

Електронне наукове фахове видання "Ефективна економіка" включено до переліку наукових фахових видань України з питань економіки (Наказ Міністерства освіти і науки України від 11.07.2019 № 975) [www. economy.nayka.com. ua](http://www.economy.nayka.com.ua) | № 12, 2019 | 27.12.2019 р.

DOI: [10.32702/2307-2105-2019.12.87](https://doi.org/10.32702/2307-2105-2019.12.87)

УДК 330.46:336.774

*I. В. Мірошніченко,
кандидат економічних наук, доцент кафедри економіко-математичного моделювання,
Державного вищого навчального закладу «Київський національний економічний
університет імені Вадима Гетьмана»
ORCID: 0000-0002-1307-7889
К. Г. Івлієва,
студентка 4 курсу економічної кібернетики,
Державного вищого навчального закладу «Київський національний економічний
університет імені Вадима Гетьмана»
ORCID: 0000-0001-5306-1075*

ОЦІНЮВАННЯ КРЕДИТНОГО РИЗИКУ МЕТОДАМИ МАШИННОГО НАВЧАННЯ

*I. Miroshnychenko
Ph.D. (Economics), Associate Professor of the Department of Economic and Mathematical
Modeling, State University Kyiv National Economic University named after Vadym Hetman,
K. Ivliieva
4th year student of economic cybernetics,
State University Kyiv National Economic University named after Vadym Hetman*

ASSESSING CREDIT RISK USING MACHINE LEARNING METHODS

Стаття присвячена вирішенню завдання мінімізації кредитних ризиків банку, що видає кредити юридичним особам. Проведений аналіз внутрішньої рейтингової моделі банку, розрахунок її дискримінаційної сили, на її основі побудована математична модель для прогнозування кредитоспроможності майбутніх позичальників та оцінка якості моделі, застосований алгоритм позбуття аномалій в даних та порівняно результати початкової та покращеної моделей. Запропонований підхід до оцінки кредитоспроможності позичальника дозволив покращити якість рейтингової моделі, а також її здатність розрізняти надійних та ненадійних клієнтів, та зменшив кількість дефолтних та не дефолтних клієнтів. Особливістю підходу, який представлений у статті, є застосування методів машинного навчання для аналізу та обробки даних у програмному забезпеченні R. Практичною цінністю проведеного дослідження є отримані знання щодо важливості правильного відбору даних, які характеризують рейтингову модель, що в подальшому слугуватиме одним із інструментів для попередження втрат від невиконання кредитних зобов'язань.

The article is devoted to solving the problem of minimizing credit risks of the bank issuing loans to legal persons. This study presents a process for the development and implementation of risk prediction model, one that builds upon a foundation of statistics, data mining and machine learning principles. Statistical and data mining techniques and methodologies have been discussed in detail. The key phases that were covered are: the analysis of the bank's internal

rating model, calculation of its discriminatory power, mathematical model based on the rating model was built to predict creditworthiness of future borrowers and assessment of model quality coefficients, the algorithm of data cleaning from anomalies and comparison of the model results was applied. The proposed approach to assessing the borrower's creditworthiness improved the quality of the rating model, as well as its ability to distinguish between reliable and unreliable clients, and reduced the number of default and no default clients. It has been defined that selected quality indexes of the classification models: Accuracy, Specificity, Sensitivity, Balanced accuracy, ROC-curve, AUROC and Gini index are higher for the improved model. A feature of the proposed approach is the application of machine learning methods to data analysis and processing in software R programming language.

The practical value of the conducted study is the acquired knowledge on the importance of correct selection of data characterizing the rating model, which in the future will serve as one of the tools for prevention of losses from non-fulfillment of credit obligations. There is a need to devise strategies to not only identify customers who exhibit negative behavior (nonpayment, fraud), but also deal with them effectively to minimize further loss and recoup any monies owed, as quickly as possible.

