості в межах тимчасового горизонту.

Методика VaR представляє три важливі елементи у своєму застосуванні: конкретний рівень величини збитку, фіксований період, протягом якого вимірюється ризик, та довірчий інтервал.

Хоча спочатку метод був створений для вимірювання ринкового ризику, було розроблено кілька моделей VaR, що зробило його універсальним показником, що використовується для різних фінансових та нефінансових організацій, що піддаються ризику.

Перевага методу VaR полягає головним чином у створенні «спільного знаменника», що дозволяє порівнювати різні види ризикової діяльності на різних економічних ринках.

Розглянемо методику оцінки з застосуванням раніше обраної моделі GARCH.

Модель GARCH приймається як повна метрика ризику в аспекті урахування тієї властивості, що умовна волатильність є функцією безперервної зміни квадратів її попередніх значень, що породжує кластери волатильності. Модель є авторегресивною, оскільки значення прибутковості залежить від

значень, які припускають, що гетероскедастичність, що спостерігається в різні періоди, може бути автокорельована [58].

Для отримання VaR за методом GARCH спочатку необхідно змоделювати умовну волатильність:

$$\sigma_t^2 = \omega + \sum_{i=1}^{L_1} \alpha_i Y_{t-i}^2 + \sum_{i=1}^{L_2} \beta_i \sigma_{t-i}^2,$$

де L - кількість лагів;

$\omega, \alpha_i, \beta_i$ - параметри, що оцінюються за максимальною правдоподібністю.

Модель також передбачає розрахунок нормальності доходності:

$$Z_t \sim N(0,1).$$

Таким чином, прибутковість активів:

$$Y_c \sim N(0, \sigma_t^2).$$

Після оцінки волатильності за допомогою GARCH, VaR розраховується як добуток оціненої умовної волатильності та відсотка нормального розподілу відповідно до наступного виразу:

$$VaR = Z_t \cdot \sigma_t.$$

В роботі [58] також вказується, що GARCH – це найбільш ефективна модель оцінки, яка забезпечує коректний розрахунок VaR при застосуванні на фондових ринках.

2.4 Показники якості оцінок прогнозів можливих втрат

Створення коректного прогнозу є важливим аспектом для застосування тієї чи іншої моделі при аналізі фінансових процесів на фондовому ринку та має вирішальне значення. Важливо оцінити точність прогнозу, використовуючи справжні прогнози. Отже, розмір залишків не є надійним показником того, наскільки більшими можуть бути справжні помилки прогнозу. Точність прогнозів можна визначити тільки з урахуванням того, наскільки добре модель працює з новими даними, які не використовувалися при припасуванні моделі.

Для оцінок якості прогнозів використовують такі характеристики: надійність, точність, достовірність, похибки прогнозів, помилка прогнозу, коефіцієнт детермінації.

Поняття «надійність» передає інформацію про те, чи може користувач покладатися на прогноз щодо використання його для прийняття рішень.

Точність прогнозу — те, наскільки точний прогноз. Точність прогнозу може бути розрахована за формулою:

$$FA = 1 - \left(\frac{y - \bar{y}}{y} \cdot 100 \right),$$

де y – фактичне значення величини;

\bar{y} – прогнозоване значення величини.

Достовірність прогнозів – це ймовірність справдження даного прогнозу для заданого варіанта або довірчого інтервалу.

Серед похибок варто виділити наступні:

1. Абсолютна похибка моделі: різниця між справжнім та прогнозним значенням показника, що отримане за допомогою моделі прогнозування.

2. Середньоквадратична похибка, яка виражається за формулою:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{n}},$$

де n – розмір вибірки.

Коефіцієнт детермінації - доля дисперсії залежної змінної, що пояснюється використаною моделлю.

Визначається за формулою:

$$R^2 = 1 - \frac{SS_{residuals}}{SS_{total}},$$

де $SS_{residuals}$ – сума квадратів залишків регресії, яка обчислюється за формулою:

$$SS_{residuals} = \sum_{i=1}^n (y_i - \bar{y}_i)^2,$$

$$SS_{total} = \sum_{i=1}^n (y_i - \underline{y})^2,$$

$$\underline{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

Помилка прогнозу — це різниця між значенням, що спостерігається, і його прогнозом.

2.5 Висновки до розділу 2

Враховуючи складність фінансового ринку та нелінійну характеристику фінансових змінних, штучна нейронна мережа як важлива форма штучного

інтелекту стала великим досягненням у багатьох фінансових галузях та може застосовуватися для аналізу фінансових процесів на фондовому ринку.

В літературі найбільшу ефективність при прогнозуванні цін на фондовому ринку продемонстрували нейронні мережі LSTM, саме тому вони обрані для моделювання в межах магістерської дисертації. Окрім цього, для моделювання дисперсії при аналізі фінансових показників на фондовому ринку обрано методологію ARCH/GARCH, яка є найбільш ефективною серед аналогів.

Найбільш розповсюдженими тестами, які застосовуються для перевірки на гетероскедастичність, є тест Бреуша-Пагана і тест Уайта. Нині існують онлайн-інструменти, за допомогою яких можна виконати перевірку на гетероскедастичність.

З літературних джерел встановлено, що модель GARCH – найбільш ефективна модель, яка забезпечує коректний розрахунок VaR при застосуванні на фондових ринках.

Побудову СППР та виконання обчислювальних експериментів з обраними в межах даного розділу магістерської дисертації рішеннями буде наведено у наступному розділі магістерської дисертації.

РОЗДІЛ 3 ПОБУДОВА СППР ТА ВИКОНАННЯ ОБЧИСЛЮВАЛЬНИХ ЕКСПЕРИМЕНТІВ

3.1 Архітектура та функціональна схема СППР

У процесі виконання роботи була розроблена і реалізована система підтримки прийняття рішень, яка призначена для аналізу ціноутворення на фондовому ринку. Вона виконує наступні задачі: побудова моделей гетероскедастичних процесів для прогнозування волатильності, оцінювання ринкових ризиків методом VaR та аналіз якості отриманої прогнозованої моделі. Головною метою створення даної програми було підвищення рівня ефективності прийняття рішень на фондовому ринку, а також проілюструвати роботу методів моделювання фінансових процесів, застосувати їх на реальних даних. Схему роботи алгоритму схематично зображено на рис.3.1.

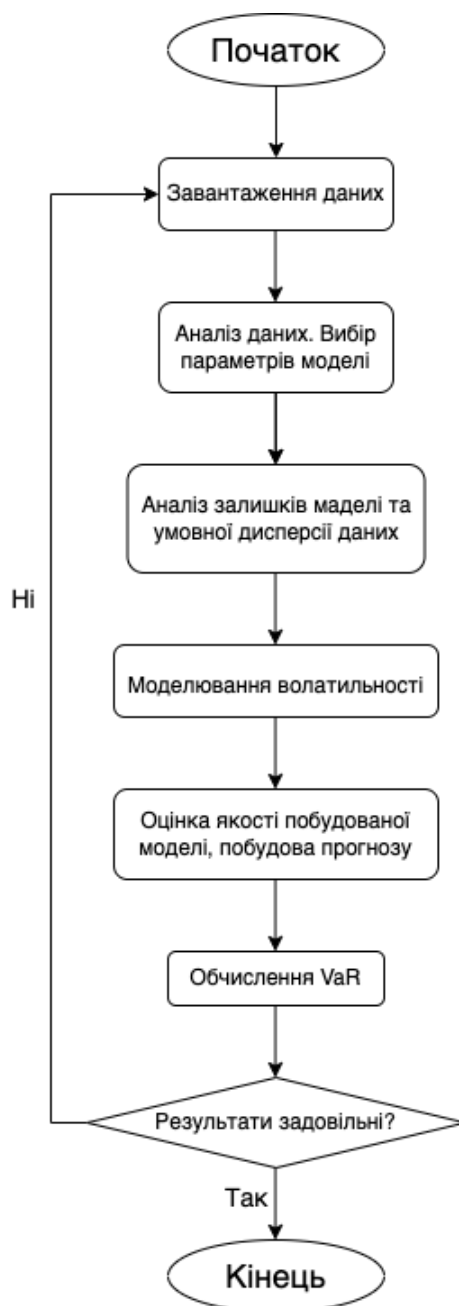


Рисунок 3.1 – Блок-схема роботи СППР

3.2 Аналіз вхідних даних

Для моделювання у даній роботі було обрано процес ціноутворення акцій міжнародної компанії TESLA, взятий з інтернет ресурсу finance.yahoo.com. Обраний датасет складається з наступних колонок: «High», «Low», «Open»,

«Close», «Volume», «Adj Close» та містить 1259 записів, а саме з 13.11.2017 – 11.11.2022. Для подальшої роботи потрібно обрати характеристику, яка буде найбільш коректною для моделювання. Після дослідження значення кожної колонки було визначено взяти саме колонку «Adj Close», адже вона показує скориговану ціну закриття, тобто ціну закриття після всіх застосованих коригувань. Двома основними з них є розбиття та дивіденди. Отже, проведемо візуальний аналіз даних, спочатку зобразимо графік даного часового ряду (рис.3.2).



