

УДК 519.21:519.6

<https://doi.org/10.17721/1812-5409.2023/2.5>

В.П. Зубченко¹, к.ф.-м.н, доц.
А.В. Авраменко²

V.P. Zubchenko¹, PhD, Associate Professor
A.V. Avramenko²

Дослідження скорингової моделі для кредитопозичальників банку

Investigation of the scoring model for bank borrowers

^{1,2}Київський національний університет імені
Тараса Шевченка, 03680, м. Київ, пр-т.
Академіка Глушкова 4д,
e-mail: volodymyr.zubchenko@knu.ua,
²avramenkoanton23@gmail.com

^{1,2}Taras Shevchenko National University of Kyiv,
03680, Kyiv, Glushkova av., 4d,
e-mail: volodymyr.zubchenko@knu.ua,
²avramenkoanton23@gmail.com

В роботі досліджено скорингові моделі як інструмент управління кредитними ризиками, їх типи та особливості. Розглянуто поняття кредитоспроможності та скорингу, типи скорингу та їх особливості, методи скорингових моделей (логістична регресія, дерева рішень, лінійне програмування, дерева рішень та інші) та етапи їх побудови. В роботі наведені основні формули, що використовуються для оцінки кредитоспроможності позичальників та підвищення точності скорингових моделей, включаючи індекс стабільності популяції, коефіцієнт R^2 , коефіцієнт Колмогорова-Смірнова та інші; розглянуто практичне застосування скорингових моделей в банках України на прикладі типового позичальника.

Ключові слова: кредитний скоринг, кредитоспроможність, платоспроможність, оцінка кредитоспроможності, позичальники, скорингова карта, класифікація ознак, методи створення, машинне навчання, тестування моделі.

In the paper we investigate scoring models as a tool for credit risk management, their importance, types, features and applications. We consider the history of scoring models and the development of the modern concepts of creditworthiness and scoring, examine scoring types and their features, methods (logistic regression, decision trees, linear programming, decision trees, and others), strengths and weaknesses of each considered method and stages of building scoring models; we stress the importance of scoring cards for building scoring models, indicate the main quantitative and qualitative features, their classification used for and describe the procedure of creating scoring cards. In the paper we consider the factors needed to be considered for more effective scoring model building process; we indicate the main formulas used to assess the creditworthiness of borrowers and improve the accuracy of scoring models, including Population Stability Index (PSI), R-Square Coefficient, Kolmogorov-Smirnov Coefficient, GINI Coefficient, and others; the practical application of scoring models in banks of Ukraine and their steps after applying scoring models are described on the example of a typical borrower.

Keywords: credit scoring, creditworthiness, solvency, creditworthiness assessment, borrowers, scoring card, classification of features, creation methods, machine learning, model testing.

Вступ

В наш час кредитування є однією з ключових економічних складових, оскільки для банків воно є основним джерелом прибутку. В сучасному світі спостерігається збільшення обсягів споживчого кредитування. Проте разом з цим, зростає і кількість неповернених кредитів, що наносить великі збитки банківським установам. Саме через це важливою задачею є мінімізація кредитних ризиків для зменшення цих збитків. З цією метою було розроблено кілька методів оцінювання кредитних ризиків, одним з яких і є кредитний скоринг.

У 1940 році Національне бюро економічних досліджень США опублікувало роботу Ральфа Янга «Особисті фінансові компанії та їх кредитна практика», в якій вперше були висловлені ідеї скорингу. Однак, зараз про цю публікацію знає лише вузьке коло фахівців, а «винахідником» скорингу вважається Девід Дюран, який адаптував розроблену Гансом Фішером методику класифікації рослин для класифікації кредитів на «погані» і «хороші». У 1941 р. Національне бюро економічних досліджень США замовило дослідницький проект, в якому Дюран на основі статистики банків і інших фінансових організацій проаналізував сотню позитивних і негативних кредитних історій, використав дискримінантний аналіз та розробив індивідуальну систему оцінки ступеня важливості параметрів.

Модель Д. Дюрана складалась з групи чинників для визначення ступеня кредитного ризику та коефіцієнтів для різних факторів, що характеризують кредитоспроможність клієнта.

У 1956 році американці – інженер Вільям Фер і математик Ерл Айзек, які працювали в Стенфордському дослідницькому інституті, описали першу кредитну скорингову модель. Вони розробили математичний алгоритм, що дозволяв прораховувати кредитні ризики у вигляді трьохзначного числа, яке є кредитним рейтингом, і таким чином визначав рівень кредитоспроможності позичальника. Вони заснували компанію Fair, Isaac and Company (FICO), яка досі є одним із лідерів серед розробників скорингових систем.

На початку 1960-х років з появою кредитних карток, системи кредитного скорингу набули більшого поширення. Великий потік клієнтів, які зверталися за кредитними картками, разом з дефіцитом трудових ресурсів привів до автоматизації кредитного процесу. Використання кредитного скорингу в банках

зменшило кількість дефолтів по позикам на 50% в порівнянні з використанням експертної думки, яку використовували до цього.

