

УДК 681.518.25

О.М. Солошенко

ВДОСКОНАЛЕННЯ МЕТОДУ ІТЕРАТИВНОЇ КЛАСИФІКАЦІЇ З ВКЛЮЧЕННЯ ВІДХИЛЕНИХ ЗАЯВОК У КРЕДИТНОМУ СКОРИНГУ

The objective of the research is the adjustment of the Iterative Reclassification Method for including the rejected applications into the credit scoring. The methodology of implementation uses partially classified data and the logistic regression generalization. The method of the adjustment the Weight Of Evidence and the Information Value indicators using the rejected loan applications is proposed at the first stage. The process of including the rejected applications into predictive power analysis of characteristics has been demonstrated, providing alternative and adjusted discretization process for the continuous variables. The method of including the adjusted Weight Of Evidence and the partially classified rejected applications into logistic regression procedure is proposed at the second stage. The general method of the Iterative Probability Recalculation is proposed at the final stage, using the adjusted logistic regression approach. The research results are the significant improvement of the Iterative Reclassification Method and the generalization of the logistic regression application. As a conclusion, the main advantages of the method are given in comparison with the classical Iterative Reclassification Method, for instance, the probabilistic duality of the rejected applications is mentioned here.

Keywords: credit scoring, reject inference, logistic regression, Weight Of Evidence, Information Value, partially classified data, data mining, binary classification, Iterative Reclassification Method.

Вступ

Сучасний ризик-менеджмент неможливий без застосування статистичних моделей для прогнозування майбутньої кредитоспроможності клієнтів [1]. До найбільш поширених технологій аналізу даних (Data Mining) з метою оцінювання кредитоспроможності позичальників належить побудова скорингових моделей [2]. Однією з найпоширеніших проблем при скоринговому моделюванні є аналіз відхилених заявок (Reject Inference) з метою уникнення неповноти інформації внаслідок деяких бізнес-правил, попередніх скорингових моделей, кредитно-ризикової політики й інших перевірок (у т.ч. ручного розгляду та суб'єктивних факторів), що відсікали вибірку заявок, застосовуючи певне узагальнене правило з їх відхилення, що, власне, й визначило формування множини прийнятих заявок [3].

Значною практичною перевагою та необхідністю включення відхилених заявок у вибірку є забезпечення стабільності скорингової карти відносно кожної змінної при її впровадженні та забезпечення відсутності зміщення цільової бінарної змінної в бік більшої частки позитивних заявок при побудові моделі. Вирішення проблеми відхилених заявок з метою включення їх до складу навчальної вибірки залежить від механізму відмови в наданні кредитів по кредитних заявках. Основними такими механізмами є повністю випадкове відхилення (Missing Completely At Random, MCAR), випадкове відхилення (Missing At Random, MAR) та неви-

падкове відхилення (Missing Not At Random, MNAR) [4]. Залежно від механізму відхилення заявок вибирається алгоритм виведення ймовірностей належності відхиленої заявки до певного класу – позитивного та негативного (Good\Bad) [5].

Недоліком більшості методів аналізу відхилених заявок є те, що вони передбачають у кінцевому підсумку виведення бінарного результату для цільової змінної, наприклад, за допомогою деякого порогу відсікання (Cut-Off Rate) й подальшого включення відхилених заявок власне з бінарною цільовою змінною у вибірку з прийнятими заявками, наприклад, за допомогою методу ітеративної класифікації (Iterative Reclassification Method) [3, 6].

Тому актуальним є завдання розроблення методу включення нечітко класифікованих відхилених заявок (Partial Good/Bad Classifications) з імовірнісним результатом належності до одного із взаємовиключних класів – завдання вдосконалення методу ітеративної класифікації з включення відхилених заявок у кредитному скорингу [3].

Статтю присвячено розробленню нового методу включення частково класифікованих відхилених заявок як на рівні аналізу характеристик та їх дискретизації, так і на рівні оцінювання параметрів логістичної регресії в цілому.