Ключові слова: *кредитний ризик; рейтингова модель; машинне навчання; дискримінаційна сила; тест Колмогорова-Смірнова; логістична регресія.*

Key words: *credit risk; rating model; machine learning; discriminatory power; Kolmogorov-Smirnov test; logistic regression.*

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями. Знання щодо ймовірності дефолту клієнта банку важко переоцінити, оскільки сьогодні на ринку банківських послуг збільшується попит на кредитні продукти. Зростання конкуренції на ринку банківських послуг вимагає від банків вдосконалення механізмів оцінювання кредитоспроможності потенційних позичальників з метою підвищення якості обслуговування клієнтів, мінімізації кредитних ризиків, вивільнення грошових коштів зі створених резервів.

Аналіз кредитоспроможності відіграє визначальну роль в процесі розгляду заявки на кредит і являє собою найбільш трудомісткий процес. Дані заходи дозволяють визначити проблемні кредити на ранній стадії їх виникнення, що в свою чергу підвищує можливості банку повернути свої кошти і запобігти можливим збиткам. Отже, якісний і комплексний аналіз кредитоспроможності позичальників — це запорука ефективної діяльності банку. Найбільш ефективним інструментом визначення ймовірності дефолту позичальника вважаються математичні та статистичні моделі, тому в даній роботі ми зосередимо увагу на них.

Аналіз останніх досліджень та публікацій. Моделі оцінки кредитного ризику позичальників досліджуються у працях вітчизняних та зарубіжних науковців: Н.В. Кузнецова [1], Г.В. Гаврилюк [2], А.Б. Камінський [3], Н. Сіддікі [4], Д.Дюрана [5].

Так, Н.В. Кузнецова розглянула інтегрований підхід до оцінювання кредитних ризиків на основі комбінації двох підходів — мережі Байеса і дерева рішень або мережі Байеса і бінарної логістичної регресії. Інтегрований підхід на основі мережі Байеса і дерева рішень показав найкращий результат з усіх моделей, розглянутих у роботі окремо.

Дослідження Г.В. Гаврилюк присвячене аналізу вагомості характеристик потенційних позичальників — фізичних осіб в оцінюванні їх кредитоспроможності. У роботі обґрунтовано доцільність застосування методу нечіткого багатокритеріального аналізу, за яким вибір найбільш кредитоспроможних позичальників за рядом критеріїв здійснюється на основі парних порівнянь альтернатив.

Стаття А. Б. Камінського присвячена аналізу та розбудові стратегій управління портфелем простроченої заборгованості за споживчими кредитами. Розглядається побудова скорингів за допомогою застосування технології карт самоорганізації (Self Organizing Maps, SOM) або карт Кохонена до навчальної вибірки, що дозволяють упоратися із задачею ідентифікації факторів та побудови скорингу пріоритезації зусиль.

У праці Н. Сіддікі «Credit Risk. Scorecards. Developing and Implementing Intelligent Credit Scoring» всебічно та детально розглянуто стратегічні карти як інструмент процесу управління кредитними ризиками, їх моделювання та процес розробки стратегічних карт.

Дослідження Д. Дюрана «Risk Elements in Consumer Instalment Financin» представляє глибокий аналіз факторів, які мають значення для визначення кредитних ризиків та визначення кредитних стандартів у сфері кредитування фізичних осіб.

Виділення не вирішених раніше частин загальної проблеми.

Грунтуючись на результатах досліджень зазначених вище авторів, можна стверджувати, що дискусія стосовно питання кредитоспроможності позичальників та пов'язаних з цим кредитних ризиків, не закінчена та потребує обговорювання. Нові практичні методи дослідження та передбачення кредитних ризиків еволюціонують. Пропонується дослідження кредитного ризику методами машинного навчання з можливістю подальшого застосування алгоритму у банках задля мінімізації кредитних ризиків.

Мета та завдання статті. Метою статті є систематизація вітчизняного та зарубіжного досвіду використання моделей визначення ймовірності дефолту позичальника в процесі оцінки кредитного ризику позичальника. Особлива увага приділяється практичному застосуванню методів машинного навчання для оцінки створеної рейтингової моделі банку. Досягнення мети дослідження потребувало вирішення ряду завдань, зокрема розкриття теоретичної сутності оцінки дефолту позичальника, побудова математичної моделі ймовірності дефолту на основі рейтингової моделі банку, пошук оптимальних способів оцінювання якості математичної моделі та глибокий аналіз задля виключення аномалій в моделі.