Успіх застосування скорингу у видачі кредитних карток забезпечив на початку 80-х років перехід методу і на інші банківські продукти, такі як персональні кредити, іпотечні кредити та кредити для малого бізнесу. Досягнення в області обчислювальної техніки відкрили можливості для побудови скорингових моделей новими способами. Були розроблені та впроваджені основні математичні методи, які використовуються і сьогодні: лінійне програмування, дерева рішень та логістична регресія.

Мета роботи – розкрити поняття кредитоспроможності та скорингу, розглянути типи скорингу та їх особливості; розглянути методи (логістична регресія, дерева рішень та лінійне програмування) та етапи побудови скорингових моделей, навести основні формули, що використовуються для оцінки кредитоспроможності позичальників та підвищення точності скорингових моделей; описати практичне застосування скорингових моделей в банках України на прикладі конкретного позичальника.

1. Загальна інформація про предмет дослідження

Основним критерієм, що визначає можливість надання банком кредиту, є кредитоспроможність та платоспроможність позичальника, після перевірки яких банки визначають фактори, що впливають на можливість неповернення кредиту.

Кредитоспроможність - наявність у клієнта передумов для проведення активної операції (отримання кредиту) і його спроможність повернути кредит у повному обсязі та в обумовлені договором строки.

Банки (фінансові установи) при оцінці кредитоспроможності враховують наступні фактори:

- ✓ фінансовий стан потенційного позичальника;
- ✓ боргове навантаження – співвідношення існуючих зобов'язань і запитуваної позики до основного доходу заявника;
- ✓ кредитну репутацію клієнта;
- ✓ вартість майна, яке є у власності позичальника;

- ✓ соціальний статус заявника, його особистісні характеристики, просування кар'єрними сходами та інші фактори.

Платоспроможність - це здатність боржника банку здійснювати розрахунки за всіма видами своїх зобов'язань відповідно до умов договорів. [1]

Кредитоспроможність позичальника, на відміну від його платоспроможності, не фіксує неплатежі за минулий період чи на певну дату, а прогнозує здатність до погашення боргу на найближчу перспективу. Отже, кредитоспроможність - це якісна оцінка позичальника, яка дається банком для розгляду питання про можливість і умови кредитування і дозволяє передбачити ймовірність своєчасного повернення позик та їх ефективного використання. [2]

Для визначення кредитоспроможності та платоспроможності кредитопозичальника банки застосовують метод кредитного скорингу.

Кредитний скоринг - математична або статистична модель, створена на підставі кредитної історії минулих клієнтів за допомогою якої банк намагається визначити, наскільки велика вірогідність того, що окремих потенційний позичальник поверне кредит в обумовлений строк; систематизована шкала значень якісних та кількісних показників, а також балів, що використовується для оцінки ймовірності повернення кредиту.

В залежності від мети проведення оцінки виділяють такі 4 типи кредитного скорингу:

- 1) *аплікаційний скоринг* (application scoring) - використовується на початку кредитних відносин, коли позичальник отримує кредит. [3] Дана скорингова модель є методом оцінки потенційних позичальників для прийняття рішення щодо надання кредиту та трансформує в кількісні значення ті ризики, які пов'язані з правильною оцінкою соціальних, демографічних, фінансових та інших даних позичальника. Її використання дозволяє отримати найбільш об'єктивну оцінку на підставі статистично підтверджених закономірностей, а не суб'єктивної думки певного кола осіб;
- 2) *поведінковий скоринг* (behavioral scoring) є частиною кредитного моніторингу протягом кредитного періоду. [3] Даний скоринг є оцінкою ймовірності повернення позичальником

вже існуючих кредитів, яка заснована на даних про історію транзакцій по його рахунках, його кредитній історії, фінансовому стані, соціальних, демографічних та інших даних;

- 3) *колекторський скоринг* (collection scoring) - оцінка ймовірності повного або часткового повернення кредиту. [3] Даний вид скорингу використовується при порушенні позичальником терміну договору та здійснюється для визначення пріоритетних заходів щодо поверненням позичальником кредиту;
- 4) *шахрайський скоринг* (fraud scoring) - статистична оцінка вірогідності шахрайських дій з боку потенційного позичальника. [3]

Скорингові моделі є інструментом оцінки *кредитного ризику*, тобто ймовірності виникнення збитків або додаткових втрат або недоотримання запланованих доходів унаслідок невиконання боржником узятих на себе зобов'язань відповідно до умов договору. [5]

Скорингові моделі застосовуються з метою:

- ✓ прискорення процедури оцінки позичальника і, як наслідок, зменшення терміну розгляду заявки та прийняття рішення;
- ✓ зменшення обсягу неповернутих кредитів та зменшення обсягів сформованих (відповідно до нормативних вимог) резервів під можливі втрати за кредитними зобов'язаннями відповідно;
- ✓ зниження впливу людського фактору на прийняття рішень за кредитними запитами;
- ✓ формування повної та якісної бази даних позичальників тощо.