Рисунок 3.2 – Графік значень цін акцій компанії TESLA

Побудуємо гістограму отриманого ряду(рис.3.3):

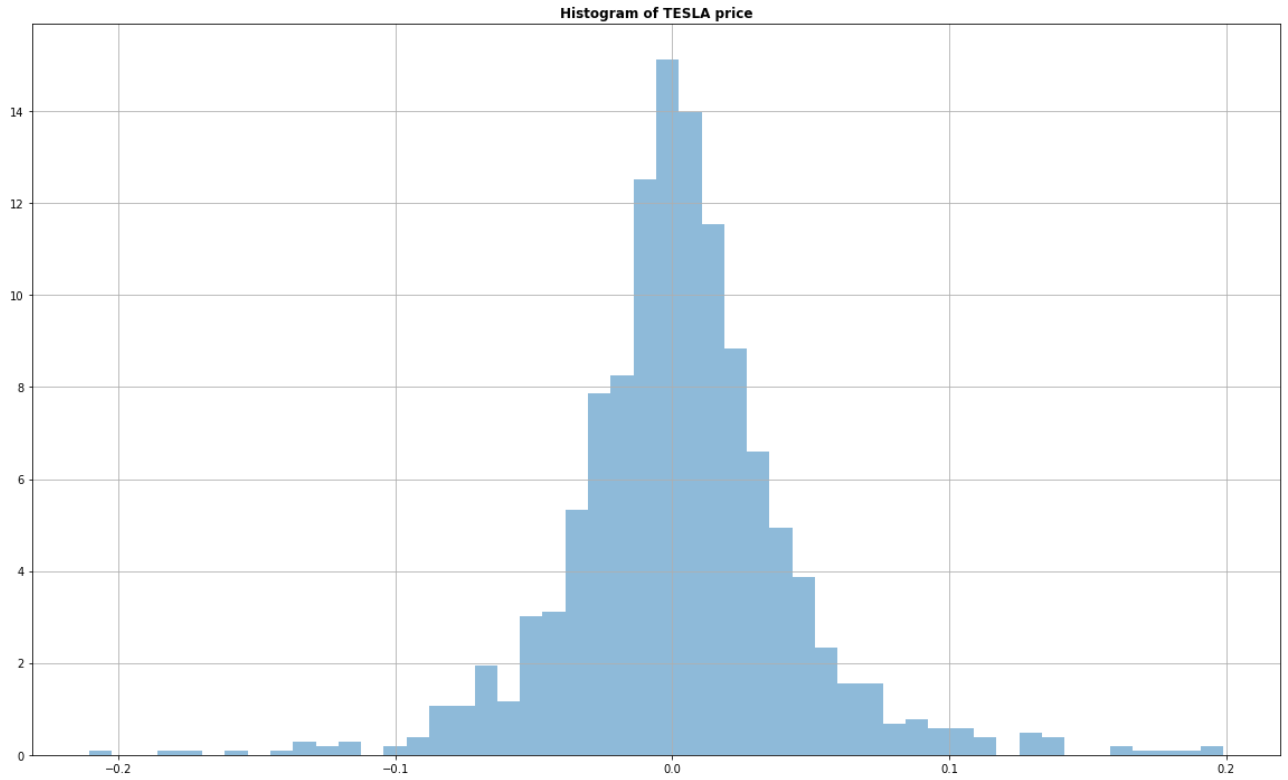


Рисунок 3.3 – Гістограма даних компанії TESLA

Тепер варто проаналізувати графіки щільності нормального розподілу та розподілу Стюдента- t на рисунках 3.4, 3.5:

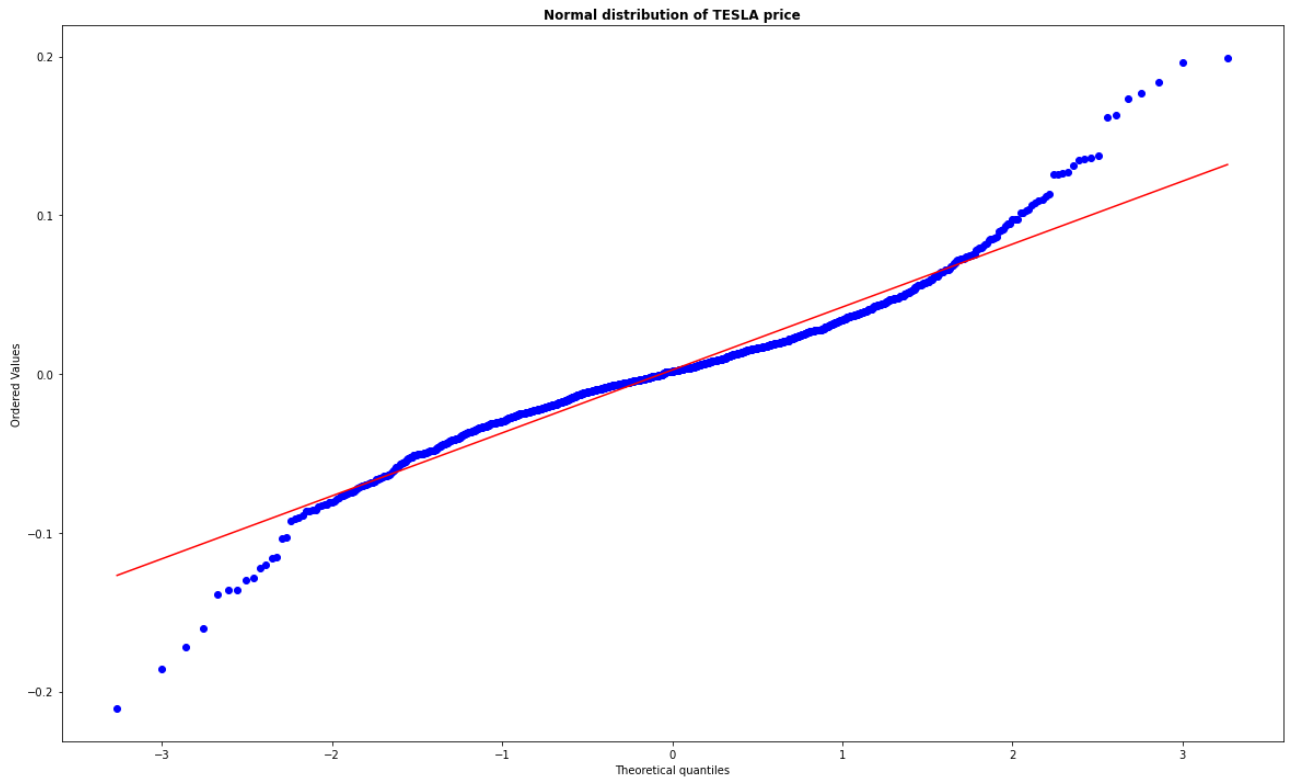


Рисунок 3.4 – Графік нормального розподілу компанії TESLA

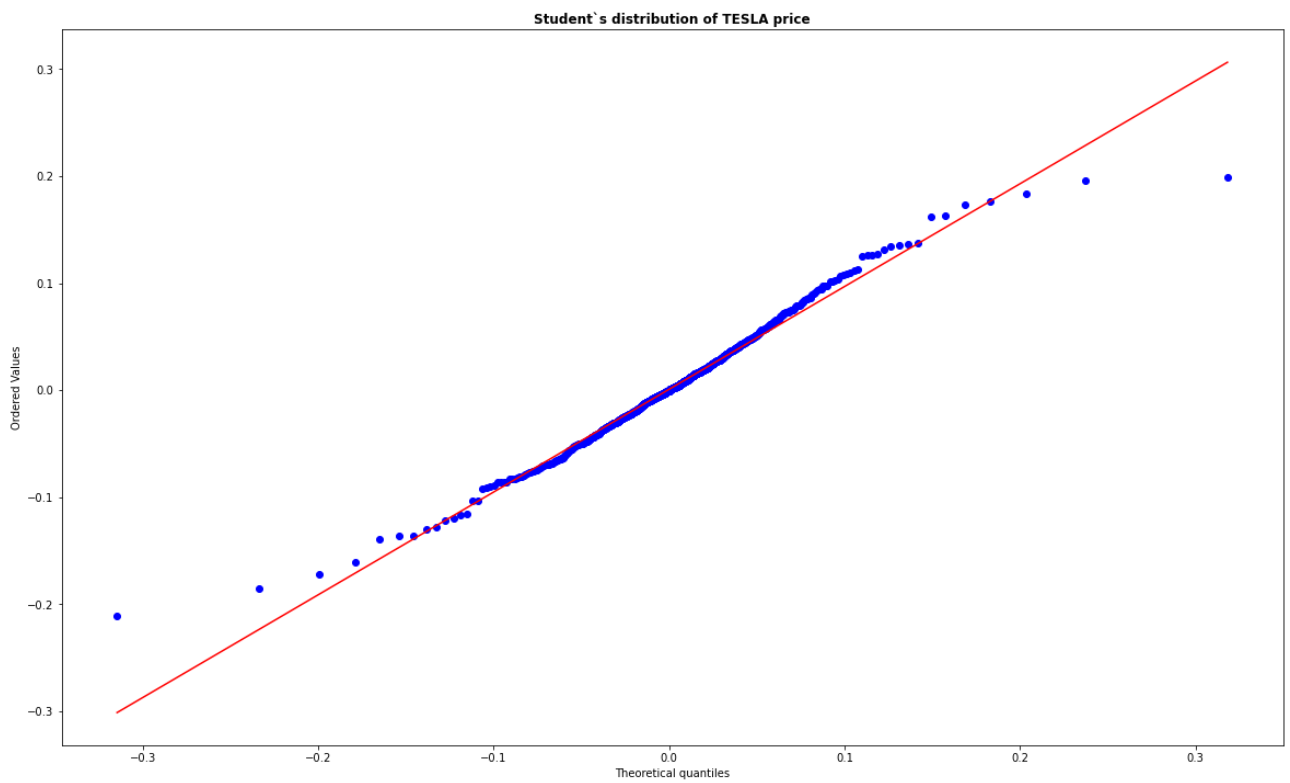


Рисунок 3.5 – Графік розподілу Стьюдента компанії TESLA

З графіків бачимо, що розподіл цін схожий на нормальний, проте являється трохи вищим, а от розподіл Стюдента краще описує ряд, особливо на хвостах розподілу.

3.3 Побудова математичних моделей фінансових процесів та дисперсії

Для побудови відповідних математичних моделей підготуємо наші дані. Спочатку розбиваємо вибірку на тестову й тренувальну, для очистки ряду від випадкових шумів застосовуємо фільтр Калмана, а також проведемо тест на гетероскедастичність(рис.3.6).

```

=====
Dep. Variable:          Adj_Close    R-squared:              0.000
Mean Model:           Constant Mean  Adj. R-squared:         0.000
Vol Model:            GARCH          Log-Likelihood:        -5505.34
Distribution:         Normal        AIC:                   11018.7
Method:              Maximum Likelihood  BIC:                   11039.1
                                         No. Observations:     1208
Date:                Wed, Nov 16 2022  Df Residuals:          1207
Time:                14:05:47        Df Model:               1
                                         Mean Model
=====
              coef    std err          t      P>|t|    95.0% Conf. Int.
-----+-----
mu            21.0683   9.064e-02    232.432   0.000   [ 20.891, 21.246]
Volatility Model
=====
              coef    std err          t      P>|t|    95.0% Conf. Int.
-----+-----
omega         0.1408   3.216e-02     4.377   1.201e-05   [7.774e-02, 0.204]
alpha[1]      1.0000   1.051e-02    95.116   0.000   [ 0.979, 1.021]
beta[1]       1.7790e-11  6.167e-06    2.885e-06  1.000   [-1.209e-05,1.209e-05]
=====

```

Рисунок 3.6 – Описова статистика та результати тесту Engle

Можна звернути увагу, що ймовірність, отримана в результаті, явно менша за поріг 0,05. Це свідчить про те, що у даному випадку необхідно

відхилити нуль-гіпотезу, яка припускає незмінність дисперсії у часі. Отже для залишків даної моделі характерна наявність ефекту ARCH.