Виклад основного матеріалу дослідження. Проблема кількісної оцінки ризиків є одним з найважливіших ключових елементів управління кредитною організацією. Міжнародні регулюючі органи (Базельський комітет з банківського нагляду) пропонують банкам вибір: скористатися спрощеними підходами, але в шкodu точності, або впровадити в свої системи управління ризиками сучасні підходи.

«Високий рівень» підходу до оцінки кредитного ризику, вперше згаданий в стандартах Базеля II, отримав назву підходу на основі внутрішніх рейтингів позичальників (IRB-підхід). У загальних рисах IRB-підхід являє собою підхід до оцінки кредитних ризиків банків на основі використання внутрішніх рейтингів позичальників, тобто рейтингів, які встановлюють самі банки [6].

У нашому розпорядженні був готовий кредитний рейтинг, за яким банк видавав кредити МСП (малим та середнім підприємствам). Зазначений підхід оцінює платоспроможність позичальників від AAA (найкраща оцінка) до C (найгірша) (див. табл. 1).

Робота з дослідження даних проводилася за наступними етапами:

1. розрахунок та оцінка дискримінаційної сили моделі (*discriminatory power*);
2. моделювання наявних даних для подальшого прогнозування статусу майбутнього позичальника;
3. оцінка якості побудованої моделі;
4. глибокий аналіз даних та виключення аномалій для покращення якості моделі.

Таблиця 1.

Фрагмент початкових даних (якісних) та модифікованих (кількісних)

Financial Rating	Non-Financial Rating	Final Rating	Default Flag		Financial Rating	Non-Financial Rating	Final Rating	Default Flag
CC	A	CCC	No Default		3	8	4	0
CC	AA	B	Default		3	9	5	1
CC	AAA	B	Default		3	10	5	1
CC	AAA	B	No Default		3	10	5	0
CCC	BB	B	Default		4	6	5	1

Дані кредитного рейтингу складалась з таких компонентів:

- **Фінансовий рейтинг (*finr*)**, що базується на фінансових показниках позичальника з межами зміни від 1 до 10.
- **Нефінансовий рейтинг (*nfinr*)** базується на якісній інформації про позичальника з межами зміни від 1 до 10.
- **Фінальний рейтинг (*fr*)** включає фінансовий та нефінансовий рейтинг, зважений на 70% та 30% відповідно. Фінальний рейтинг використовується банком у процесі кредитування для прийняття кредитних рішень та моніторингу кредитів. Також, відомий **статус клієнта (*df*)**, що оформив кредит – «дефолт» чи «не дефолт». Це змінна, яка підлягає оцінюванню, грунтуючись на трьох факторах, описаних вище.

У якості інструменту для обробки, аналізу та моделювання даних було використано мова програмування *R*.

Розрахунок та оцінка дискримінаційної сили моделі

Процедура оцінювання дискримінаційної сили моделі дає змогу зрозуміти, наскільки добре створена рейтингова модель та чи спроможна вона класифікувати позичальників на кредитоспроможних та некредитоспроможних.

Для підтвердження припущень про нормальний розподіл показника *fr*, який є основним для прийняття рішення про статус кредитоспроможності клієнта, був використаний тест Колмогорова-Смірнова,

що базувався на гіпотезі H_0 : показники компоненти fr належать до однієї генеральної сукупності. Спростування гіпотези H_0 буде означати, що рейтингова модель здатна правильно розподіляти клієнтів на «дефолтних» і «не дефолтних». За умови правильності нерівності (1), гіпотеза H_0 відхиляється [6].

$$KS > c(\alpha) \sqrt{\frac{n+n'}{nn'}} \quad (1)$$

$$KS = \max |F^{good} - F^{bad}|$$

де α – рівень значущості; $c(\alpha)$ – табличний коефіцієнт, що залежить від α (α —довірчий рівень, обраний за 0,05); n – розмір вибірки дефолтних позичальників (7427); n' – розмір вибірки недефолтних позичальників (2001); F^{good} , F^{bad} – кумулятивні функції розподілу недефолтних та дефолтних позичальників за fr відповідно.