Етапами побудови скорингової моделі можна зазначити:

- ⇒ визначення основних характеристик та припущень;
- ⇒ підготовка вхідних даних для скорингової моделі, перевірка їх якості і достатності;
- ⇒ побудова моделі;
- ⇒ тестування моделі та оцінка коректності її результатів;
- ⇒ опис та затвердження моделі;
- ⇒ впровадження моделі;
- ⇒ моніторинг моделі та, за необхідності, внесення змін.

2. Методика побудови скорингових моделей

2.1. Методи і моделі для вирішення задачі оцінювання кредитних ризиків

Основними методами, що застосовуються для створення скорингових моделей, є лінійне програмування, дерева рішень та логістична регресія.

Лінійне програмування в контексті оцінки кредитоспроможності клієнта використовується для побудови лінійної ймовірнісної моделі.

Лінійна ймовірнісна модель – модель, у якій ймовірність виникнення певної події (наприклад, виплати кредиту або затримки у платежах) моделюється як лінійна функція від набору показників або змінних. Ці показники можуть включати різні характеристики кредитопозичальників, такі як вік, зайнятість, дохід, кредитна історія та інші.

Лінійну ймовірнісну модель можна представити наступним чином:

$$P(Y = 1|X) = a_0 + a_1X_1 + a_2X_2 + \dots + a_kX_k,$$

де $P(Y = 1|X)$ представляє ймовірність події Y (наприклад, виплата кредиту) при заданих значеннях показників $X_1, X_2, \dots, X_k$. $a_0, a_1, a_2, \dots, a_k$ є коефіцієнтами моделі, які визначають вплив кожного показника на ймовірність виникнення події. $X_1, X_2, \dots, X_k$ є значеннями показників кредитопозичальників.

Лінійні ймовірнісні моделі є простими в реалізації та інтерпретації та досить стійкими до викидів даних, проте мають низьку ефективність при роботі з даними, які характеризуються нелінійними залежностями.

Логістична регресія – це статистичний регресійний метод, що застосовують у випадку, коли залежна змінна є категорійною (може набувати тільки два значення), тобто за допомогою логістичної регресії можна оцінити ймовірність того, що подія наступить для конкретного дослідження, наприклад, чи буде дефолт по кредиту, чи ні. В останні роки логістична регресія стала найбільш поширеним методом побудови скорингових моделей для розрахунку рейтингу позичальників і управління кредитними ризиками.

Логістична регресія дозволяє оцінити ймовірність того, що відбудеться якась конкретна подія для певного випробування (наприклад, повернення кредиту).

Логістична регресія описується формулою:

$$F(X) = \frac{1}{1 + e^{-(w_0 + w_1x_1 + \dots + w_nx_n)}}, \text{ де}$$

$x_1, x_2, \dots, x_n$ – незалежні змінні/показники;
 $w_1, w_2, \dots, w_n$ – коефіцієнти логістичної регресії;
 n – кількість незалежних змінних/показників.

Існує кілька методів знаходження коефіцієнтів логістичної регресії. Найчастіше використовують метод максимальної правдоподібності. Його використовують в статистиці, щоб одержати оцінки параметрів генеральної сукупності за даними вибірки. В основі методу знаходиться функція правдоподібності (likelihood function), яка виражає щільність ймовірності спільної появи результатів вибірки.

До переваг логістичної регресії можна віднести ефективність в навчанні та зрозумілість в інтерпретації результатів. Окрім цього, в порівнянні з лінійною регресією логістична регресія потребує менше вхідних даних, бо для її проведення потрібно менше припущень. При цьому, одним із обмежень логістичної регресії є те, що вона не дозволяє описати складні нелінійні зв'язки між незалежними та залежними змінними. Іншими недоліками логістичної регресії є те, що вона потребує великого обсягу навчальних даних і є чутливою до викидів.[4]

Дерева рішень представляють собою структуру, зазвичай створену за допомогою алгоритмів машинного навчання, яка моделює прийняття рішень шляхом розгалуження на основі набору правил та умов. Дерево рішень можна проілюструвати як ієрархічну структуру, в якій спостереження класифікуються на основі тверджень «якщо... то...». У деревах рішень вибір класифікаторів та їх граничних значень здійснюється автоматично на основі їх статистичної важливості. Дерево рішень починається з найважливішого класифікатора (кореневого вузла дерева), який розбиває всі залежні значення на дві групи (істина або хиб). Усі наступні класифікатори (оператори if) називаються внутрішніми вузлами. Залежні значення присвоюються відповідним групам операторів then, які називаються листками.

Ключовою перевагою дерев рішень є їх зрозуміле представлення у того, як були отримані результати моделі. Ієрархічна та послідовна структура дерев рішень дозволяє проілюструвати нелінійні зв'язки між незалежними та залежними змінними. Ще однією перевагою застосування дерева рішень є те, що цей метод є ефективним у роботі з відсутніми значеннями. Головний недолік дерев рішень полягає в тому, що вони можуть бути

нестійкими до перенавчання та змін у навчальних даних – це означає, що невеликі коригування вхідних даних можуть кардинально змінити структуру дерева рішень.[5]

Останнім часом набирають популярність нові підходи до побудови скорингових моделей, що використовують можливості штучного інтелекту, такі як нейронні мережі та метод випадкового лісу.