Наступним кроком для моделювання волатильності дисперсії буде визначення порядку самої моделі. Для цього побудуємо графіки АКФ та ЧАКФ для квадратів залишків моделі (Рисунок 3.7).

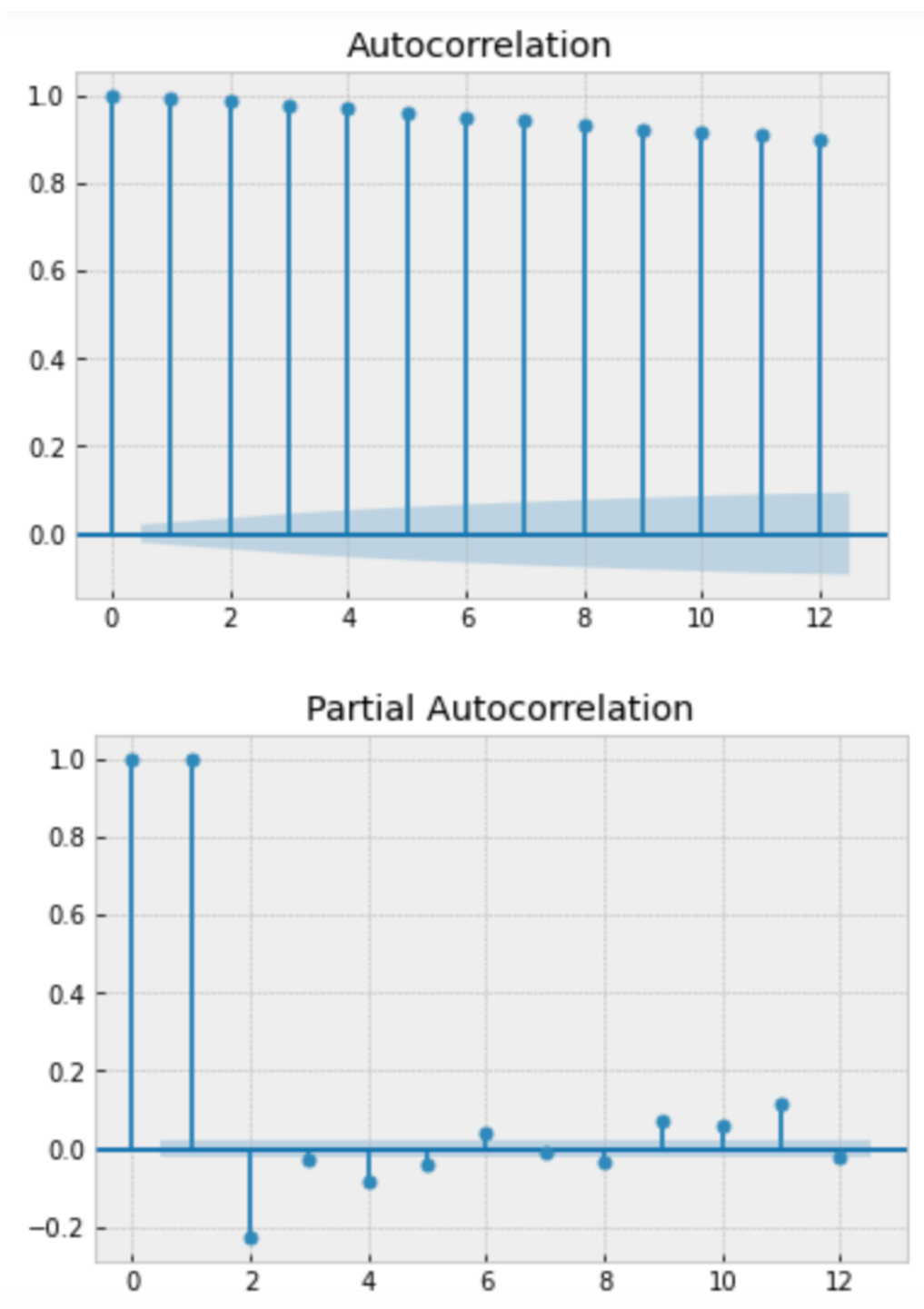


Рисунок 3.7 – Графіки АКФ та ЧАКФ для квадратів залишків моделі

Аналізуючи графіки, бачимо що можна брати параметри більше 1, тому порівняємо декілька моделей з різними параметрами, щоб обрати найкращу.

Таблиця 3.1 – Результати GARCH моделі

Model	MAE	MAPE	THEIL
GARCH (1,1)	45.503398699012	11.60310685149867	0.25971130284802
GARCH (2,4)	43.702794762397	11.51956147319662	0.25971123179284
GARCH (4,8)	43.259252938935	11.73202941090763	0.25708675806683
GARCH (4,12)	42.703393537471	11.60310954523587	0.25711302848347

Отримані результати прогнозів в таблиці 3.1 виявилися задовільними. В процесі дослідження також перевірялися моделі EGARCH і TARЧН, які показали майже такі ж оцінки, тому було вирішено не додавати їх у дану роботу. Як найкращу модель було обрано саме GARCH (4,12).

Також у даній магістерській роботі проведено дослідження щодо прогнозування часових рядів за допомогою рекурентних нейронних мереж. Для моделювання справжньої волатильності було обрано саме нейронну мережу LSTM, яка працює по принципу системи довгої короткочасної пам'яті. Представимо результати роботи даного методу в таблиці 3.2.

Таблиця 3.2 – Результати роботи LSTM моделі

Model	MAE	MAPE	THEIL
LSTM	7.3834419570547	0.42678149293463	0.06897247272352

Далі для легшого аналізу зобразимо результати найкращої GARCH моделі та LSTM моделі в таблиці 3.3.

Таблиця 3.3 – Порівняння результатів роботи моделей

Model	MAE	MAPE	THEIL
GARCH (4,12)	42.703393537471	11.60310954523587	0.25711302848347
LSTM	7.3834419570547	0.42678149293463	0.06897247272352

Отже, можна зробити висновок, що в даному випадку рекурентна нейронна мережа справляється краще, тобто має кращі оцінки та результати роботи.

3.4 Результати оцінювання ринкового ризику

Останнім кроком дослідження буде розрахунок ринкового ризику методом Value-at-Risk. Для розрахунку будемо використовувати прогнози, отримані саме за допомогою моделі LSTM. Згадуємо, що на початку розділу було зроблено дослідження про схильність до розподілу Стюдента, тобто розраховуємо показник саме для цього випадку. Як саме розраховується показник описано в другому розділі, тож не будемо на цьому зупинятися. Зобразимо результати в таблиці 3.4.

Таблиця 3.3 – Порівняння результатів роботи моделей

Показник	Результат
95%	478.834

99%	435.057
-----	---------

3.5 Висновки до розділу 3

Враховуючи складність фінансового ринку та нелінійну характеристику фінансових змінних, штучна нейронна мережа, як важлива форма штучного інтелекту, стала великим досягненням у багатьох фінансових галузях та може застосовуватися для аналізу фінансових процесів на фондовому ринку.

Згідно з літературними джерелами найбільшу ефективність при прогнозуванні цін на фондовому ринку продемонстрували нейронні мережі LSTM, саме тому вони обрані для моделювання в межах магістерської дисертації. Окрім цього, для моделювання дисперсії при аналізі фінансових показників на фондовому ринку обрано методологію ARCH/GARCH, яка є найбільш ефективною серед аналогів.

У ході дослідження мовою програмування Python було реалізовано таку систему підтримки прийняття рішень, яка дозволяє виконувати моделювання волатильності певного процесу та розраховувати ринковий ризик. Ця система використовує моделі GARCH та LSTM, а також методологію VaR для обчислення ринкового ризику. Результати дослідження показали, що найбільш ефективним методом виявилася саме нейронна мережа LSTM. Окрім того, було визначено значення показника VaR для 95% і 99% ймовірностей.

РОЗДІЛ 4 РОЗРОБКА СТАРТАП ПРОЕКТУ

4.1 Постановка задачі стартап-проекту

Своєчасне оцінювання ризику необхідне на усіх етапах прийняття рішень в багатьох сферах діяльності. Принципи ризик-менеджменту однакові для всіх сфер діяльності незважаючи на принципову відмінність напрямків. Мету ризик-менеджменту можна визначити таким чином: визначення характеру протікання процесу протягом усього історичного періоду та прогнозування його поведінки у майбутньому, ідентифікацію основних ризикових компонент та оцінювання їх впливу на подальшу діяльність установи.

Вагомі відмінності починаються на етапі підбору методів аналізу та математичного апарату. Тому для отримання хороших результатів в сфері аналізу відокремлюють системи для різних напрямків. Тому для оцінки ринкових ризиків розробляються окрема система. Дана система дає можливість завантажити дані в зручному форматі, швидко проаналізувати взаємозв'язки та спрогнозувати дані, оцінити ризики для заданого рівня довіри. Завдяки такій системі можна вчасно виявити можливі втрати та своєчасно зреагувати на них.

Для розробки стартап-проекту та виведення його на ринок необхідно провести детальне дослідження, яке передбачає виконання наведених нижче чотирьох кроків.

Перший крок. Маркетинговий аналіз стартап-проекту:

- розробка опису ідеї проекту, визначення основних напрямків використання товару чи послуги, формування основних відмінностей від конкурентів;
- аналіз ринкових можливостей;

- розробка стратегії для виведення товару на ринок.

Другий крок. Організація стартап-проекту, що включає:

- календарний план реалізації та запуску продукту;
- плановий обсяг виробництва, та відповідно матеріальних ресурсів і персоналу;
- витрати для реалізації проекту та запуску його.

Третій етап. Фінансово-економічний аналіз та аналіз ризиків:

- обсяг інвестиційних витрат;
- розрахунок фінансових показників проекту та визначення інвестиційної привабливості проекту;
- визначення основних ризиків проекту.

Четвертий етап. Комерціалізація проекту:

- визначення цільової групи інвесторів;
- складання інвестиційної пропозиції;
- визначення основних каналів для просування офerti інвесторам.

4.2 Карта проекту

Проект полягає у створенні системи прийняття рішень для автоматичного оцінювання ринкових ризиків фінансового ринку. Дана програма аналізує дані введені дані і видає оцінку ризиків. В таблиці 4.1 представлено основну інформацію по проекту.