Постановка задачі

Об'єктом дослідження є алгоритм включення відхилених заявок у вибірку для побудо-

ви скорингових моделей, зокрема метод ітеративної класифікації з включення відхилених заявок у кредитному скорингу.

Предметом дослідження є особливості та недоліки методу ітеративної класифікації з включення відхилених заявок у кредитному скорингу, зокрема використання лише бінарних оцінок.

Мета роботи – вдосконалити метод ітеративної класифікації з включення відхилених заявок у кредитному скорингу, а саме: 1) розробити метод перерахунку показників ваги значення змінної (Weight Of Evidence, WOE) та значення інформації (Information Value, IV) на рівні аналізу характеристик і дискретизації; 2) модифікувати метод максимальної правдоподібності з метою врахування підмножини відхилених заявок, що нечітко класифіковані.

Вдосконалення методу ітеративної класифікації на етапі аналізу характеристик

Для розширення області визначення WOE та IV як існуючих функцій багатьох змінних з множини додатних раціональних чисел [3] на множини додатних дійсних чисел, введемо необхідні для оперування з нечітко класифікованими даними [3] нові поняття: нечітко покращена вага значення змінної (Fuzzy Adjusted Weight Of Evidence, FAWOE) та нечітко покращене значення інформації (Fuzzy Adjusted Information Value, FAIV).

Проблема коректного включення відхилених заявок полягає в тому, що інформація про обслуговування кредиту по відхилених заявках відсутня (рис. 1), тобто має місце невизначене значення цільової бінарної змінної належності до класу позичальників (Good\Bad) [5].

Нехай для кожної з відхилених заявок залежно від характеристик було оцінено ймовірність $p_1(x_i) = P\{y_i = 1 | x_i\}$ належності до класу позитивних заявок, наприклад, ітеративно за допомогою алгоритму максимізації математичного сподівання (Expectation Maximization, EM) [5]. Тоді ймовірність належності до класу негати-

вних заявок дорівнює: $p_0(x_i) = 1 - p_1(x_i)$. Зазначимо, що EM-алгоритм дає змогу спочатку оцінити дві умовні ймовірності всієї множини характеристик – на підмножині позитивних заявок $p_G(x_i) = P\{x_i | y_i = 1\}$ та на підмножині негативних заявок $p_B(x_i) = P\{x_i | y_i = 0\}$, враховуючи як прийняті, так і відхилені заявки. Також завданням EM-алгоритму є оцінювання співвідношення позитивних і негативних заявок (їх безумовних ймовірностей) із врахуванням відхилених заявок $\pi_G = P\{y = 1\}$ та $\pi_B = P\{y = 0\}$. Тоді за теоремою Байєса матимемо

$$p_1(x_i) = \frac{\pi_G p_G(x_i)}{\pi_G p_G(x_i) + \pi_B p_B(x_i)}$$

Унаслідок застосування алгоритму максимізації математичного сподівання ми маємо не лише n прийнятих заявок з відомим бінарним результатом $y_i \in \{0,1\}$, а також m відхилених заявок з оціненим ймовірнісним результатом $p_i \in (0;1)$, $p_i = p_1(x_i)$.

Формули для обчислення показників IV та WOE на множині прийнятих заявок з чітко визначеним бінарним результатом мають такий вигляд [3]:

$$WOE_i = \ln\left(\frac{g_i}{b_i}\right),$$

$$IV = \sum_{i=1}^c (g_i - b_i) \ln\left(\frac{g_i}{b_i}\right) = \sum_{i=1}^c (g_i - b_i) WOE_i,$$

де g_i – відносна кількість позитивних заявок у сегменті (кластері) до загальної кількості позитивних заявок:

$$g_i = \frac{G_i}{\sum_{i=1}^c G_i} = \frac{G_i}{G},$$

тобто розподіл позитивних заявок по дискретних або дискретизованих значеннях змінної, тому

$$\sum_{i=1}^c g_i = 1;$$

b_i – відносна кількість негативних заявок у сегменті (кластері) до загальної кількості негативних заявок:

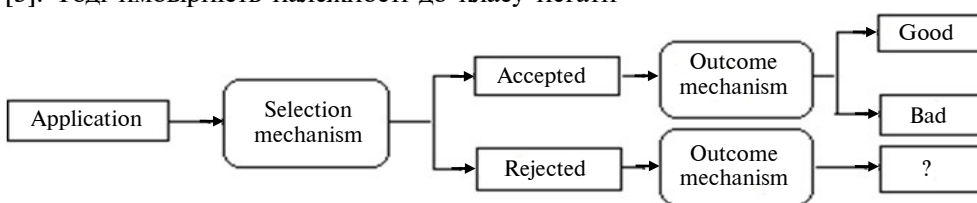


Рис. 1. Процес формування множини відхилених заявок, що створює проблему невизначеності цільової змінної [5]

$$b_i = \frac{B_i}{\sum_{i=1}^c B_i} = \frac{B_i}{B},$$

тобто розподіл негативних заявок по дискретних або дискретизованих значеннях змінної, тому

$$\sum_{i=1}^c b_i = 1.$$

Для кожної з характеристик множина m відхилених заявок розподіляється по сегментах значень цієї змінної, а кожен елемент множини відхилених заявок, у свою чергу, має ймовірність належності до множини позитивних заявок. Позначимо кількість відхилених заявок в i -му кластері деякої характеристики як m_i , тоді

$$m_1 + m_2 + \dots + m_c = \sum_{i=1}^c m_i = m.$$

Введемо внутрішню нумерацію з індексом j для кожного i -го сегмента значень змінної p_{ij} .

Удосконалення методу полягає в підрозумуванні нечітко покращеного (Fuzzy Adjusted) показника WOE (FAWOE) із врахуванням аналізу відхилених заявок (Reject Inference) таким чином:

$$\text{WOE}_i^* = \ln \left(\frac{\frac{g_i G + \sum_{j=1}^{m_i} p_{ij}}{G + \sum_{i=1}^c \sum_{j=1}^{m_i} p_{ij}}}{\frac{b_i B + \sum_{j=1}^{m_i} (1 - p_{ij})}{B + \sum_{i=1}^c \sum_{j=1}^{m_i} (1 - p_{ij})}} \right),$$

тобто через абсолютні кількості для множини прийнятих заявок і через імовірності для відхилених заявок маємо

$$\text{WOE}_i^* = \ln \left(\frac{\frac{G_i + \sum_{j=1}^{m_i} p_{ij}}{\sum_{i=1}^c G_i + \sum_{i=1}^c \sum_{j=1}^{m_i} p_{ij}}}{\frac{B_i + \sum_{j=1}^{m_i} (1 - p_{ij})}{\sum_{i=1}^c B_i + \sum_{i=1}^c \sum_{j=1}^{m_i} (1 - p_{ij})}} \right).$$

Удосконалення методу на цьому етапі полягає в тому, що кожен елемент множини від-

хилених заявок вносить дійсну частку ($0 \leq p_{ij} \leq 1$) до кількості позитивних заявок в i -му кластері, що відповідає його ймовірності належності до позитивних заявок, а також вносить дійсну частку ($0 \leq 1 - p_{ij} \leq 1$) до кількості негативних заявок в i -му кластері, що відповідає ймовірності належності до негативних заявок ($1 - p_{ij}$). Введемо такі позначення:

G_i^* – прогнозована кількість позитивних заявок серед відхилених заявок в i -му сегменті; B_i^* – прогнозована кількість негативних заявок серед відхилених заявок в i -му сегменті.