Нейронні мережі – інструмент нелінійного машинного навчання. Вони складаються з вхідного, прихованого та вихідного шарів нейронів. Навчальні дані передаються через вхідні нейрони до нейронів прихованих шарів. На наступному кроці нейрони прихованих шарів намагаються розпізнати деякі закономірності в даних та зв'язки між ознаками та залежною змінною. Після обробки цих даних вони передаються далі на вихідний рівень. Оцінки вихідного рівня порівнюються з фактичними даними та обчислюється похибка моделі.

Нейронні мережі здатні виявляти складні нелінійні моделі даних з великою кількістю параметрів і спостережень. З іншого боку, навчання нейронних мереж потребує багато часу і великих обчислювальних потужностей, а результати створених ними моделей є складними для інтерпретації.[5]

Випадкові ліси — це група методів машинного навчання, які використовуються для задач класифікації та регресії. Основна ідея методу випадкових лісів полягає в побудові набору дерев рішень із випадково вибраними підвибірками даних та ознак. Кожне дерево в лісі незалежно конструюється шляхом випадкового вибору підвибірки даних (bootstrapping) та випадкового вибору підмножини ознак. Випадковий вибір ознак є другим методом, який використовується випадковими лісами для диверсифікації дерев рішень. У цьому підході не всі параметри використовуються в навчанні дерева рішень. Кожне з дерев рішень випадкового лісу буде побудовано на навчальній вибірці з частиною випадково вибраних незалежних змінних. Як наслідок, випадковий вибір параметрів моделі не дозволяє моделі надавати надмірно великого значення певним параметрам та нехтувати іншими.

Метод випадкового лісу має кілька переваг і недоліків. Цей підхід дозволяє зменшити дисперсію окремих дерев рішень і запобігти переобладнанню моделі. Ще одна

перевага випадкового лісу полягає в тому, що він може обробляти відсутні значення. Крім того, метод випадкового лісу не вимагає попередньої нормалізації змінних або вибору ознак і ефективно обробляє набори даних із великою кількістю вхідних змінних. Найбільшим недоліком цього підходу є висока складність аналізу моделі через велику кількість випадкових дерев та потреба у великих обчислювальних потужностях.[5]

2.2. Визначення основних характеристик, підготовка вхідних даних, перевірка їх якості

На даному етапі побудови скорингової моделі здійснюється обробка і аналіз характеристик-показників про поведінку аналогічних позичальників в минулому, експертних знань, значної кількості даних, отриманих з різних (в т. ч. зовнішніх) джерел. Показники, що використовуються, поділяються на кількісні та якісні.

До складу якісних показників можуть включатись, зокрема, але не виключно, наступні показники:

- сімейний стан;
- вік позичальника;
- освіта позичальника;
- стаж роботи;
- стаж роботи на останньому місці роботи;
- посада;
- наявність/відсутність власного майна та/або нерухомості.

До складу кількісних показників можуть включатись, зокрема, але не виключно наступні показники:

- коефіцієнт платоспроможності позичальника;
- розмір доходу позичальника;
- показник сукупного чистого доходу;
- частка погашення кредиту в чистому доході позичальника;
- коефіцієнт боргового навантаження (відношення місячного платежу позичальника за кредитом до його місячного доходу).

Для формування повноцінної вибірки та суттєвого підвищення якості даних, здійснюється цілеспрямований, систематичний збір і аналіз даних.

На основі зібраних даних створюється *скорингова карта*, тобто набір характеристик (вік, дохід, професія, досвід роботи, наявність майна тощо) позичальника та певні ваги, виражені в балах. Клієнту банку нараховується

певна кількість балів залежно від інформації, яку він повідомив про себе.

Максимальна сума кредиту, яку банк готовий надати позичальнику, розраховується залежно від кількості балів. Логістична регресія є найбільш поширеною статистичною моделлю оцінювання бінарної залежної змінної. Оцінка коефіцієнтів логістичної регресії як балів є головним кроком у розробці карти балів.

В подальшому здійснюється класифікація всіх позичальників на три основні групи:

- a) хороші (далі - Goods);
- b) погані (далі - Bads);
- c) невизначені (далі - Indetermined).

Класифікація позичальників залежить від мети, відповідно до якої розроблюється скорингова модель (behavioral scoring, fraud scoring, collection scoring або application scoring).

Позичальники, по яких відсутні випадки виникнення простроченої заборгованості для Application scoring класифікуються, як «Goods». Позичальники, за кредитними операціями яких виникла прострочена заборгованість понад 90 днів для Application scoring, класифікуються, як «Bads». Як Bads також можуть бути визначені позичальники по яких відбулось списання кредитної заборгованості за рахунок сформованого (відповідно нормативних вимог) резерву, рефінансування (пов'язаного з фінансовими труднощами позичальника), погашення кредитної заборгованості тощо.

Позичальники, які не можуть бути класифіковані як «Goods» та «Bads» класифікуються для Application scoring, як «Indetermined» та не враховуються при розробці скорингової моделі. До Indetermined, наприклад, можливо віднести позичальників за якими виникла прострочена заборгованість від 30 до 60 днів.