Таблиця 4.1 – Інформаційна карта проекту

Назва проекту	Система аналізу ринкових ризиків на базі методів машинного навчання.
Автори проекту	Радіо Ольга Володимирівна
Коротка анотація	Система дозволяє оцінювати ризики, на основі яких користувач приймає рішення спрямоване на ріст прибутку.
Термін реалізації проекту	12 місяців
Необхідні ресурси	Інтелектуальні: спеціаліст з аналізу даних; розробник; тестувальник; продуктивний менеджер проекту. Матеріальні: комп'ютери/ноутбуки; сервер; офісне приміщення. Фінансові: заробітні плати працівникам; апаратне забезпечення; оренда приміщення.
Головні цілі та завдання проекту	Огляд підходів для статистичного аналізу даних, методів прогнозування та оцінювання ризиків. Створення програмного забезпечення для цієї системи.
Очікувані результати	Система автоматизованого моделювання ринкових процесів та оцінювання ризиків стане помічником у багатьох сферах діяльності. Вона дозволить вести більш ефективну економічну діяльність, а також при керуванні бізнесом. Завдяки якісному оцінюванню фінансового ризику компанії зможуть як застрахувати себе від майбутніх збитків, так і виконувати прибуткові ризикові операції з найменшими втратами.

4.3 Команда стартап- проекту

Обов'язки та ключові ролі початкової команди проекту висвітлені у таблиці 4.2.

Таблиця 4.2 – Команда проекту

Посада	Функціональні обов'язки та досвід роботи	Роль
Спеціаліст з аналізу даних	Вибір архітектури проекту; пошук нових галузей в яких доцільно буде використовувати сервіс; статистичний аналіз великих об'ємів даних та пошук зв'язків та закономірностей за допомогою сучасних методів.	Рішення бізнес задач з використанням передових методів обробки даних.
Розробник	Створення програмного забезпечення проекту та вдосконалення існуючих рішень.	Відповідальність за якість та функціональні можливості продукту.
Тестувальник	Пошук імовірних недоліків та помилок у функціонуванні продукту; пропозиції щодо вдосконалення.	Відповідальність за якість та функціональні можливості продукту.
Продуктовий менеджер проекту	Бренд-менеджмент; перемовини з замовниками системи; моніторинг та аналіз даного сектору послуг.	Пошук клієнтів, нових ніш на ринку, просування бренду.

4.4 Розробка бізнес-моделі проекту

Бізнес-модель проекту продемонстровано в таблиці 4.3.

Таблиця 4.3 – Бізнес-модель проекту.

Назва	Зміст
Споживчі сегменти	Дослідницькі центри; банки; університети; страхування; енергетика та інші компанії, специфіка роботи яких пов'язана із прогнозуванням та ризиком.
Ціннісні пропозиції	Ведення більш ефективної економічної діяльності, при керуванні бізнесом. Завдяки якісному оцінюванню фінансового ризику компанії зможуть як застрахувати себе від майбутніх збитків, так і виконувати прибуткові ризикові операції з найменшими втратами. Вирішення складних задач та виявлення основних можливостей для майбутнього розвитку. Продукт, який не потребує особливих налаштувань перед першим використанням.
Канали збуту	Прямі канали: <ul style="list-style-type: none"> - сайт-візитка; - конференції аналітиків/бізнес аналітиків; - презентації для цільового сегменту - розроблених рішень; - презентації в вузах. Партнерські канали: служба роботи з клієнтами.
Взаємовідносини з клієнтами	Служба підтримки. Супроводження продукту протягом певного терміну. Особливий підхід для кожного клієнта. Формат договорів на довгосрокову підписку.
Потоки надходження доходу	Розробка спеціалізованого продукту (за умови надання клієнтом бази даних). Покупка ліцензії на певний термін (універсальна web-версія продукту), довгосрокова підписка для використання необхідної частини розробленого рішення; разові виплати за проведення аналізу.

Ключові ресурси	Технічні ресурси, інтелектуальні ресурси, трудові ресурси.
Ключові види діяльності	Аналіз та прогнозування; візуалізація; клієнтська аналітика; керування даними; прийняття рішень; керування ризиками та їх оцінювання.
Ключові партнери	Компанії, що спеціалізуються на аудиті і консалтингу - актуальна інформація щодо стану економіки, показників, що впливають на прибутки компаній; досвід, що мають компанії в комерційній, технологічній та ін. сферах, а також в проектуванні та поширенні бізнес-рішень. Центри даних - постачання даних для навчання моделей, перевірки їх прогнозної якості, обмін сучасними підходами та моделями. Учбові центри - обмін досвідом, молодими спеціалістами.

4.5 Аналіз ринкових можливостей запуску стартап-проекту

Визначення ринкових можливостей, які можна використати під час ринкового впровадження проекту, та ринкових загроз, які можуть перешкодити реалізації проекту, дозволяє спланувати напрями розвитку проекту із урахуванням стану ринкового середовища, потреб потенційних клієнтів та пропозицій проектів-конкурентів.

Спочатку проводиться аналіз попиту: наявність попиту, обсяг, динаміка розвитку ринку (табл. 4.4).

Таблиця 4.4 – Попередня характеристика потенційного ринку стартап-проекту

Показники стану ринку	Характеристика
Кількість головних гравців, од	3
Динаміка ринку	Зростає
Наявність обмежень для входу	Вихід на ринок прямого конкурента, або суттєве зниження вартості послуг непрямих конкурентів (консалтингових компаній).
Специфічні вимоги до стандартизації та сертифікації	Немає
Середня норма рентабельності в галузі (або по ринку), %	300

Рентабельність — поняття, що характеризує економічну ефективність виробництва, за якої за рахунок грошової виручки від реалізації продукції (робіт, послуг) повністю відшкодовує витрати на її виробництво й одержується прибуток як головне джерело розширеного відтворення [24].

Суть одного із найважливіших методів оцінки економічної ефективності інвестицій полягає у розрахунку їх середньої рентабельності за формулою [25]:

$$R = \frac{P}{1 + n} * 100 ,$$

де P - прибуток за час експлуатації проекту; n - час експлуатації проекту.

Інвестувати грошові засоби доцільно тоді, коли від цього можна отримати більший прибуток, ніж від їх зберігання у банку. Порівнюючи середньорічну рентабельність інвестицій зі ставкою банківського відсотка, можна дійти висновку, що вигідніше [28].

Середня норма рентабельності в галузі (або по ринку) порівнюється із банківським відсотком на вкладення. За умови, що останній є вищим, можливо, має сенс вкласти кошти в інший проект.

За результатами аналізу таблиці робиться висновок щодо того, чи є ринок привабливим для входження за попереднім оцінюванням.

Надалі визначаються потенційні групи клієнтів, їх характеристики, та формується орієнтовний перелік вимог до товару для кожної групи (табл. 4.5).

Таблиця 4.5 – Характеристика потенційних клієнтів стартап-проекту

Потреба, що формує ринок	Цільова аудиторія	Відмінності у поведінці різних потенційних цільових груп	Вимоги споживачів до товару
Необхідне програмне забезпечення для прогнозування показників та оцінювання ринкових ризиків.	Потенційними цільовими групами є дослідницькі центри, університети та компанії, специфіка роботи яких пов'язана із аналізом, ризикменеджментом та прогнозуванням (державний сектор, охорона здоров'я, рітейл, банки).	Відмінність у підходах та застосованих моделях у прогнозуванні та ризикменеджменті, відмінність у сферах діяльності клієнтів.	Рішення має бути швидким, ефективним, мати зрозумілий юзерфрендлі інтерфейс.

Після визначення потенційних груп клієнтів проводиться аналіз ринкового середовища: складаються таблиці факторів, що сприяють ринковому впровадженню проекту, та факторів, що йому перешкоджають (табл. 4.6-4.7). Фактори в таблиці подаються в порядку зменшення значущості.

Таблиця 4.6 – Фактори загроз

Фактор	Зміст загрози	Можлива реакція компанії
Конкуренція	Вихід на ринок великої компанії.	Розширити рекламну кампанію до виходу нового гравця на ринок. Передбачити додаткові переваги власного ПЗ для того, щоб повідомити про них саме після виходу міжнародної компанії на ринок. Обрати нову цільову аудиторію і зосередитися на ній або наявність виходу з ринку.
Зміна потреб користувача	Дане ПЗ не задовольняє всі потреби користувачів. Користувачам необхідне ПЗ з іншим функціоналом та підходами до аналізу.	Передбачити можливість додавання нового функціоналу до існуючого рішення. Передбачити можливості зміни напрямку продукту взагалі, так звану точку Pivot.
Зростання попиту	Занадто швидкі темпи зростання попиту на даного типу ПЗ.	Розширення штату, пошук нових методів для задоволення як найбільшої кількості клієнтів; передбачення гнучкості рішення.

Таблиця 4.7 – Фактори можливостей

Фактор	Зміст можливості	Можлива реакція компанії
Конкуренція	Відсутність аналогічного продукту для вітчизняного користувача.	Локалізація та адаптація сервісу для локальних груп. Адаптація до вітчизняних особливостей.
Розробка нових методів прогнозування та оцінювання ризиків.	Розробка нових статистичних методів, що будуть швидші та ефективніші для конкретних цілей.	Покращити ПЗ додаванням нового функціоналу, розширення існуючих можливостей.
Зростання попиту.	Можливість залучити більшу кількість клієнтів.	Розширення інфраструктури, гнучкість рішення, розширення існуючих можливостей, підходів та моделей.

Надалі проводиться аналіз пропозиції: визначаються загальні риси конкуренції на ринку (табл. 4.8).

Таблиця 4.8 – Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства
Тип конкуренції: монополістична конкуренція	Існує декілька фірмконкурентів	Підтримка якості продукту та постійні нововведення
За рівнем конкурентної боротьби: міжнародний	Фірми-конкуренти з інших країн	Створити основу ПЗ таким чином, щоб можна було легко переробити дане ПЗ для використання у інших країнах та галузях
За галузевою ознакою: міжгалузева	Продукт може використовуватись для різних галузей	Постійне вдосконалення продукту, що не має привязки до сфери, гнучкість рішення
За видами товарів: товарно-видова	Види товарів є однаковими: ПЗ для статистики та бізнес-аналізу	Створити ПЗ, враховуючи недоліки конкурентів
За характером переваг: нецінова	Вдосконалення технології створення ПЗ, щоб собівартість була нижчою	Використання менш дорогих технологій для розробки, ніж використовують конкуренти
За інтенсивністю: марочна	Велике значення для клієнтів має бренд компаній конкурентів, які вже завоювали місце на ринку	Рекламні стратегії, підтримка компаній партнерів

Таблиця 4.9 – Аналіз конкуренції в галузі за М. Портером

Складові аналізу	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товаризамінники
	SAS SPSS Eviews	Наявність вже існуючих рішень	–	Контроль якості продукту	Наявність більш широкого функціоналу, зручнішого інтерфейсу та авторитет
Висновки	Доволі інтенсивна конкуренція і боротьба з вже закріпившимися на ринку гравцями	Є можливості входу в ринок, але є потенційні конкуренти. Строки виходу на ринок – 6 місяців	–	Клієнти диктують усі умови роботи на ринку	Необхідно випускати ПЗ не гірше, ніж у конкурентів та розширяти функціонал

За результатами аналізу таблиці робиться висновок щодо принципової можливості роботи на ринку з огляду на конкурентну ситуацію. Також робиться висновок щодо характеристик (сильних сторін), які повинен мати проект, щоб бути конкурентоспроможним на ринку. Другий висновок враховується при формулюванні переліку факторів конкурентоспроможності (табл. 4.10).