Тоді $G_i^* = \sum_{j=1}^{m_i} p_{ij}$ і $B_i^* = \sum_{j=1}^{m_i} (1 - p_{ij})$. Очевидно,

що виконується така рівність: $G_i^* + B_i^* = m_i$, тобто дійсні числа G_i^* та B_i^* в сумі становлять цілу кількість відхилених заявок у деякому i -му сегменті значень змінної. Тепер покращений показник FAWOE у поняттях додаткових прогнозованих дійсних кількостей позитивних і негативних заявок записується таким чином:

$$\text{FAWOE}_i = \text{WOE}_i^* = \ln \left(\frac{\frac{G_i + G_i^*}{\sum_{i=1}^c (G_i + G_i^*)}}{\frac{B_i + B_i^*}{\sum_{i=1}^c (B_i + B_i^*)}} \right),$$

де $G_i^* = \sum_{j=1}^{m_i} p_{ij}$ і $B_i^* = \sum_{j=1}^{m_i} (1 - p_{ij})$.

Очевидно, що сума позитивних і негативних заявок (із врахуванням прогнозованих та фактичних заявок) дорівнює загальній сумі заявок (прийнятих і відхилених):

$$\begin{aligned} & \sum_{i=1}^c (G_i + G_i^* + B_i + B_i^*) = \\ & = \sum_{i=1}^c (G_i + B_i) + \sum_{i=1}^c (G_i^* + B_i^*) = \\ & = n + \sum_{i=1}^c m_i = n + m. \end{aligned}$$

Як наслідок, нечітко покращений (Fuzzy Adjusted) показник IV (FAIV) буде обчислюватись так:

$$FAIV = IV^* = \sum_{i=1}^c \left(\frac{G_i + G_i^*}{\sum_{i=1}^c (G_i + G_i^*)} - \frac{B_i + B_i^*}{\sum_{i=1}^c (B_i + B_i^*)} \right) \times \sum_{i=1}^c \ln \left(\frac{\frac{G_i + G_i^*}{\sum_{i=1}^c (G_i + G_i^*)}}{\frac{B_i + B_i^*}{\sum_{i=1}^c (B_i + B_i^*)}} \right).$$

Враховуючи формулу FAIV, зберігається властивість зв'язку з FAWOE:

$$FAIV = IV^* = \sum_{i=1}^c \left(\frac{G_i + G_i^*}{\sum_{i=1}^c (G_i + G_i^*)} - \frac{B_i + B_i^*}{\sum_{i=1}^c (B_i + B_i^*)} \right) FAWOE_i.$$

Формули FAWOE і FAIV відображають узагальнення існуючих формул WOE і IV та збігаються з ними при відсутності відхилених заявок в аналізі характеристик, що є однією з найголовніших властивостей введених понять:

$$FAWOE_i | \{\forall i \in \{1, \dots, c\} : G_i^* = 0, B_i^* = 0\} =$$

$$= \ln \left(\frac{\frac{G_i}{\sum_{i=1}^c G_i}}{\frac{B_i}{\sum_{i=1}^c B_i}} \right) = WOE_i,$$

$$FAIV | \{\forall i \in \{1, \dots, c\} : G_i^* = 0, B_i^* = 0\} =$$

$$= \sum_{i=1}^c \left(\frac{G_i}{\sum_{i=1}^c G_i} - \frac{B_i}{\sum_{i=1}^c B_i} \right) \sum_{i=1}^c \ln \left(\frac{\frac{G_i}{\sum_{i=1}^c G_i}}{\frac{B_i}{\sum_{i=1}^c B_i}} \right) = IV.$$

Запропоноване вдосконалення на етапі аналізу характеристик посідає чільне місце в загальному процесі вдосконалення методу ітеративної класифікації з включення відхилених заявок у кредитному скорингу, а саме є пер-

шою складовою частиною ітеративного методу, аналогічною до кроку підрахунку математичного сподівання (Expectation Step) EM-алгоритму (рис. 2).

Наступним кроком є використання оцінок FAWOE на етапі другої складової частини методу – методу моделювання.



Рис. 2. Два етапи вдосконалення методу ітеративної класифікації з включення відхилених заявок у кредитному скорингу

Перехід на наступну ітерацію в методі відбувається через повторне оцінювання ймовірнісної цільової змінної для кожного елемента вибірки відхилених заявок.