Для остаточного визначення «Bads» при здійсненні класифікації позичальників може бути застосований один із трьох методів, в залежності від доступних даних:

- ✓ метод консенсусу (спільне обговорення різними зацікавленими службами банку для досягнення консенсусу щодо найкращого визначення «Bads», виходячи з досвіду та операційної діяльності відповідних служби);
- ✓ статистичний метод, який може будуватися на аналізі змін простроченої заборгованості (Roll Rate) з метою визначення рівня простроченої

заборгованості, після якого більшість позичальників будуть проблемними (так званої «точки неповернення») та порівняльному аналізу рівня поточної простроченої заборгованості з найгіршим;

- ✓ комбінація методу консенсусу та статистичного методу.

Одним з перших кроків при побудові скорингової моделі та формування вибірки є визначення періоду в минулому, за який беруться позичальники для аналізу («вікно спостереження») та визначення кінцевої точки, тобто дати, станом на яку визначається статус позичальника.

З метою підвищення ефективності процесу розробки скорингових моделей попередній вибір показників необхідно здійснювати, враховуючи наступні фактори:

- надійність та стабільність;
- зручність збору даних;
- інтерпретованість;
- доступність в майбутньому;
- розрахунок коефіцієнтів (будь-який отриманий коефіцієнт повинен бути обґрунтований).

Для побудови скорингової моделі вибірка має бути істотною і достатньою, а саме: 1500 – 2000 кейсів кожного з класів (Goods/Bads) буде достатньо для побудови скорингових моделей високої якості. Продуктивність скорингової моделі істотно погіршується у випадку, коли кількість кейсів менша за 100 кожного з класів позичальників. В протипагу, вибірка, що включає більше 5000 кожного з класів (Goods/Bads), не має значного ефекту від її використання при побудові скорингової моделі.

2.3. Масштабування та тестування

Для зручності використання та інтерпретації результатів оцінки необхідно застосовувати масштабування. В ході масштабування підбирається діапазон і формат балів скорингової моделі, а також швидкість зміни позитивних шансів у міру зростання балів.

Формат скорингових балів може бути різним, включаючи десятковий або дискретний вираз, наприклад:

- a) бали можуть виражати ймовірність того, що конкретний Позичальник буде поганим;

- б) бали можуть у вигляді числової шкали з заданим мінімум/максимум (наприклад, від 0 до 1, від 1 до 100).

Після розробки скорингової моделі необхідно заміряти її прогностичну здатність та коректність її побудови. Фінальні бали по кредитним запитам необхідно розподілити по кластерах за принципом кривої дистрибуції (розподіл Гаусса, або нормальний розподіл) з інтуїтивно логічним розподілом Approval Rate та Bad Rate (чим нижчий скоринговий бал, тим вищий ризик і нижчий Approval Rate, та навпаки – чим вищий скоринговий бал, тим нижчий ризик і вищий Approval Rate; Тренд Bad Rate повинен бути максимально рівним). При цьому відхилення розподілу вліво або вправо є допустимими за умови, якщо сильні поведінкові показники сумісно були використані в скоринговій моделі, наприклад, показники з кредитного запиту.

Кластери скорингових балів повинні бути сегментовані з врахуванням потреб бізнесу та завдання скорингу. Наприклад, для Application scoring скорингові бали можуть бути розподілені на:

- 1) Зелену зону (автоматичне погодження);
- 2) Сіру зону (потребує більш глибокого ручного аналізу);
- 3) Червону зону (автоматична відмова).

Існує декілька статистичних показників для виявлення відповідності моделі стандарту:

1. Індекс стабільності популяції (Population Stability Index, PSI);
2. Коефіцієнт R^2 ;
3. Log Likelihood;
4. Коефіцієнт GINI.

Індекс стабільності популяції показує величину зміни популяції між вибіркою недавніх позичальників та очікуваними величинами (виходячи з навчальної вибірки) та розраховується за формулою:

$$SI = \sum (\% \text{ фактичний} - \% \text{ очікуваний}) * \ln \left(\frac{\% \text{ фактичний}}{\% \text{ очікуваний}} \right), \text{ де}$$

% фактичний, % очікуваний – розподіл випадків для недавньої вибірки позичальників і навчальної вибірки відповідно для кожного показника.

Чим нижче значення SI, тим стабільніший показник. Найкраще значення прагне до 0.

Коефіцієнт R^2 демонструє точність наближення прогнозованого значення до вихідних даних та розраховується за наступною формулою:

$$R^2 = \frac{\text{Var}[\hat{y}]}{\text{Var}[y]} = \frac{\sum_{i=1}^N (\hat{y}_i - \widehat{y_{mean}})^2}{\sum_{i=1}^N (y_i - y_{mean})^2},$$

де $\hat{y}_i$ – прогнозне значення; y_i – первісне (фактичне) значення. Чим краще модель, тим вищий R^2 , значення якого змінюється з 0 до 1.