Отже, було сформовано дві вибірки: F^{good} , яка містить всі фінальні рейтингові бали (fr) та має статус дефолту (df) «1» та F^{bad} , яка містить всі фінальні рейтингові бали (fr) та має статус не дефолту (df) «0» [7]. Візуалізацію теста можна побачити на рис. 1.

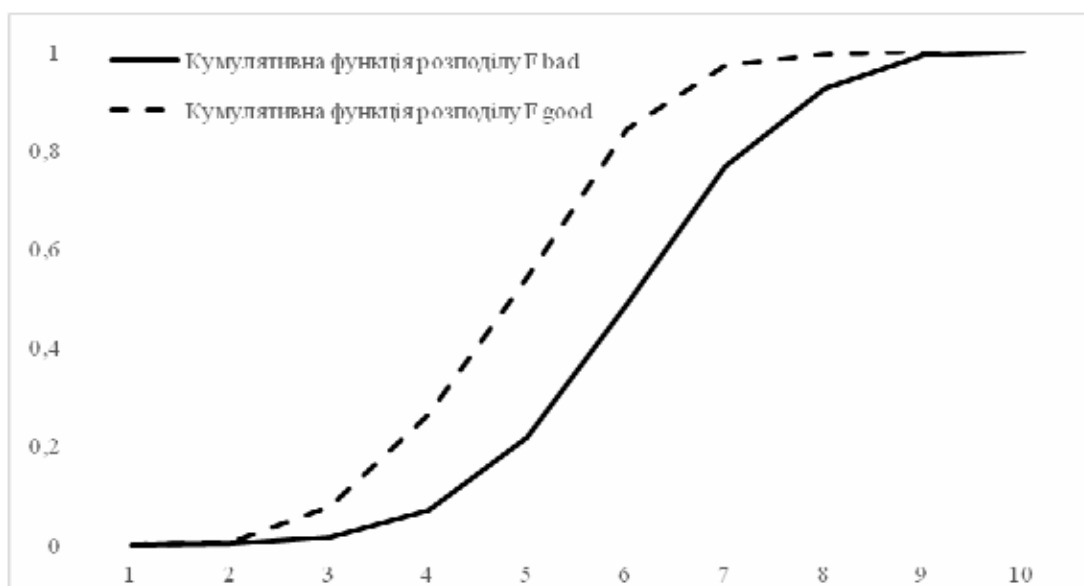


Рис. 1. Графік кумулятивних функцій до КС тесту.

Джерело: розраховано за [7]

За допомогою мови програмування R був проведений розрахунок лівої частини нерівності (1) (функція ks.test) [8], права частина розраховувалась за наявних даних вибірки. За нашими розрахунками:

$$KS = 0,034744, \text{ в той час, як } (\alpha) \sqrt{\frac{n+n'}{nn'}} = 0,030728.$$

Можемо зробити висновок, що гіпотеза H_0 відхиляється. Оскільки нульовий гіпотезою тесту Колмогорова-Смірнова є рівність розподілів двох вибірок, то це означає, що існуючий кредитний рейтинг здатний відрізнити «поганих» і «хороших» позичальників, тому що розподіли вибірок не рівні.

Моделювання наявних даних для подальшого прогнозування статусу майбутнього позичальника

Ми пропонуємо побудувати математичну модель, яка дала б змогу оцінити статус клієнта, виходячи з попередніх даних його рейтингів, на основі бінарної класифікації вихідної змінної. Для цього було використано математичну модель логістичної регресії:

$$p = \frac{e^{(\beta_0 + \beta_0 x_1 + \dots + \beta_0 x_n)}}{1 + e^{(\beta_0 + \beta_0 x_1 + \dots + \beta_0 x_n)'}}$$

де p – ймовірність дефолту позичальника, x_i – пояснювальні змінні, β_i – коефіцієнти регресії пояснювальних змінних, n – кількість пояснювальних змінних.

Логістична регресія — це вид нелінійної множинної регресії, яка аналізує функціональну залежність між декількома незалежними змінними (регресорами) і залежною змінною. Бінарна логістична регресія застосовується у тому випадку, коли вихідна змінна може приймати тільки два значення [1].