Вдосконалення методу ітеративної класифікації на етапі моделювання за допомогою логістичної регресії

Суть методу полягає в покращенні логарифма функції правдоподібності [7] на всій множині заявок $m + n$ (*rejected + accepted*) із врахуванням множини відхилених заявок (*rejected*), яка доповнює множину прийнятих заявок (*accepted*), що можна записати таким чином:

$$L^*(c) = \sum_{i=m+1}^{m+n} \left\{ (1 - y_i) \ln \left(1 - \frac{1}{1 + \exp \left\{ - \left(c_0 + \sum_{j=1}^k c_j x_{ij} \right) \right\}} \right) \right\} +$$

$$\begin{aligned}
& + y_i \ln \left(\frac{1}{1 + \exp \left\{ - \left(c_0 + \sum_{j=1}^k c_j x_{ij} \right) \right\}} \right) \Bigg\} + \\
& + \sum_{i=1}^m \left\{ (1 - p_i) \ln \left(\frac{1}{1 + \exp \left\{ - \left(c_0 + \sum_{j=1}^k c_j x_{ij} \right) \right\}} \right) \right\} + \\
& + p_i \ln \left(\frac{1}{1 + \exp \left\{ - \left(c_0 + \sum_{j=1}^k c_j x_{ij} \right) \right\}} \right) \Bigg\},
\end{aligned}$$

де x_{ij} – елемент матриці вимірів (індекс i визначає номер виміру, а індекс j – номер змінної), c_j – невідомі коефіцієнти логістичної регресії для відповідних змінних (причому c_0 – коефіцієнт зміщення – *intercept*, а k – кількість змінних). У розробленому методі $x_{ij} = \text{FAWOE}_j(x_{ij}^*)$, де індекс i визначає номер заявки (прийнятої або відхиленої), а індекс j – номер змінної, x_{ij}^* – фактичне значення параметра (числове чи текстове), якому ставиться у відповідність значення FAWOE-перетворення.

Зауважимо, що для координат вектора результатів цілі числа $y_i \in \{0; 1\}$ відповідають прийнятим заявкам, але для відхилених заявок $p_i = p_1(x_i)$ – це дійсні числа (ймовірності належності до множини позитивних заявок). Тому аналітичні формули для першої та другої похідних на класичній області визначення цільової змінної тут не підходять, через що використовуємо метод Ньютона [7] з числовим обчисленням градієнта та матриці Гессе або з відповідними аналітичними формулами на розширеній області визначення:

$$\mathbf{c}_{p+1} = \mathbf{c}_p - ((L^*)''(\mathbf{c}_p))^{-1} (L^*)'(\mathbf{c}_p),$$

$$\text{де } (L^*)''(\mathbf{c}_p) = \left[\frac{\partial^2 L^*(\mathbf{c}_p)}{\partial c_i \partial c_j} \right]_{i,j=1\dots k} \in \text{Mat}(k, k), (L^*)'(\mathbf{c}_p) -$$

вектор-градієнт розмірності k .

Умова завершення ітераційного процесу така

$$\|\mathbf{c}_{p+1} - \mathbf{c}_p\| \leq \varepsilon.$$

Запропонований метод можна назвати методом нечітко покращеної логістичної регресії (Fuzzy Adjusted Logistic Regression, FALR).

Суть покращення логарифма функції правдоподібності для логістичної регресії полягає у включенні в загальну суму двох дуальних доданків для кожної відхиленої заявки, якій відповідає значення ймовірності, а не бінарний результат.

При відсутності відхилених заявок у моделюванні покращений логарифм функції правдоподібності дорівнює класичному, що є однією з найголовніших властивостей введеного методу FALR:

$$\begin{aligned}
& L^*(\mathbf{c}) \mid i \in \{m+1, \dots, m+n\} = \\
& = \sum_{i=m+1}^{m+n} \left\{ (1 - y_i) \ln \left(\frac{1}{1 + \exp \left\{ - \left(c_0 + \sum_{j=1}^k c_j x_{ij} \right) \right\}} \right) \right\} + \\
& + y_i \ln \left(\frac{1}{1 + \exp \left\{ - \left(c_0 + \sum_{j=1}^k c_j x_{ij} \right) \right\}} \right) \Bigg\} = L(\mathbf{c}).
\end{aligned}$$

Ця складова частина методу аналогічна кроку максимізації ЕМ-алгоритму (Maximization Step).