Log Likelihood використовується для оцінки якості моделі тільки з точки зору бінарної класифікації. Його можна порівняти з натуральним логарифмом функції Likelihood на передбачуваний вибірці.

$$\begin{aligned} \text{Log Likelihood} &= \\ &= \sum_{i, \text{where } y_i=1} \ln \left(\frac{1}{1 + e^{-(\hat{w}_0 + \hat{w}_1 x_{i1} + \hat{w}_2 x_{i2} + \dots + \hat{w}_n x_{in})}} \right) + \\ &+ \sum_{i, \text{where } y_i=0} \ln \left(\frac{1}{1 + e^{-(\hat{w}_0 + \hat{w}_1 x_{i1} + \hat{w}_2 x_{i2} + \dots + \hat{w}_n x_{in})}} \right) - \ln \left(\frac{1}{1 + e^{-(\hat{w}_0 + \hat{w}_1 x_{i1} + \hat{w}_2 x_{i2} + \dots + \hat{w}_n x_{in})}} \right) \end{aligned}$$

Log Likelihood завжди нижче нуля. Краща модель має вищий Log Likelihood (абсолютне значення Log Likelihood повинно наближатися до нуля).

Коефіцієнт GINI – це міра оцінки якості скорингової моделі. Дозволяє судити про здатність скорингової моделі відрізнати «хороших» Позичальників від «поганих» та розраховується за формулою:

$$GINI = \frac{\sum_{i=1}^n (CB_i + CB_{i-1}) * SB_i G_i - \frac{1}{2} * \frac{TotalB}{Total}}{\frac{1}{2} * \frac{TotalB}{Total}}, \text{ де}$$

CB_i – сукупна частка «Bads» в i – й Bin від загальної кількості «Bads»;

$SB_i G_i$ – частка «Bads» та «Goods» в i – й Bin від загальної кількості «Bads» та «Goods»;

TotalB – загальна кількість «Bads»;

Total – загальна кількість «Bads» та «Goods».

Найвище значення коефіцієнту GINI є кращим, тобто більш високе значення даного коефіцієнту показує, що конкретна скорингова модель краще відрізняє «Bads» і «Goods». Найкраще значення прагне до 100% (або 1).

Дуже високий GINI не є реалістичним та, скоріш за все, вказує на те, що модель треба переробити.

Коефіцієнт GINI порівнює криву Лоренца (Lorenz Curve) з лінією досконалої випадковості (Line of perfect randomness). На рисунку нижче, GINI – це відношення площі під кривою Лоренца і вище лінії досконалої випадковості (A) до всієї площі над лінією досконалої випадковості (A + B). (рис.1)

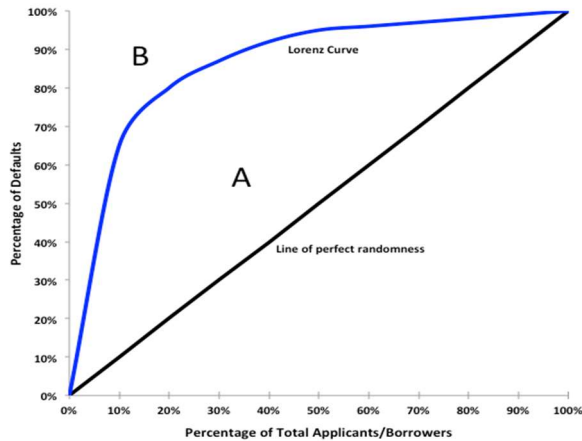


Рисунок 1

Коефіцієнт Колмогорова-Смірнова (КС) є показником розподілу для бінарної класифікації. Коефіцієнт Колмогорова-Смірнова визначається за формулою:

$KS = |F_g(score) - F_b(score)|$, де $F_g(score)$, $F_b(score)$ - кумулятивний розподіл скорингового бала для «Goods» і «Bads».

Теоретично коефіцієнт Колмогорова-Смірнова може приймати значення від 0 до 100, але на практиці при значенні КС > 70 вважається, що є проблеми в розрахунках або із самою моделлю.

3. Практичне застосування скорингових моделей в банках України

Розглянемо на прикладі кейс одного з банків України.

В банк «Стабільність» звернувся потенційний позичальник Петренко Петро Іванович із запитом на проведення активної операції, а саме - кредитування придбання нового транспортного засобу, при цьому запитуваний кредит складає 1 200 000 грн. на 84 місяці під 15%.

Довідково: Петренко П.І. 1975р.н. проживає в м. Кропивницький, 12 років працює лікарем-стоматологом в ТОВ «Smile 32», одружений в 2011р., має у власності будинок. Згідно довідки про доходи за останні 6 місяців його доходи складають 306 000 грн. (доходи дружини не враховуються до розрахунку), щомісячні обов'язкові витрати (податкові, комунальні платежі) – 15 000 грн., рахунків в банку немає.

Банк «Стабільність» пропонує для фізичних осіб банківський продукт «Автокредитування» - кредит на придбання нового автомобіля.

Умови кредитування:

- строк - від 1 до 7 років;
- процентна ставка – від 0.01% до 18%.
- максимальна сума кредиту – до 2 000 000 грн.