При значенні $p = 0,5$ прийнято вважати, що клієнт має рівні шанси стати “банкрутом” або “не банкрутом”. Чим ближче значення логістичної функції до $p = 1$, тим вище ймовірність банкрутства компанії та навпаки [9].

Для кожного з клієнтів банку з наявних даних відомий його статус – дефолт (1) або не дефолт (0). Мета полягає в тому, щоб знайти такі коефіцієнти β_0, \dots, β_n , за яких модельні значення ймовірності дефолту відповідали спостережуваній ймовірності дефолту. За пояснювальні змінні було обрано три змінні: фінансовий рейтинг ($finr$), нефінансовий рейтинг ($nfinr$) та фінальний рейтинг (fr); вихідна змінна – статус клієнта (df). У машинному навчанні прийнято розділяти дані про тренувальні та тестові набори: якщо ви намагаєтеся оцінити свою систему на базі даних, на якій ви навчили модель, ви не отримаєте достовірні результати. Важливо працювати з новими даними.

Весь масив даних було розділено на навчальну (80%) та тестову (20%) вибірки. Так, щоб у кожній множині частка дефолтів була приблизно однакова [10].

Після оцінювання параметрів моделі, математична модель набула вигляду:

$$P^I = (df = 1 | finr, nfinr, fr) = F\left(\frac{e^{2,337 - 0,666 * finr + 0,003 * nfinr + 0,011 * fr}}{1 + e^{2,337 - 0,666 * finr + 0,003 * nfinr + 0,011 * fr}}\right)$$

Оцінка якості побудованої моделі

Перевірка якості моделі відбувалася на тестовій вибірці. Основними критеріями ефективності математичної моделі були обрані показники чутливості (Sensitivity), специфічності (Specificity), точності (Accuracy), ROC-крива (ROC curve) та коефіцієнт Джині (Gini coefficient) [4].

Ці показники розраховуються на основі так званої матриці помилок (confusion matrix). Кращим результатом класифікації є такий, для якого кількість правильно класифікованих випадків максимальна, а кількість неправильно класифікованих випадків мінімізована (див табл. 2).

Таблиця 2.
Загальний вигляд матриці помилок

		Актуальні значення	
		TRUE	FALSE
Предиковані	TRUE	True Positives (TP) (істинно позитивне)	False Negatives (FN) (хибно негативне)
	FALSE	False Positives (FP) (хибно позитивне)	True Negatives (TN) (істинно негативне)

З матриці помилок (табл.3), для нашої моделі бачимо, що завдяки моделі ми можемо передбачити 1 779 «не дефолтів» від 1 845 їх загальної кількості і лише 107 «дефолтів» від 512 загальної кількості. Це не є хорошим результатом прогнозування [4].

Таблиця 3.
Матриця помилок для побудованої моделі бінарної логістичної регресії

		Актуальні значення	
		TRUE	FALSE
Предиковані	TRUE	1779	66
	FALSE	405	107

Чутливість (Sensitivity)—частка істинно позитивних випадків, які були правильно ідентифіковані моделлю.

$$Sensitivity = \frac{TP}{TP + FN}$$

Показник Sensitivity дорівнює 0,82, що є досить високим, але це означає, що модель може прогнозувати лише 82% “не дефолтів” правильно.

Специфічність (Specificity) — частка істинно негативних випадків, які були правильно ідентифіковані моделлю.

$$Specificity = \frac{TN}{TN + FP}$$

Показник Specificity становить 0,61. Це означає, що модель може прогнозувати лише 61% «дефолтів» правильно.

Точність (Accuracy) - це частка справжніх результатів (як істинних позитивних, так і істинних негативних) серед загальної кількості досліджених випадків (Total cases).

$$\text{Accuracy} = \frac{TN + TP}{\text{TotalCases}}$$

Збалансована Точність (Balanced Accuracy)—показник, який розраховується у випадку, якщо вибірка не пропорційна, тобто різна кількість спостережень у класах.

$$\text{Balanced Accuracy} = \frac{1}{2} \left(\frac{TP}{P} + \frac{TN}{N} \right)$$

Показник дорівнює 0,7165, що не погано, але недостатньо, щоб сказати, що модель хорошої якості.