Загальна суть вдосконалення методу ітеративної класифікації

Розглянемо варіант комплексного застосування запропонованого методу аналізу характеристик і методу покращення логістичної регресії.

За основу беремо метод ітеративного аналізу відхилених заявок (Iterative Reclassification Method) [3]. Кроки методу ітеративного аналізу відхилених заявок із врахуванням методу покращення WOE та модифікації логістичної регресії будуть відрізнятися від класичного методу ітеративної класифікації (Iterative Reclassification Method), де насамперед поняття ітеративної класифікації необхідно замінити на поняття повторного оцінювання ймовірностей – назовемо його методом ітеративного обчислення ймовірностей (Iterative Probability Recalculation Method):

1) побудова скорингової моделі лише на прийнятих заявках;

2) прогнозування ймовірностей p_i для всіх m відхилених заявок за допомогою моделі, побудованої лише на прийнятих заявках, аналогічно до кроку підрахунку математичного сподівання (Expectation Step) в ідеї ЕМ-алгоритму;

3) перерахування нечітко покращеного WOE за допомогою запропонованої формули FAWOE із врахуванням ймовірностей p_i відхилених заявок, аналогічно до кроку підрахунку математичного сподівання (Expectation Step) в ідеї EM-алгоритму. Паралельно повторно оцінюються значення FAIV для змінних;

4) повторне оцінювання коефіцієнтів c_j логістичної регресії за допомогою врахування ймовірностей відхилених заявок процедурою FALR, аналогічно до кроку максимізації (Maximization Step) в ідеї EM-алгоритму;

5) прогнозування ймовірностей p_i для всіх m відхилених заявок за допомогою моделі, побудованої на прийнятих заявках і на відхилених заявках, аналогічно до кроку підрахунку математичного сподівання (Expectation Step) в ідеї EM-алгоритму;

б) повертаємось на крок 3.

Умовою завершення процесу оцінювання коефіцієнтів можна вважати збіжність вектора коефіцієнтів логістичної регресії.

У запропонованому методі ітеративного обчислення ймовірностей (Iterative Probability Recalculation Method) кроки 1 і 2 відповідають ініціалізації ймовірностей відхилених заявок, крок 3 – удосконаленому методу аналізу характеристик (див. рис. 2), крок 4 – удосконаленому методу моделювання за допомогою логістичної регресії (рис. 2), кроки 5 та 6 – зворотному зв'язку. Кроки 3–6 – ітераціям на рис. 2. У класичному методі ітеративної класифікації (Iterative Reclassification Method) на кроках прогнозування – 2 і 5 – використовуються бінарні оцінки, що обчислюються на основі прогнозованих ймовірностей з використанням порогу відсікання (Cut-Off Rate). Також у всьому алгоритмі, замість нечітко покращених (Fuzzy Adjusted) оцінок і методів, використовуються класичні, визначені лише на бінарних оцінках, що є головним недоліком класичного методу порівняно із запропонованим.

Висновки

Запропоновано вдосконалення методу ітеративної класифікації – метод ітеративного обчислення ймовірностей для процедури аналізу відхилених заявок, що включає дві складові частини: вдосконалений метод аналізу характеристик та вдосконалений метод моделювання за допомогою логістичної регресії.

Введено поняття нечітко покращеної ваги значення змінної, нечітко покращеного значення інформації та запропоновано вдосконалення методу дискретизації й аналізу характеристик для перерахунку показників ваги значення змінної і значення інформації.

Запропоновано метод нечітко покращеної логістичної регресії, що є вдосконаленням методу моделювання за допомогою логістичної регресії через оцінювання максимальної правдоподібності з метою врахування підмножини відхилених заявок, що нечітко класифіковані.