Таблиця 4.10 – Обґрунтування факторів конкурентоспроможності

Фактор конкурентоспроможності	Обґрунтування
Ціна	Більш доступна ціна збільшує кількість потенційних клієнтів
Функціонал	Методи і моделі, що підходять для багатьох сфер діяльності
Виконання ПЗ у кросплатформеному вигляді	Можливість використання ПЗ на будь якій платформі
Мобільність	Розробка веб-версій з базовим функціоналом

За визначеними факторами конкурентоспроможності (табл. 4.10) проводиться аналіз сильних та слабких сторін стартап-проекту (табл. 4.11).

Таблиця 4.11 – Порівняльний аналіз сильних та слабких сторін проекту

Фактор конкурентоспроможності	Бали 1-20	Рейтинг товарів конкурентів у порівнянні з даним						
		-3	-2	-1	0	+1	+2	+3
Ціна	10		+					
Функціонал	15				+			
Виконання ПЗ у кросплатформеному вигляді	20					+		
Мобільність	10			+				

Фінальним етапом ринкового аналізу можливостей впровадження проекту є складання SWOT-аналізу (матриці аналізу сильних (Strength) та слабких (Weak) сторін, загроз (Troubles) та можливостей (Opportunities) (табл.4.12) на основі виділених ринкових загроз та можливостей, та сильних і слабких сторін (табл. 4.11).

Таблиця 4.12 – SWOT-аналіз стартап-проекту

Сильні сторони: ціна, функціонал, мобільність	Слабкі сторони: кросплатформеність, недостатня кількість спеціалістів підкованих у найсучасніших методах
Можливості: конкуренція, розробка нових методів прогнозування, зростання попиту	Загрози: конкуренція, зміна потреб користувачів, занадто швидке зростання попиту

На основі SWOT-аналізу розробляються альтернативи ринкової поведінки (перелік заходів) для виведення стартап-проекту на ринок та орієнтовний оптимальний час їх ринкової реалізації з огляду на потенційні проекти конкурентів, що можуть бути виведені на ринок.

Визначені альтернативи аналізуються з точки зору строків та ймовірності отримання ресурсів (табл. 4.13).

Таблиця 4.13 – Альтернативи ринкового впровадження стартаппроекту.

Альтернативи ринкової поведінки	Ймовірність отримання ресурсів	Строки реалізації
Створення повноцінного вебсервісу	65%	11 місяців
Створення ПЗ	80%	12 місяців
Перехід на безкоштовне розповсюдження	50%	5 місяців

З означених альтернатив обирається та, для якої: а) отримання ресурсів є більш простим та ймовірним; б) строки реалізації – більш стислими.

4.6 Розроблення ринкової стратегії проекту

Розроблення ринкової стратегії першим кроком передбачає визначення стратегії охоплення ринку: опис цільових груп потенційних споживачів (табл. 4.14).

Таблиця 4.14 – Вибір цільових груп потенційних споживачів

Опис профілю цільової групи потенційних клієнтів	Готовність споживачі в сприйняти продукт	Орієнтовний попит в межах цільової групи	Інтенсивність конкуренції в сегменті	Простота входу в сегмент
Компанії (українські та міжнародні), специфіка роботи яких пов'язана із аналізом даних, прогнозуванням та ризикменеджментом	Висока	Високий	Сильна	Складно
Державні підприємства, специфіка роботи яких пов'язана із аналізом даних, прогнозуванням та ризикменеджментом.	Помірна	Високий	Сильна	Складно
Університети та дослідницькі центри, специфіка роботи яких пов'язана із аналізом даних, прогнозуванням та ризикменеджментом.	Помірна	Помірний	Помірна	Середня складність
Як цільові групи обрано: 1 та 3.				

За результатами аналізу потенційних груп споживачів (сегментів) автори ідеї обирають цільові групи, для яких вони пропонуватимуть свій товар, та визначають стратегію охоплення ринку:

- якщо компанія зосереджується на одному сегменті – вона обирає стратегію концентрованого маркетингу;

- якщо працює із кількома сегментами, розробляючи для них окремо програми ринкового впливу – вона використовує стратегію диференційованого маркетингу;

- якщо компанія працює із всім ринком, пропонуючи стандартизовану програму (включно із характеристиками товару/послуги) – вона використовує масовий маркетинг.

Для роботи в обраних сегментах ринку необхідно сформувавши базову стратегію розвитку (табл. 4.15).

Таблиця 4.15 – Визначення базової стратегії розвитку

Обрана альтернатива розвитку проекту	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку
Створення програмного забезпечення для аналізу даних, прогнозування та ризикменеджменту.	Ринкове позиціювання (Позиція компанії чи продукту показує чим він унікальний, чим відрізняється від конкурентів, чим корисний споживачу.)	Простота інтерфейсу, універсальність та ефективність продукту.	Стратегія диференціації.

Наступним кроком є вибір стратегії конкурентної поведінки (табл. 4.16).

Таблиця 4.16 – Визначення базової стратегії конкурентної поведінки

Чи є проект «першопрохідцем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки
Ні	Як нових, так і вже існуючих.	Так. Проте розробляти більш гнучкі та універсальні методи, розширювати функціонал новими методами та підходами.	Стратегія заняття конкурентної ніші.

На основі вимог споживачів з обраних сегментів до постачальника (стартап-компанії) та до продукту (див. табл. 4.6), а також в залежності від обраної базової стратегії розвитку (табл. 4.15) та стратегії конкурентної поведінки (табл. 4.16) розробляється стратегія позиціонування (табл. 4.17), що полягає у формуванні ринкової позиції (комплексу асоціацій), за яким споживачі мають ідентифікувати торгівельну марку/проект.

Таблиця 4.17 – Визначення стратегії позиціонування

Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартаппроекту	Вибір асоціацій, які мають сформувану комплексну позицію власного проекту (три ключових)
Ефективність, зручність інтерфейсу, швидкість роботи.	Диференціації	Простота користувацького інтерфейсу, що дозволяє пришвидшити та спростити роботу, швидкість роботи, що дозволяє підвищити швидкість експериментів, ефективність результатів, що дозволяє розробити якісний прогноз та оцінку. гнучкість та мобільність	Стабільність роботи, якість роботи, швидкість роботи, зручність роботи мобільність, гнучкість.

Результатом виконання підрозділу має стати узгоджена система рішень щодо ринкової поведінки стартап-компанії, яка визначатиме напрями роботи стартап-компанії на ринку.

4.7 Розробка маркетингової програми стартап-проекту

Першим кроком є формування маркетингової концепції товару, який отримає споживач.

Таблиця 4.18 – Визначення ключових переваг концепції потенційного товару

Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами
Якість прогнозу оцінки	Якісний прогноз будь-яких показників, визначення ризикових компонент	Відсутність самостійно будувати модель та обирати модель прогнозу. Комплексний підхід до оцінювання ризику.
Спрощення інтерфейсу користувача	Простота роботи з ПЗ.	Користувачам не потрібно замислюватись над тим, як саме побудувати прогноз. Незалежно, від даних (категоріальних, числових) ПЗ виконає аналіз, запропонує методи та підходи прогнозування та оцінювання ризиків.
Мобільність	Можливість використовувати базову веб-версію.	Користувачі не обов'язково мають бути прив'язані до конкретного місця, а зможуть виконати необхідний базовий аналіз у будь-який час та з будь-якого девайсу.

Надалі розробляється трирівнева маркетингова модель товару: уточнюється ідея продукту та/або послуги, його фізичні складові, особливості процесу його надання (табл. 4.19).

Таблиця 4.19 – Опис трьох рівнів товару

Рівні товару	Сутність та складові		
Товар за задумом	Зручність, швидкість та мобільність отримання практичного результату щодо прогнозування процесів та оцінювання ризиків.		
Товар у реальному виконанні	Властивості/характеристики	М/Нм	Вр/Ех/Тл/Е/Ор
	Якість.		
	Швидкість обробки.		
	Мобільність.		
	Ціна.		
	Якість: функціональне тестування (наявності багів), тестування перформансу, стабільності та глобалізації.		
	Пакування: відсутнє.		
Товар із підкріпленням	1-місячна пробна безкоштовна версія та безкоштовне встановлення на етапі бетатестування.		
	Постійна підтримка для користувачів.		
За рахунок чого потенційний товар буде захищено від копіювання: патент, система захисту.			

Після формування маркетингової моделі товару слід особливо відмітити – чим саме проект буде захищено від копіювання. Захист може бути організовано за рахунок захисту ідеї товару (захист інтелектуальної власності), або ноу-хау,

чи комплексне поєднання властивостей і характеристик, закладене на другому та третьому рівнях товару.

Наступним кроком є визначення цінових меж, якими необхідно керуватись при встановленні ціни на потенційний товар (остаточне визначення ціни відбувається під час фінансово-економічного аналізу проекту), яке передбачає аналіз ціни на товари-аналоги або товари субститути, а також аналіз рівня доходів цільової групи споживачів (табл. 4.20). Аналіз проводиться експертним методом.

Таблиця 4.20 – Визначення меж встановлення ціни

Рівень цін на товаризамінники	Рівень цін на товари-аналоги	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар/послугу
50000 грн	1000000 грн	250000 грн	30000 – 40000 грн

Таблиця 4.21 – Формування системи збуту

Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту

Купують ПЗ та роблять щорічні внески для подовження ліценції (аналог довгосрокові підписки)	Продаж	0 (напрямую)	Прямий канал збуту
---	--------	--------------	--------------------

Визначення ідеї та теми рекламного звернення зумовлює всі наступні етапи планування рекламної діяльності. Тому цей етап є основним для професіоналів — творчих працівників рекламних агенцій чи рекламних підрозділів підприємств.

Таблиця 4.22 – Концепція маркетингових комунікацій.

Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
Купівля ПЗ через Інтернет, робота з ПЗ на комп'ютерах з різними ОС.	Електронна пошта, Інтернет, мобільний зв'язок	Швидкодія, простота використання, ефективність, мобільність.	Показати переваги ПЗ, у тому числі і перед конкурентами	Деморолик. Використання сайтвізитки.