При цьому банк використовує наступну експертну скорингову модель для оцінки платоспроможності потенційного позичальника за напрямом кредитування на придбання нового транспортного засобу (див. рис. 2)

Для оцінки фінансового стану позичальника банком використовуються наступні кількісні показники:

- коефіцієнт платоспроможності позичальника (відношення сукупного середньомісячного доходу позичальника до суми середньомісячних сукупних витрат позичальника та щомісячним платежем по кредиту);
- частка погашення кредиту в чистому сукупному доході позичальника (відношення щомісячного платежу по кредиту до середньомісячних сукупних чистих надходжень позичальника (доходи позичальника за мінусом його витрат);
- показник сукупного чистого доходу позичальника (визначається, як відношення залишку коштів (щомісячних доходів) на одну особу позичальника та членів його родини, що знаходяться на утриманні, після сплати постійних витрат та місячного платежу за кредитом до показника прожиткового мінімуму);
- показник накопичень позичальника на інших рахунках в банку (відношення загальної суми залишків грошових коштів на депозитних рахунках позичальника, відкритих в банках до суми кредиту);

До якісних показників, які використовуються банком при оцінці фінансового стану відносяться наступні показники:

- наявність власного майна у позичальника;
- наявність постійної роботи;
- загальний стаж роботи;
- сімейний стан;
- місце проживання;
- вік позичальника на момент погашення кредиту.

Таким чином, скорингова модель потенційного позичальника Петренка П.І. матиме наступний вигляд (див. рис. 3):

Коефіцієнт платоспроможності позичальника	Значення	1	1.1	1.2	1.3	1.4	1.6	1.8
	Кількість балів						170	
Частка погашення кредиту в сукупному доході позичальника	Значення	1	0.9	0.75	0.6	0.5	0.4	0.3
	Кількість балів				130			
Чистий дохід позичальника	Значення	0.5	0.7	0.9	1	1.2	1.3	1.4
	Кількість балів							160
Наявність накопичень на рахунках в банку	Значення	0	0.1	0.11	0.15	0.18	0.2	0.25
	Кількість балів	0						
Наявність власної нерухомості	Значення	не має	нерухомість/транспорт	житло	нерухомість та транспорт	житло та транспорт	нерухомість	власне майно
	Кількість балів	5	15	20	28	38	42	45
Наявність постійної роботи	Значення	безробітний	студент/пенсіонер	від 0,5 до 1 року	від 1 до 2 років	від 2 до 3 років	від 3 до 5 років	від 5 років
	Кількість балів	0	14	18	23	27	33	37
Стаж роботи	Значення	менше 1 року	від 1 року до 2 років	від 2 років до 3 років	від 3 до 5 років	від 5 до 10 років	від 10 до 15 років	більше 15 років
	Кількість балів	0	14	19	25	30	37	40
Сімейний стан	Значення	неодружений на утриманні більше 2 осіб	не-одружений	одружений на утриманні більше 3 осіб	неодружений на утриманні до 2 осіб	одружений на утриманні до 3 осіб	одружений на утриманні до 10 років	одружений більше 10 років
	Кількість балів	10	15	20	27	35	45	50
Місце проживання	Значення	без житла	оренда житла	житло обтяжене	житло родичів	житло у кредиті	власне житло подружжя	власне житло
	Кількість балів	0	14	20	28	33	36	40
Вік позичальника на момент погашення кредиту	Значення	25	30	35	30	45	55	65
	Кількість балів	16	22	27	44	50	35	30

Рисунок 2

Коефіцієнт платоспроможності позичальника	Значення	1	1.1	1.2	1.3	1.4	1.6	1.8
	Кількість балів						170	
Частка погашення кредиту в сукупному доході позичальника	Значення	1	0.9	0.75	0.6	0.5	0.4	0.3
	Кількість балів				130			
Чистий дохід позичальника	Значення	0.5	0.7	0.9	1	1.2	1.3	1.4
	Кількість балів							160
Наявність накопичень на рахунках в банку	Значення	0	0.1	0.11	0.15	0.18	0.2	0.25
	Кількість балів	0						
Наявність власної нерухомості	Значення	не має	нерухомість/транспорт	житло	нерухомість та транспорт	житло та транспорт	нерухомість	власне майно
	Кількість балів	5	15	20	28	38	42	45
Наявність постійної роботи	Значення	безробітний	студент/пенсіонер	від 0,5 до 1 року	від 1 до 2 років	від 2 до 3 років	від 3 до 5 років	від 5 років
	Кількість балів	0	14	18	23	27	33	37
Стаж роботи	Значення	менше 1 року	від 1 року до 2 років	від 2 років до 3 років	від 3 до 5 років	від 5 до 10 років	від 10 до 15 років	більше 15 років
	Кількість балів	0	14	19	25	30	37	40
Сімейний стан	Значення	неодружений на утриманні більше 2 осіб	не-одружений	одружений на утриманні більше 3 осіб	неодружений на утриманні до 2 осіб	одружений на утриманні до 3 осіб	одружений на утриманні до 10 років	одружений більше 10 років
	Кількість балів	10	15	20	27	35	45	50
Місце проживання	Значення	без житла	оренда житла	житло обтяжене	житло родичів	житло у кредиті	власне житло подружжя	власне житло
	Кількість балів	0	14	20	28	33	36	40
Вік позичальника на момент погашення кредиту	Значення	25	30	35	30	45	55	65
	Кількість балів	16	22	27	44	50	35	30

Рисунок 3

Загальний бал даного позичальника склав 679 од.