Метод ітеративного обчислення ймовірностей має істотні переваги перед методом ітеративної класифікації. Однією з головних переваг запропонованого методу та його відмінністю від класичного методу ітеративної класифікації є дуальність відхилених заявок, тобто відсутність необхідності їх бінарної ітеративної класифікації з певним порогом відсікання, що полягає в оперуванні лише ймовірностями належності до класу клієнтів з позитивною кредитною історією, не виключаючи можливості їх належності також до класу клієнтів з негативною кредитною історією. На відміну від методу ітеративної класифікації, розроблений метод виключає необхідність підрахунку порогу відсікання для процедури ітеративної класифікації завдяки поняттю частково класифікованих даних, що застосовується до множини відхилених заявок.

Основними відмінностями введених понять FAWOE, FAIV та FALR перед класичними є визначеність на нечітких ймовірнісних оцінках відхилених заявок і тотожність відповідним класичним поняттям при відсутності нечітко класифікованих вхідних даних. Однією з головних властивостей введених понять є збереження зв'язку FAWOE і FAIV при розгляді узагальненого випадку з включенням відхилених заявок.

Метод ітеративного обчислення ймовірностей дає можливість розв'язувати задачі побудови аплікаційних скорингових моделей на якісно новому, узагальненому, ймовірнісно-бінарному рівні.

Перспективи подальших досліджень включають апробацію та теоретичне обґрунтування інших модифікацій методів аналізу відхилених заявок, зокрема, аналіз збіжності, стабільності та точності часткової класифікації даних, що полягає в оцінюванні штучної множини відхилених заявок з відомим бінарним результатом, а також аналіз якості прогнозів скорингових

моделей при включенні як частини прийнятих заявок, так і відхилених заявок як навчальної вибірки, при проведенні аналізу предикативності на інших прийнятих заявках, що включені в тестову вибірку.

Іншим напрямом перспективних досліджень є вдосконалення та узагальнення інших методів моделювання з бінарною цільовою змінною з метою включення ймовірнісних цільових оцінок і показників FAWOE в інші методи, відмінні від логістичної регресії [8].

Список літератури

1. *Энциклопедия* финансового риск-менеджмента: Учеб. пособие / Под ред. А.А. Лобанова и А.В. Чугунова. – М.: Альпина Паблишер, 2003. – 786 с.
2. *Lyn C. Thomas et al.*, Credit Scoring and its Applications: SIAM monographs on mathematical modeling and computation. Philadelphia: University City Science Center, SIAM, 2002, 248 p.
3. *Naeem Siddiqi*. Credit risk scorecards: developing and implementing intelligent credit scoring. Hoboken: John Wiley & Sons, Inc., 2006, 196 p.
4. *A.J. Feelders* “Credit scoring and reject inference with mixture models”, *Int. Journal of Intel. Sys. in Accounting, Finance and Management*, no. 8, pp. 271–279, 1999.
5. *Je-Men Mok*, Reject Inference in Credit Scoring. Amsterdam: BMI paper, 2009, 38 p.
6. *Руководство* по кредитному скорингу: Учеб. пособие / Под ред. Элизабет Мэйз; пер. с англ. И.М. Тикота; науч. ред. Д.И. Вороненко. – Минск: Гревцов Паблишер, 2008. – 464 с.
7. *Paul D. Allison*, Logistic Regression Using the SAS® System: Theory and Application. Cary, NC: SAS Institute Inc., North Carolina, 1999, 306 p.
8. *Бідюк П.І., Кузнецова Н.В., Терентьев О.М.* Система підтримки прийняття рішень для аналізу фінансових даних // *Наук. вісті НТУУ “КПІ”*. – 2011. – № 1. – С. 48–61.

Рекомендована Радою
Навчально-наукового комплексу
“Інститут прикладного системного
аналізу” НТУУ “КПІ”

Надійшла до редакції
22 жовтня 2013 року