ROC-крива (Receiver Operating Characteristic) – графік, що показує ефективність моделі бінарного вибору (див. рис. 2). ROC-крива показує співвідношення між TP та FP за прийнятою лінією розділу (cut-off). Величина cut-off для зручності була обрана 0,5 (тобто, якщо показник >0,5 – позичальник отримує статус дефолтного, менше – не дефолтного).

Нанесемо на інтервалі $(-\infty; +\infty)$ відповідні $X = 1 - \text{Specificity}$ та $Y = \text{Sensitivity}$ – це і буде ROC-крива, яка завжди приймає значення $[0;1]$.

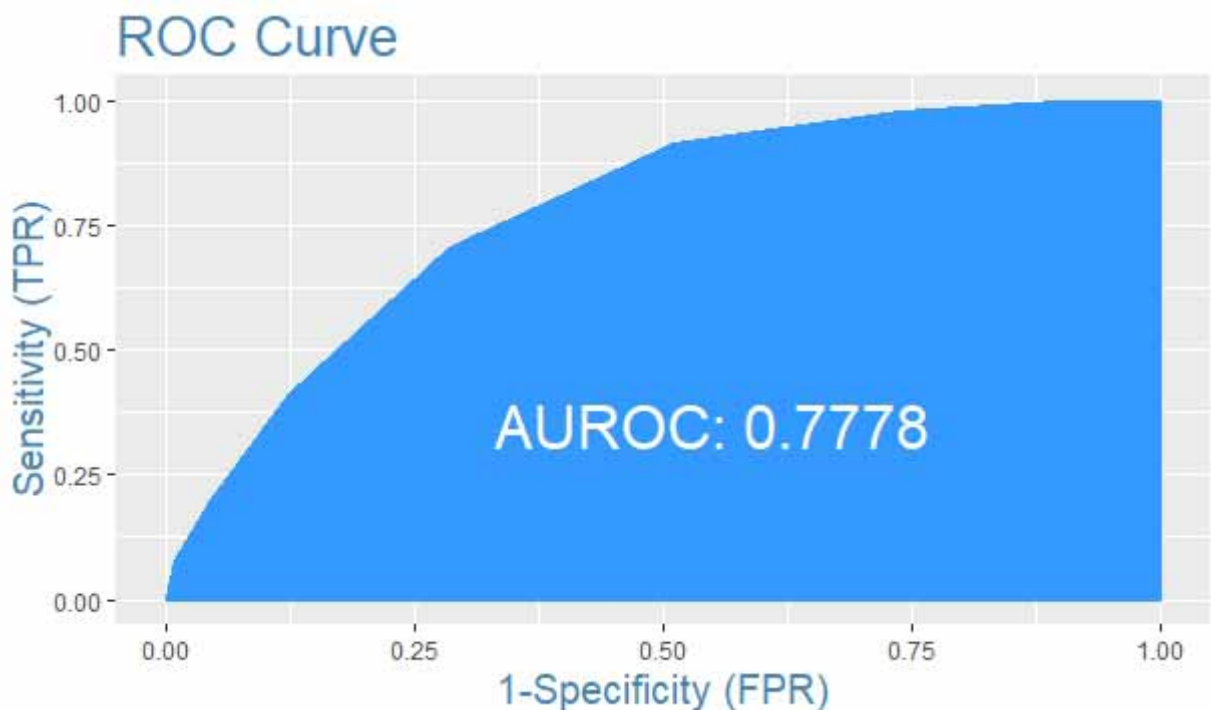


Рис. 2. Графік ROC-кривої

AUC (Area under the ROC Curve) – площа під ROC-кривою, міра дискримінаційної потужності моделі $[0,5;1]$.

$$AUC = \int f(x)dx = \left[\sum_i \frac{x_i + x_{i+1}}{2} \right] \cdot (Y_{i+1} + Y_i)$$

Отриманий показник дорівнює 0,7789, що говорить про непогану дискримінаційну силу моделі.