4.8 Висновки до розділу 4

Отже, можливість ринкової комерціалізації проекту присутня, оскільки на даного виду продукти наявний попит, динаміка ринку збільшується, рентабельність роботи на ринку є також значно високою. Існують перспективи впровадження з огляду на потенційні групи клієнтів, бар'єри входження, стан конкуренції, конкурентоспроможність проекту.

Цільовими групами вибрано компанії (українські та міжнародні), специфіка роботи яких пов'язана із аналізом даних, прогнозуванням та ризик менеджментом. Як альтернативу (варіант) впровадження для ринкової реалізації проекту доцільно обрати розроблення програмного забезпечення. Подальша імплементація проекту є доцільною.

ВИСНОВОК

В ході проходження практики було досліджено математичні методи і моделі, що дозволяють моделювати складні системи і процеси, оскільки фінансові процеси і є складними системами. Було досліджено величезну кількість джерел на дану тематику. В даній магістерській дисертації було розглянуто питання прогнозування волатильності та оцінки ризику акцій фінансового ринку. А також методів для прогнозування нестационарних часових рядів. Було оглянуто і досліджено велику кількість джерел на дану тематику. Також було досліджено методи фільтрації даних від шуму і проаналізовано для чого це необхідно, а також застосовано саме метод Калмана. Було розглянуто методи моделювання гетероскедастичних систем. А також способи оцінки систем на гетероскедастичність. Звичайно було розглянуто особливості ринкових і зокрема фондових ризиків і методів їх оцінки. А саме методика VaR і використання волатильності для оцінки ризиків. Також, в ході написання роботи було проведено аналіз даних і пошук оптимальних методів для побудови системи підтримки прийняття рішень для оцінювання ризику акцій. За результатами досліджень було обрано модель, що дає найкращі результати. І на основі даної моделі було побудовано систему підтримки прийняття рішень. Система була створена на основі емпіричної вибірки даних по цінам на акції компанії TESLA. Реалізовано систему на мові програмування Python, оскільки дана мова містить велику кількість бібліотек для роботи з часовими рядами і вказаними вище методами. Зокрема зручна робота з нейронними мережами. В кінці було виконано моделювання волатильності та оцінювання ризику акцій і представлено значення показника VaR для 95% і 99% ймовірностей. Дану роботу можна і далі продовжувати. До неї можна додати ще методології для оцінки

волатильності: нечітку логіку, метод подібних траєкторій, тощо. Також можна узагальнити на теорію портфелів і оцінювати ризики для портфеля акцій.

ПЕРЕЛІК ПОСИЛАНЬ

1. Конспект лекцій з курсу “Системи прийняття рішень” для студентів напряму підготовки 6.030502 “Економічна кібернетика” , спеціальності 051 “Економіка” / С. В. Гринчуцька – Тернопіль, ТНТУ імені І. Пулюя, 2017. - 130с.
2. Системи і методи підтримки прийняття рішень. навч. посіб. для здобувачів ступеня магістра за освітніми програмами «Системний аналіз та управління», «Системний аналіз фінансового ринку» спеціальності 124 «Системний аналіз» / П.І.Бідюк, О.Л.Тимощук, А.Є.Коваленко; Л.О.Коршевніук КПІ ім. Ігоря Сікорського ;. Електронні текстові дані. Київ : КПІ ім. Ігоря Сікорського, 2020. - 259 с.
3. Системи підтримки прийняття рішень : навчальний посібник для самостійного вивчення дисципліни / [уклад.: С. М. Братушка, С. М. Новак, С. О. Хайлук] ; Державний вищий навчальний заклад “Українська академія банківської справи Національного банку України”. – Суми : ДВНЗ “УАБС НБУ”, 2010. – 265 с.
4. Проектування інформаційних систем підтримки прийняття рішень: навч. посіб. / [Бідюк П.І., Коршевніук Л.О., Коваленко А.Є., Гожий О.П.]. — К.: НТУУ «КПІ», 2013. - 608с.
5. Ситник В. Ф. Системи підтримки прийняття рішень: Навч. посіб. / В.Ф. Ситник. — К.: КНЕУ, 2009. — 614 с.
6. Бідюк П.І., Коршевніук Л.О.Проектування ком’ютерних інформаційних систем підтримки прийняття рішень: навч. Навчальн. посіб. / [Бідюк П.І., Коршевніук Л.О.]. — К.: ННК ІПСА НТУУ «КПІ», 2010. - 310с.
7. What is a decision support system (DSS)? : Режим доступу: <https://www.techtarget.com/searchcio/definition/decision-support-system> (дата звернення: 15.10.2022).

8. Фондовий ринок: Підручник : у 2 кн. Кн. 2 / В.Д. Базилевич, В.М. Шелудько, В.В. Вірченко та ін. ; за ред. В.Д. Базилевича; Київ. нац. ун-т ім. Т. Шевченка. Київ : Знання, 2016. - 686 с.
9. Діденко В.М. Фондовий ринок : навч. посіб. ; за ред. В.М. Діденка. Чернівці : ТОВ «Видавництво Наші книги». 2012. - 624 с.
10. Поважний О. С. Цінні папери і фондовий ринок : навч. посіб. Львів : Магнолія. 2012. - 361 с.
11. Justin Sirignano, Rama Cont. Universal features of price formation in financial markets: perspectives from Deep Learning. 2018. – 21 p.
12. R. Cont. Statistical modeling of high frequency financial data: Facts, models and challenges, IEEE Signal Processing, 28, 2011. – pp.16-25.
13. M. Benzaquen, J. Donier, and J.P. Bouchaud. Unravelling the trading invariance hypothesis. Market Microstructure and Liquidity, 2016. – 45 p.
14. T. Andersen, O. Bondarenko, A. Obizhaeva and P. Kyle. Intraday Trading Invariance in the E-Mini S&P 500 Futures Market, Working Paper, 2017. – 58 p.
15. A. Kyle and A. Obizhaeva. Market microstructure invariance: Empirical hypotheses. Econometrica, 84(4), 2016. – pp. 1345-1404.
16. Безпалько О. В. Сутність ціни як економічної категорії та підходи до її формування / О. В. Безпалько // Економічний простір. – Дніпропетровськ : ДНУ, 2011. – № 53. – С. 32–38.
17. Market Risk: Режим доступу: <https://corporatefinanceinstitute.com/resources/knowledge/trading-investing/market-risk/> (дата звернення: 17.10.2022).
18. Прогнозування та аналіз часових рядів. Методичні вказівки до практичних занять та самостійної роботи студентів спеціальності 051 «Економіка» освітня програма «Економічна кібернетика», «Економічна аналітика» / Укл.: Юрченко М. Є. – Чернігів: ЧНТУ, 2018. – 88 с.

19. Часові ряди: Режим доступу: <https://kstat.pnu.edu.ua/wp-content/uploads/sites/63/2018/04/%D0%A7%D0%B0%D1%81%D0%BE%D0%B2%D1%96-%D1%80%D1%8F%D0%B4%D0%B8.pdf> (дата звернення: 18.10.2022).
20. Azoff, M. E. *Neural Network Time Series Forecasting of Financial Markets*/ M. E. Azoff. – Cornwall : John Wiley and Sons, 2014. – 212 p.
21. Wang Z, Wang L, Szolnoki A, Perc M. Evolutionary games on multilayer networks: a colloquium. *Eur Phys J B*. 2015. – pp. 1–15.
22. Jo T. VTG schemes for using back propagation for multivariate time series prediction. *Appl Soft Comput*. 2013. – pp. 692–702.
23. Teixeira LA, Oliveira A (2010) A method for automatic stock trading combining technical analysis and nearest neighbor classification. *Expert Syst Appl* 37(10):6885–6890.
24. Shynkevich Y, McGinnity TM, Coleman SA, Belatreche A, Li Y. Forecasting price movement using technical indicators: investing the impact of varying window length. *Neurocomputing*, 2017. – pp.71–88.
25. Tsinaslanidis PE, Kugiumtzis D. A prediction scheme using perceptually important points and dynamic time warping. *Expert Syst Appl.*, 2014. – pp. 684–694.
26. Lu CJ. Integrating independent component analysis-based denoising scheme with neural network for stock price prediction. *Expert Syst Appl.*, 2010. – pp. 705–712.
27. Pal SS, Kar S. Fuzzy time series model for unequal interval length using genetic algorithm. In: *Information technology and applied mathematics*, vol 699. Springer, 2019. - pp 205–216.
28. Rajab S, Sharma V. An interpretable neuro-fuzzy approach to stock price forecasting. *Soft Comput*, 2019. - pp. 921–936.
29. Zhou F, Zhou HM, Yang Z, Yang L. A strategy combining empirical mode decomposition and factorization machine based neural network for stock market trend prediction. *Expert Syst Appl*, 2019. – pp. 136–151.

30. Yang F, Chen Z, Li J, Tang L. A novel hybrid stock selection method with stock prediction. *Appl Soft Comput*, 2019. – pp. 820–831.
31. Boyacioglu MA, Avci D. An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: the case of the Istanbul stock exchange. *Expert Syst Appl*, 2010. – pp.79–91.
32. Niaki STA, Hoseinzade S. Forecasting S&P 500 index using artificial neural networks and design of experiments. *J Ind Eng Int.*, 2019. – pp. 12-44.
33. Rezaee MJ, Jozmaleki M, Valipour M. Integrating dynamic fuzzy C-means, data envelopment analysis and artificial neural network to online prediction performance of companies in stock exchange. *Phys A*, 2018. – pp. 78–93.
34. Pang X, Zhou Y, Wang P, Lin W, Chang V. An innovative neural network approach for stock market prediction. *J Supercomput*, 2018. – pp.1–21.
35. Kara Y, Boyacioglu MA, Baykan OK. Predicting direction of stock price index movement using artificial neural networks and support vector machines: the sample of the Istanbul Stock Exchange. *Expert Syst Appl*, 2011. – pp.44-68.
36. Sagheer, A. & Kotb, M. Time series forecasting of petroleum production using deep lstm recurrent networks. *Neurocomputing* 323, 2019. – pp. 203–213,.
37. Zheng, Y., Liu, Q., Chen, E., Ge, Y. & Zhao, J. Exploiting multi-channels deep convolutional neural networks for multivariate time series classification. *Front. Comput. Sci.* 10, 2018. – pp. 96–112,
38. Kao LJ, Chiu CC, Lu CJ, Yang JL. Integration of nonlinear independent component analysis and support vector regression for stock price forecasting. *Neurocomputing*, 2013. – pp. 534–542.
39. Niaki STA, Hoseinzade S. Forecasting S&P 500 index using artificial neural networks and design of experiments. *J Ind Eng Int* 9, 2013. – pp.1–9.
40. Esfahanipour A, Aghamiri W. Adapted neuro-fuzzy inference system on indirect approach TSK fuzzy rule base for stock market analysis. *Expert Syst Appl*, 2010. – pp. 474–478.