У спрощеному вигляді наступні етапи щодо кредитної операції виглядають таким чином: здійснюється аналіз отриманого загального балу конкретного позичальника згідно скорингової моделі. У разі невідповідності визначеного загального балу позичальника його мінімально необхідному розміру (встановленому в банку) кредитна заявка потенційного позичальника відхиляється.

У разі прийняття попереднього рішення про проведення кредитної операції, банком, відповідно до встановлених внутрішніх процедур та з урахуванням нормативних вимог НБУ, на підставі отриманих результатів скорингової моделі (величини показників) визначається кредитний рейтинг та клас позичальника? Після чого приймається остаточне рішення про проведення кредитної операції (укладаються договори, надаються кошти, формується кредитна справа позичальника тощо).

В подальшому здійснюється моніторинг кредитної операції протягом всього терміну кредитування та, відповідно до нормативно визначених термінів, проводиться оцінка фінансового стану позичальника і, за

необхідності, коригування його класу та кредитного рейтингу, враховуючи показники, які були отримані при проведенні поточних розрахунків.

Висновки

В ході роботи були розглянуті етапи побудови скорингових моделей та деякі методи, які використовують для цієї мети, такі як лінійні ймовірнісні моделі, логістична регресія, дерева рішень, нейронні мережі та метод випадкового лісу, кожен із яких має свої переваги та недоліки:

- Лінійні ймовірнісні моделі є простими у застосуванні та інтерпретації, але мають обмежену сферу використання, однак відрізняються низькою ефективністю при роботі з нелінійними залежностями даних;
- Логістична регресія здатна працювати з невеликою кількістю вхідних даних і є простою в інтерпретації, але є чутливою до викидів та потребує великої навчальної вибірки.
- Дерева рішень здатні проілюструвати нелінійні зв'язки між незалежними та залежними змінними та демонструють високу ефективність у роботі з відсутніми даними, але вразливі до перенавчання та змін у навчальних даних.
- Нейронні мережі здатні виявляти складні нелінійні моделі даних з великою кількістю параметрів і спостережень, але потребують великих обчислювальних потужностей і є складними для інтерпретації.
- Метод випадкового лісу ефективно обробляє набори даних із великою кількістю вхідних змінних, але є складним у застосуванні та потребує великих обчислювальних потужностей.

Загалом, скорингові моделі є необхідним інструментом для банків у процесі кредитування. Вибір конкретної моделі залежить від багатьох факторів, таких як тип даних, складність задачі, нормативні обмеження та потужність доступних обчислювальних ресурсів. Розуміння переваг та недоліків різних скорингових моделей допомагає банкам зробити обґрунтований вибір та досягти кращих результатів у визначенні кредитоспроможності клієнтів.

Список використаних джерел

1. Положення про визначення банками України розміру кредитного ризику за

активними банківськими операціями, затверджене постановою Правління НБУ від 30.06.2016р. №351, із змінами.

2. Вдовенко Л. О. Економічна сутність та значення кредитоспроможності підприємств // Облік і фінанси.– 2012. – № 1.

3. Бучко І. Є. Скоринг як метод зниження кредитного ризику банку // Вісник Університету банківської справи Національного банку України. – 2013. – №2.

4. Bohdan Popovych. Application of AI in Credit Scoring Modeling – Springer Gabler, 1st ed., 2022.

5. Положення про організацію системи управління ризиками в банках України та банківських групах, затверджене постановою Правління НБУ від 11.06.2018р. №64, із змінами.

6. Кузнєцова Н. В. Порівняльний аналіз характеристик моделей оцінювання ризиків кредитування / ред. Н. В. Кузнєцова, П. І. Бідюк / Наукові вісті НТУУ «КПІ», 2010. №1

References

1. Regulation on the determination by banks of Ukraine of the amount of credit risk for active banking operations, approved by the resolution of the NBU Board dated June 30, 2016. No. 351, with changes.
2. VDOVENKO L. O. Economic essence and the meaning of creditworthiness of enterprises // Accounting and finance. – 2012. – No. 1.
3. BUCHKO I. E. Scoring as a method of reducing the bank's credit risk // Bulletin of the University of Banking of the National Bank of Ukraine. – 2013. – No. 2.
4. BOHDAN POPOVYCH. Application of AI in Credit Scoring Modeling - Springer Gabler, 1st ed., 2022.
5. Regulations on the organization of the risk management system in banks of Ukraine and banking groups, approved by the resolution of the NBU Board dated June 11, 2018. No. 64, with changes.
6. KUZNETSOVA N. V. Comparative analysis of the characteristics of lending risk assessment models / ed. N. V. Kuznetsova, p. I. Bidyuk / Scientific news of NTUU "KPI", 2010. No. 1.

Надійшла до редколегії 08.08.2023