Коефіцієнт Джині (коефіцієнт точності). Цей коефіцієнт оцінює ступінь наближеності валідованої рейтингової системи до ідеальної. Під ідеальною системою мається на увазі така, яка безпомилково визначає, чи буде позичальник мати статус дефолтного.

$$\text{Gini} = 2 * AUC - 1 = 0.5578$$

Показник Джині є відносно низьким, тож можна зробити висновок, що предиктивна сила моделі низька [6].

Глибокий аналіз даних та виключення аномалій для покращення якості моделі

З метою покращення моделі був проведений більш глибокий аналіз даних, завдяки якому виявлено розбіжності в даних: однаковим «вхідним даним» відповідають різні «вихідні». Ті самі клієнти мали різний статус за умови однакових рейтингів, що погано впливає на можливість побудованої моделі виявляти дефолт та коректно класифікувати позичальників (таблиця 3).

Таблиця 3.
Приклад розбіжностей класифікації клієнтів за однакових вхідних показників

Financial Rating	Non-Financial Rating	Final Rating	Default Flag
BB	AAA	BBB	Default
BB	AAA	BBB	No Default

Для того, щоб скорегувати початковий набір даних, що мають негативний вплив на навчання моделі та її якість, було розроблено алгоритм чистки даних. Якщо статус «дефолт» у групі однакових вхідних даних з'являється частіше, ніж статус «не дефолт», то зберігаються значення вхідних показників зі статусом «дефолт», в той час як всі інші альтернативні значення видаляються з дослідження, як аномальні. Аналогічна процедура проводиться для альтернативних груп показників [11]. За результатом такої чистки даних, вибірка зменшилась на 45%, тобто залишились лише унікальні значення показників без аномальних даних.

Після очистки даних було побудовано нову математичну модель та оцінено її якість.

$$P^{II} = (df = 1 | finr, nfinr, fr) = F \left(\frac{e^{18,358 - 4,082 * finr + 2,723 * nfinr + 6,147 * fr}}{1 + e^{18,358 - 4,082 * finr + 2,723 * nfinr + 6,147 * fr}} \right)$$

Порівняльна таблиця результатів наведена у таблиці 4.

Таблиця 4.
Порівняння основних показників якості початкової та покращеної моделі

Показники якості	Початкова модель	Покращена модель
<i>Accuracy</i>	0,8002	0,9985
<i>Sensitivity</i>	0,8146	0,9992
<i>Specificity</i>	0,6185	0,9688
<i>Balanced Accuracy</i>	0,7165	0,9840
<i>AUROC</i>	0,7789	0,9996
<i>Gini</i>	0,5578	0,9992

Завдяки описаній процедурі очистки даних, які негативно впливають на модель та в результаті знаходження коефіцієнтів показників рейтингу, можна прогнозувати статус дефолту майбутніх позичальників, мінімізуючи кредитні ризики. Похибка становить 3%. Тобто, статус позичальника буде визначений з точністю 97%, що дає можливість ефективно оцінювати потенційних клієнтів та мінімізувати кредитні ризики банку.

Висновки. Проведено дослідження внутрішньої рейтингової моделі банку, розраховано її дискримінаційну силу, побудовано на її основі логіт-модель, розроблений алгоритм виключення аномалій з моделі та порівняно результати показників якості першої моделі та покращеної. Даний підхід показав свою ефективність застосування до проблеми мінімізації кредитних ризиків.

Важливо впроваджувати в діяльність банків сучасні більш продуктивні підходи до вирішення проблеми обробки великих даних, аналізу та валідації рейтингових моделей. Кредитні ризики несуть у собі найбільшу небезпеку для комерційних банків у контексті забезпечення та збереження їх фінансової стійкості, тому впровадження нових, більш ефективних методів оцінки, управління, а отже, попередження кредитних ризиків має стати пріоритетним напрямом розвитку банківської системи.

Перспективи подальших розвідок у даному напрямі. Позбуття незбалансованості даних, тобто нерівномірності розподілу класів у вибірці. Плануємо застосувати алгоритм SMOTE, який штучно генерує нові дані класу меншин, використовуючи метод найближчих сусідів для цих випадків.

Література.