41. Rather AM, Agarwal A, Sastry VN. Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Syst Appl* 42(6), 2015. – pp. 323–329.
42. Alpaydin E. *Introduction to machine learning*. MIT Press, Cambridge, 2009. – 30 p.
43. Rout AK, Dash PK, Dash R, Bisoi R. Forecasting financial time series using a low complexity recurrent neural network and evolutionary learning approach. *J King Saud Univ Comput Inf Sci*, 2017. – pp. 536–552.
44. Ghasemiyeh R, Moghdani R, Sana SS. A hybrid artificial neural network with metaheuristic algorithms for predicting stock price. *Cybern Syst* 48(4), 2017. – pp. 365–392.
45. Svalina I, Galzina V, Lujic R, Simunovic G. An adaptive network-based fuzzy inference system (ANFIS) for the forecasting: the case of close price indices. *Expert Syst Appl* 40(15), 2015. – 48 p.
46. Lin CS, Chiu SH, Lin TY. Empirical mode decomposition-based least squares support vector regression for foreign exchange rate forecasting. *Econ Model* 29(6), 2014. – pp.258–268.
47. Sadaei HJ, Enayatifar R, Lee MH, Mahmud M. A hybrid model based on differential fuzzy logic relationships and imperialist competitive algorithm for stock market forecasting. *Appl Soft Comput* 40, 2017. – pp. 132–149.
48. Asadi S, Hadavandi E, Mehmanpazir F, Nakhostin MM. Hybridization of evolutionary Levenberg–Marquardt neural networks and data pre-processing for stock market prediction. *Knowl Based Syst* 35, 2017. – pp. 245–258.
49. Chien YWC, Chen YL. Mining associative classification rules with stock trading data-A GA-based method. *Knowl Based Syst* 23, 2010. – pp. 605–614.
50. Reza H, Shahrabi J, Hadavandi E. A bat-neural network multi-agent system (BNNMAS) for stock price prediction: case study of DAX stock price. *Appl Soft Comput* 29, 2015. – pp.196–210.
51. Patel J, Shah S, Thakkar P, Kotecha K. Predicting stock market index using fusion of machine learning techniques. *Expert Syst Appl* 42(4), 2015. – pp.216–236.

52. Pulido M, Melin P, Castillo O. Particle swarm optimization of ensemble neural networks with fuzzy aggregation for time series prediction of the Mexican Stock Exchange. *Inf Sci* 280, 2016. – pp. 188–204.

53. Kumar BS, Ravi V. A survey of the applications of text mining in financial domain. *Knowl Based Syst* 114, 2016. – pp.128–147.

54. І. Г. Лук'яненко, В. М. Жук. Аналіз часових рядів. Побудова ARIMA, ARCH/GARCH моделей з використанням пакета E.Views 6.0. Практичний посібник для роботи в комп'ютерному класі. Київ, 2013. – 188 с.

55. Long Short-Term Memory Networks: Режим доступу: <https://www.mathworks.com/help/deeplearning/ug/long-short-term-memory-networks.html> (дата звернення: 19.10.2022).

56. Єріна А. М. Економічна статистика : практикум / А. М. Єріна, О. К. Мазуренко, З. О. Пальян. – К. : ТОВ «УВПК:ЕксОб», 2012. – 232 с.

57. Variable transformations: Режим доступу: <https://www.xlstat.com/en/solutions/features/variable-transformations> (дата звернення: 19.10.2022).

58. Risk Management and VaR: Comparison of the accuracy of risk measurement for different assets. *Base Revista de Administração e Contabilidade da UNISINOS*, vol. 17, no. 4, 2020. - pp. 664-686.

ДОДАТОК А ЛІСТИНГ ПРОГРАМИ

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import statsmodels.api as sm
import arch
from statsmodels.graphics.tsaplots import plot_acf
from numpy import exp,sqrt
from scipy.stats import norm
import scipy.stats as scs
import statsmodels.tsa.api as smt
import os
import sys
import pandas as pd
import numpy as np
import statsmodels.formula.api as smf
import statsmodels.tsa.api as smt
import statsmodels.api as sm
import scipy.stats as scs
from arch import arch_model
import matplotlib.pyplot as plt
import matplotlib as mpl
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
```

```
data = pd.read_csv('TSLA.csv', delimiter=',')
data.index = list(data['Date'])
forecast_horizon = 150
data

df = data[['Date', 'Adj_Close']]
df['daily_returns'] = np.log(df.Adj_Close / df.Adj_Close.shift(1)) * 100 # Daily log
returns
df['monthly_std'] = df['daily_returns'].rolling(21).std() # Standard deviation across
trading month
df['annual_vol'] = df['monthly_std'] * np.sqrt(252) # Annualize monthly standard
deviation
df = df.dropna().reset_index(drop=True)

train = df[:-forecast_horizon]
train.index = list(train['Date'])
test = df[-forecast_horizon:]
test.index = list(test['Date'])
train, test

describe_stat = scs.describe(train.Adj_Close)
data.Adj_Close.plot(figsize=(12,6),label='TESLA Price')
plt.grid()
plt.legend()

plt.figure(figsize=(20,12))
data.Adj_Close.pct_change().hist(bins=50, density=True, histtype="stepfilled",
alpha=0.5)
```

```

plt.title("Histogram of TESLA price", weight="bold")

import scipy

Q = data.Adj_Close.pct_change().dropna()

scipy.stats.probplot(Q, dist=scipy.stats.norm,
plot=plt.figure(figsize=(20,12)).add_subplot(111))

plt.title("Normal distribution of TESLA price", weight="bold");

tdf, tmean, tsigma = scipy.stats.t.fit(Q)

scipy.stats.probplot(Q, dist=scipy.stats.t, sparams=(tdf, tmean, tsigma),
plot=plt.figure(figsize=(20,12)).add_subplot(111))

plt.title("Student`s distribution of TESLA price", weight="bold");

# Kalman filter

from pykalman import KalmanFilter

data.Adj_Close.fillna(0, inplace=True)

kf = KalmanFilter(initial_state_mean=0)

state_means,_ = kf.filter(data.Adj_Close.values)

state_means = state_means.flatten()

indexDate=data.index

new_data=pd.Series(state_means,index=indexDate)

#plt.plot(data.Adj_Close.values[:300])

#plt.plot(new_data.values[:300])

#plt.show()

df = pd.DataFrame(new_data, columns = ['Adj_Close'])

```



```

df['Date'] = new_data.index

df['daily_returns'] = np.log(df.Adj_Close / df.Adj_Close.shift(1)) * 100 # Daily log
returns

df['monthly_std'] = df['daily_returns'].rolling(21).std() # Standard deviation across
trading month

df['annual_vol'] = df['monthly_std'] * np.sqrt(252) # Annualize monthly standard
deviation

df = df.dropna().reset_index(drop=True)

train = df[0:len(df)-forecast_horizon]
train.index = list(train['Date'])

test = df[-forecast_horizon:-1]
test.index = list(test['Date'])

describe_stat = scs.describe(test.Adj_Close)

am=arch_model(train.Adj_Close)
res = am.fit(disp='off')
print(res.summary())

#heterosked test

import statsmodels.stats.diagnostic as smd

smd.het_arch(res.resid)
scs.describe(res.resid)
resid_all = res.resid/res.conditional_volatility
scs.describe(resid_all)

def tsplot(y, lags=None, figsize=(10, 8), style='bmh'):

```

```

if not isinstance(y, pd.Series):
    y = pd.Series(y)
with plt.style.context(style):
    fig = plt.figure(figsize=figsize)
    layout = (1, 2)
    #ts_ax = plt.subplot2grid(layout, (0, 0), colspan=2)
    #acf_ax = plt.subplot2grid(layout, (1, 0))
    #pacf_ax = plt.subplot2grid(layout, (1, 1))
    #qq_ax = plt.subplot2grid(layout, (2, 0))
    #pp_ax = plt.subplot2grid(layout, (2, 1))
    #y.plot(ax=ts_ax)
    #ts_ax.set_title('Time Series Analysis Plots')
    smt.graphics.plot_acf(y, lags=lags, alpha=0.5)
    smt.graphics.plot_pacf(y, lags=lags, alpha=0.5)
    #sm.qqplot(y, line='s', ax=qq_ax)
    #qq_ax.set_title('QQ Plot')
    #kde_pdf = scs.gaussian_kde(y)
    #x= np.linspace(min(y), max(y), len(y))
    #bins=np.arange(min(y), max(y),0.5)
    #pp_ax.plot(x,scs.norm.pdf(x,scs.tmean(y),scs.tstd(y)),label="parametric
distribution",color="red")
    #pp_ax.plot(x,kde_pdf(x),'r',label="KDE estimation",color="blue")
    #pp_ax.hist(y,bins=bins,density = True,color="gray",alpha=.6, label='data
histogram')
    #pp_ax.set_title("Probability density function")
    #pp_ax.legend()
    #plt.tight_layout()
plt.show()

```

```
return

#_ = tsplot(train.Adj_Close, lags=100)
am = arch_model(train.Adj_Close)
res = am.fit(dispatch='off')
#print(res.summary())

#import statsmodels.stats.diagnostic as smd

#smd.het_arch(res.resid)
_ = tsplot(res.resid**2, lags=12)

#GARCH

def mae(y_true, y_pred):
    y_true, y_pred = np.array(y_true), np.array(y_pred)
    return np.mean(np.abs(y_true - y_pred))

def mape(y_true, y_pred):
    y_true, y_pred = np.array(y_true), np.array(y_pred)
    return np.mean(np.abs((y_true - y_pred) / y_true))

def theil(y_true, y_pred):
    y_true, y_pred = np.array(y_true), np.array(y_pred)
    y_true_sq = np.sqrt(np.sum(y_true**2)/y_true.size)
    y_pred_sq = np.sqrt(np.sum(y_pred**2)/y_pred.size)
    y_sq = np.sqrt(np.mean((y_true - y_pred)**2))
```

```
return y_sq/(y_true_sq+y_pred_sq)
```

```
def mdl_conclusion(res):
```

```
    res_ser = pd.Series()
```

```
    res_ser = pd.Series(res.params).append([res.pvalues, pd.Series(res.model,
index=['MDL']),pd.Series([res.nobs, res.loglikelihood, res.aic, res.bic],
index=['Num', 'LL', 'AIC', 'BIC'])])
```

```
    return res_ser
```

```
from arch.univariate import ConstantMean, GARCH, Normal, StudentsT
```

```
am = ConstantMean(train.Adj_Close)
```

```
am.volatility = GARCH(p=13, q=1, power=2)
```

```
am.distribution = StudentsT()
```

```
#am.distribution = Normal()
```

```
res = am.fit(disp='off')
```

```
print(res.summary())
```

```
#res=res.resid**2
```

```
end_loc = 1000
```

```
res_test = am.fix(res.params.values, last_obs=end_loc)
```

```
pred_test = res_test.forecast(horizon=1, align = 'origin').variance
```

```
print(pred_test.shift(1)[end_loc:])
```

```
sample_var = res.conditional_volatility[end_loc:]
```

```
forecasted_var = pred_test.shift(1)[end_loc:]
```

```
#forecasted_var = pd.DataFrame(forecasts).T['h.1']
```

```
#df = pd.concat([sample_var,forecasted_var],1)
```

```
#df.columns = ['actual', 'forecasted']
plt.plot(sample_var, label = 'actual')
plt.plot(forcasted_var, label = 'forecasted')
plt.legend()
plt.show()
print('mae:', mae(sample_var, forcasted_var))
print('mape:', mape(sample_var, forcasted_var))
print('theil:', theil(sample_var, forcasted_var))

#egarch = EGARCH(p=1, q=1)
#egarch = EGARCH(p=1, o=1, q=1)
#exponential arch process: earch = EGARCH(p=5)

#from arch.univariate import EGARCH

#am = ConstantMean(train.Adj_Close)
#am.volatility = EGARCH(p=3, o=1, q=12)
#am.distribution = Normal()
#am.distribution = StudentsT()
#res = am.fit(dispatch='off')
#print(res.summary())
#end_loc = 1500
#res_test = am.fix(res.params.values, last_obs=end_loc)
#pred_test = res_test.forecast(horizon=1, align = 'origin').variance
#print(pred_test.shift(1)[end_loc:])
#sample_var = res.conditional_volatility[end_loc:]
#forcasted_var = pred_test.shift(1)[end_loc:]
```

```

#forecasted_var = pd.DataFrame(forecasts).T['h.1']
#df = pd.concat([sample_var,forecasted_var],1)
#df.columns = ['actual', 'forecasted']
#plt.plot(sample_var, label = 'actual')
#plt.plot(forecasted_var, label = 'forecasted')
#plt.legend()
#plt.show()
#print('mae:', mae(sample_var, forecasted_var))
#print('mape:', mape(sample_var, forecasted_var))
#print('theil:', theil(sample_var, forecasted_var))

#am = ConstantMean(train.Adj_Close)
#am.volatility = GARCH(p=3, o=1, q=12, power=1.0)
#am.distribution = StudentsT()
#am.distribution = Normal()
#res = am.fit(disp='off')
#print(res.summary())
#end_loc = 1500
#res_test = am.fix(res.params.values, last_obs=end_loc)
#pred_test = res_test.forecast(horizon=1, align = 'origin').variance
#print(pred_test.shift(1)[end_loc:])
#sample_var = res.conditional_volatility[end_loc:]
#forecasted_var = pred_test.shift(1)[end_loc:]
#forecasted_var = pd.DataFrame(forecasts).T['h.1']
#df = pd.concat([sample_var,forecasted_var],1)
#df.columns = ['actual', 'forecasted']
#plt.plot(sample_var, label = 'actual')

```

```

#plt.plot(forcasted_var, label = 'forcasted')
#plt.legend()
#plt.show()
#print('mae:', mae(sample_var, forcasted_var))
#print('mape:', mape(sample_var, forcasted_var))
#print('theil:', theil(sample_var, forcasted_var))

```

```

###VAR

```

```

def VaR_gaus(p, volatility, mean):

```

```

    v_a_r=pd.Series()
    coef = scs.norm.ppf(p)
    v_a_r=mean + coef*volatility
    return v_a_r

```

```

def ES_gaus(p, volatility, mean):

```

```

    es=pd.Series()
    coef = -scs.norm.pdf(scs.norm.ppf(1-p))/(p)
    es=mean + coef*volatility
    return es

```

```

def VaR_t(p, volatility, mean, nu):

```

```

    v_a_r=pd.Series()
    coef = np.sqrt((nu-2)/nu)*scs.t.ppf(p, df = nu)
    v_a_r = mean + coef * volatility
    return v_a_r

```

```

def ES_t(p, volatility, mean, nu):
    es=pd.Series()

    coef = np.sqrt((nu-2)/nu)*(-scs.t.pdf(scs.t.ppf(1-p,df = nu), df =nu)/p)*((nu +
(scs.t.ppf(p, df = nu))**2)/(nu-1))

    es=mean + coef * volatility

    return es

#quantile for 95% (also need for 99%)
p=0.05-scs.norm.pdf(scs.norm.ppf(1-p))/(p)

#alpha = 1-p = 0.95
nu=1000

np.sqrt((nu-2)/nu)*(-scs.t.pdf(scs.t.ppf(1-p,df = nu), df =nu)/p)*((nu + (scs.t.ppf(p, df
= nu))**2)/(nu-1))

import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout, LSTM
import numpy
import matplotlib.pyplot as plt
import pandas
import math

from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import mean_squared_error

# fix random seed for reproducibility

numpy.random.seed(7)

```



```
# MinMaxNormalization
```

```
dataset = forecasted_var['h.1'].apply(lambda x: (x - min(forecasted_var['h.1'])) /
(max(forecasted_var['h.1']) - min(forecasted_var['h.1'])))
```

```
target = sample_var.apply(lambda x: (x - min(sample_var)) / (max(sample_var) -
min(sample_var)))
```

```
train_size = int(len(dataset) * 0.7)
```

```
test_size = len(dataset) - train_size
```

```
train = dataset[0:train_size]
```

```
test = dataset[train_size:len(dataset)]
```

```
target_train = target[0:train_size]
```

```
target_test = target[train_size:len(target)]
```

```
len(train), len(test), len(target_train), len(target_test)
```

```
# convert an array of values into a dataset matrix
```

```
def create_dataset(dataset, look_back=1):
```

```
    dataX, dataY = [], []
```

```
    for i in range(len(dataset)-look_back-1):
```

```
        a = dataset[i:(i+look_back)]
```

```
        dataX.append(a)
```

```
        dataY.append(dataset[i + look_back])
```

```
    return numpy.array(dataX), numpy.array(dataY)
```

```
look_back = 1
```

```
trainX, trainY = create_dataset(list(target_train), look_back)
```

```
testX, testY = create_dataset(list(target_test), look_back)
```

```
# reshape input to be [samples, time steps, features]

trainX = numpy.reshape(trainX, (trainX.shape[0], 1, trainX.shape[1]))
testX = numpy.reshape(testX, (testX.shape[0], 1, testX.shape[1]))

# create and fit the LSTM network

model = Sequential()
model.add(LSTM(4, input_shape=(1, look_back)))
model.add(Dense(1))
model.compile(loss='mean_squared_error', optimizer='adam')
model.fit(trainX, trainY, epochs=100, batch_size=1, verbose=2)

# make predictions

trainPredict = model.predict(trainX)
testPredict = model.predict(testX)
trainPredict = trainPredict.reshape(1, len(trainPredict))[0]
testPredict = testPredict.reshape(1, len(testPredict))[0]
pred_train = []
pred_test = []
y_train = []
y_test = []
trainScore = math.sqrt(mean_squared_error(trainY, trainPredict))
print('Train Score: %.2f RMSE' % (trainScore))
testScore = math.sqrt(mean_squared_error(testY, testPredict))
```

```

print('Test Score: %.2f RMSE' % (testScore))
trainScore = math.sqrt(mean_squared_error(y_train, pred_train))
print('Train Score: %.2f RMSE' % (trainScore))
testScore = math.sqrt(mean_squared_error(y_test, pred_test))
print('Test Score: %.2f RMSE' % (testScore))
print('mae:', mae(y_train, pred_train))
print('mae_test:', mae(y_test, pred_test))
print('mape:', mape(y_train, pred_train))
print('mape_test:', mape(y_test, pred_test))
print('theil:', theil(y_train, pred_train))
print('theil_test:', theil(y_test, pred_test))
p = 0.01
nu = 1000

from scipy.stats import norm

var = []
var_t = []
for i in range(len(pred_test)):
    mu = np.mean(data)
    vol = pred_test[i]
    #var.append(norm.ppf(confidence_level=p, mu, vol))
    var.append(VaR_gaus(p, vol, mu))
    var_t.append(VaR_t(p, vol, mu, nu))

data['Return'] = 100 * (data['Adj_Close'].pct_change())
data.dropna(inplace=True)

```

```
fig,ax = plt.subplots(figsize=(8,4))
#ax.spines('top','right').set_visible(False)
plt.plot(data['Return'], label = 'Daily Returns')
plt.legend(loc='upper right')
plt.title('Daily Returns Over Time')

import math
daily_volatility = data['Return'].std()

monthly_volatility = math.sqrt(21) * daily_volatility

annual_volatility = math.sqrt(252) * daily_volatility

from tabulate import tabulate

print(tabulate(['Tesla',daily_volatility,monthly_volatility,annual_volatility]),headers
= ['Daily Volatility %', 'Monthly Volatility %', 'Annual Volatility %'],tablefmt =
'fancy_grid',stralign='center',numalign='center',floatfmt=".2f"))

from arch import arch_model
from arch.__future__ import reindexing

garch_model = arch_model(data['Return'], p = 12, q = 3, mean = 'constant', vol =
'GARCH', dist = 'normal')

gm_result = garch_model.fit(dispatch='off')
print(gm_result.params)
```

```

print('\n')

gm_forecast = gm_result.forecast(horizon = 5)
print(gm_forecast.variance[-1:])

rolling_predictions = []
test_size = 150

for i in range(test_size):
    train = data['Return'][:-test_size-i]
    model = arch_model(train, p=1, q=1)
    model_fit = model.fit(dispatch='off')
    pred = model_fit.forecast(horizon=1)
    rolling_predictions.append(np.sqrt(pred.variance.values[-1, :][0]))

rolling_predictions = pd.Series(rolling_predictions, index=data['Return'].index[-150:])

fig, ax = plt.subplots(figsize=(10,4))
#ax.spines[['top', 'right']].set_visible(False)
plt.plot(rolling_predictions)
plt.title('Rolling Prediction')

fig, ax = plt.subplots(figsize=(13,4))
ax.grid(which="major", axis='y', color='#758D99', alpha=0.3, zorder=1)
#ax.spines[['top', 'right']].set_visible(False)
plt.plot(data['Return'][-150:])

```

```
plt.plot(rolling_predictions)
plt.title('Tesla Volatility Prediction - Rolling Forecast')
plt.legend(['True Daily Returns', 'Predicted Volatility'])
```