- Кузнєцова Н.В. Інтегрований підхід до оцінювання кредитних ризиків. *Труди Одеського політехнічного університету*. 2010. №1-2. с. 187–192.
- Гаврилюк Г. В. Аналіз вагомості критеріїв в оцінюванні кредитоспроможності фізичних осіб. *Нейро-нечіткі технології моделювання в економіці*. 2017. №6. с. 3–23.

3. Камінський А. Б., Сікач В. О. Нейромережеві технології в управлінні портфелем простроченої заборгованості. *Моделювання та інформаційні системи в економіці*. 2011. №84. с. 5–19.
4. Siddiqi N. *Credit Risk Scorecards Developing and Implementing Intelligent Credit Scoring*. John Wiley & Sons, Inc. 2006.
5. Durand, D. *Risk Elements in Consumer Instalment Financing*, National Bureau of Economic Research, USA, New York. 1941.
6. Стежкин А.А. О методах валидации рейтинговых систем в рамках подхода внутренних рейтингов к оценке кредитного риска банков. *Финансовая аналитика: проблемы и решения*. 2015. №32. С. 29-41.
7. Řezáč M., Řezáč F., How to Measure the Quality of Credit Scoring Models. *Czech Journal of Economics and Finance*. 2011. №61(5). pp. 486-507.
8. Marsaglia G., Tsang W.W., Wang J., Evaluating Kolmogorov's distribution. *Journal of Statistical Software*. 2003. 8/18.
9. Сушко В.І., Павлюк Т.С. Класифікація моделей оцінки ймовірності банкрутства підприємств. *Економіка: теорія та практика*. 2014. №1. с.72 - 83.
10. Dobbin K.K., Richard M.S. Optimally splitting cases for training and testing high dimensional classifiers. *BMC Medical Genomics*. 2011. № 4(1):31.
11. Van den Broeck J., Herbst A.J., Solveig A.C. Data Cleaning: Detecting, Diagnosing, and Editing Data Abnormalities. *PLoS Medicine*. 2005. №2(10). pp. 966–970.

References.

1. Kuznyetsova, N.V. (2010), “Integrated approach to estimation of credit risks”, *Trudi Odesskogo politehnicheskogo Universiteta*, vol. 1-2, pp. 187- 192.
2. Havrylyuk, H. V. (2017), “Analysis of the significance of the criteria in assessing the creditworthiness of individuals”, *Nejro-nechitki tekhnolohii modeliuвання v ekonomitsi*, vol. 6, pp. 3–23.
3. Kamins'kyj, A.B. and Sikach, V.O. (2011), “Neural network technologies in managing the portfolio of overdue debt”, *Modeliuвання ta informatsijni systemy v ekonomitsi*, vol. 84, pp. 5–19.
4. Siddiqi, N. (2006), *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*, John Wiley and Sons, USA, New Jersey.
5. Durand, D. (1941), *Risk Elements in Consumer Instalment Financing*, National Bureau of Economic Research, USA, New York.
6. Stezhkin, A.A. (2015), “About the validation methods for rating systems within the framework of internal ratings to assess the credit risk of banks”, *Finansovaja analitika: problemy i reshenija*, vol. 32, pp. 29-41.
7. Řezáč, M. and Řezáč, F. (2011), “How to Measure the Quality of Credit Scoring Models”, *Czech Journal of Economics and Finance*, vol. 61(5) pp. 486-507.
8. Marsaglia, G., Tsang, W.W. and Wang J. (2003), “Evaluating Kolmogorov's distribution”, *Journal of Statistical Software*, vol. 8/18.
9. Sushko, V.I. and Pavliuk, T.S. (2014), “Classification of models for assessing the probability of bankruptcy of enterprises”, *Ekonomika: teoriia ta praktyka*, vol.1, pp. 72–83.
10. Dobbin, K.K. and Richard, M.S. (2011), “Optimally splitting cases for training and testing high dimensional classifiers”, *BMC Medical Genomics*, vol. 4:31.
11. Van den Broeck, J., Herbst, A.J. and Solveig, A.C. (2005), “Data Cleaning: Detecting, Diagnosing, and Editing Data Abnormalities”, *PLoS Medicine*, vol.2(10), pp. 966–970.

Стаття надійшла до редакції 18.12.2019 р.