

выгоду от инноваций?

URL: <https://www.crn.ru/news/detail.php?ID=117807>.

6. Big data в промышленности: инновации, к которым придется привыкать. URL: <http://www.ogcs.com.ua/index.php/articles/121-big-data-v-promyshlennosti-innovatsii-k-kotorym-pridetsya-privykat>.

**Мірошниченко І.В.**

к. е. н.

**Івлієва К.Г.**

*ДВНЗ «Київський національний економічний університет  
імені Вадима Гетьмана», м. Київ*

### **ОЦІНЮВАННЯ КРЕДИТНОГО РИЗИКУ МЕТОДАМИ МАШИННОГО НАВЧАННЯ**

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 з 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.

У нашому розпорядженні була готовий кредитний рейтинг, за яким банк видавав кредити МСП (малим та середнім підприємствам). Робота з дослідження даних проводилася за наступними етапами: розрахунок та оцінка дискримінаційної сили моделі (*discriminatory power*); моделювання наявних даних для подальшого прогнозування статусу майбутнього позичальника; оцінка якості побудованої моделі; глибокий аналіз даних та їх чистка

для покращення якості моделі.

На вході ми маємо кредитний рейтинг (складений банком), який оцінює платоспроможність позичальників від AAA (найкраща оцінка) до C (найгірша) (див. табл. 1).

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 (*fnr*)**, що базується на фінансових показниках позичальника з межами зміни від 1 до 10.
- **Нефінансовий рейтинг (*nfnr*)** 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, ґрунтуючись на трьох факторах, описаних вище.

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

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

спроможних 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 «недефолтних». За умови правильності нерівності (1),  $H_0$  відхиляється [1].

$$KS > c(\alpha) \sqrt{\frac{n+n'}{nn'}}, \quad (1)$$
$$KS = \max |F^{good} - F^{bad}|$$

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

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

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

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.

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:

$$p = \frac{\exp(\beta_0 + \beta_1 \cdot x_1 + \dots + \beta_n \cdot x_n)}{1 + \exp(\beta_0 + \beta_1 \cdot x_1 + \dots + \beta_n \cdot x_n)}$$

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

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.

Для кожного з клієнтів банку з наявних даних відомий його статус – дефолт або не дефолт. Мета полягає в тому, щоб знайти такі коефіцієнти  $\beta_0, \dots, \beta_n$ , за яких модельні значення ймовірності дефолту відповідали спостережуваній ймовірності дефолту. За пояснювальні змінні було обрано три змінні: фінансовий рейтинг, нефінансовий рейтинг та фінальний рейтинг; вихідна змінна – статус клієнта. Весь масив даних було розділено на навчальну (80%) та тестову (20%) вибірку.

Після оцінювання параметрів моделі, математична модель набула вигляду:

$$P(df = 1 | finr, nfinr, fr) = F(2,337 - 0,666 \cdot finr + 0,003 \cdot nfinr + 0,011 \cdot fr)$$

Для оцінювання якості та адекватності побудованої моделі були розраховані показники, що наведено у таблиці 2.

Таблиця 2

**Показники ефективності математичної моделі**

Чутливість ( <i>sensitivity</i> )	Спецефічність ( <i>specificity</i> )	Точність ( <i>accuracy</i> )
0,82	0,61	0,80

З метою оцінювання якості математичної моделі та інтерпретації перерахованих показників побудовано ROC-криву (*Receiver Operating Characteristic*) та оцінено показник AUC (*Area under the ROC Curve*). Отриманий показник дорівнює 0,7789, що говорить про непогану дискримінаційну силу моделі. Крім цього, було розраховано коефіцієнт Джині (коефіцієнт точності):

$$Gini = 2 * AUC - 1 = 0.5578$$

Показник Джині є відносно низьким, тож можна узагальнити, що предиктивна сила моделі низька.

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**

<b>Financial Rating</b>	<b>Non-Financial Rating</b>	<b>Final Rating</b>	<b>Default Flag</b>
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. За результатом такої чистки даних, вибірка зменшилась на 24,5%.

Після очистки даних було побудовано нову математичну модель та оцінено її якість. Порівняльну таблицю результатів можна подивитися у таблиці 4.

**Порівняння основних показників якості початкової  
та покращеної моделі**

<b>Показники якості</b>	<b>Початкова модель</b>	<b>Покращена модель</b>
Accuracy	0,8002	0,9906
Sensitivity	0,8146	0,9959
Specificity	0,6185	0,8816
Balanced Accuracy	0,7165	0,9388
AUC	0,7789	0,9989
Gini	0,5578	0,9978

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.

**Список використаних джерел**

1. Стежкин А. А. О методах валидации рейтинговых систем в рамках подходовнутренних рейтингов к оценке кредитного риска банков / А. А. Стежкин // Финансовая аналитика: проблемы и решения. – 2015. URL: <https://cyberleninka.ru/article/v/o-metodah-validatsii-reytingovyh-sistem-v-ramkah-podhoda-vnutrennih-reytingov-k-otsenke-kreditnogo-riska-bankov>.

**Пийвода К.І.  
Петренко І.П.**  
к.е.н., доцент

*ДВНЗ «Київський національний економічний університет  
імені Вадима Гетьмана», м. Київ*

**КРАУДФАНДИНГ В ІНВЕСТУВАННІ**

Українська держава протягом тривалого часу прагне до економічного розвитку та процвітання. Необхідною умовою цього процесу є наявність ресурсів для фінансування нових проєктів, які у майбутньому зможуть принести позитивний соціальний та економічний ефект. Проте через нестабільність банківської системи, складну економічну ситуацію в країні та бойові дії на Сході України інвестори не хочуть вкладати кошти у проєкти, оскільки присутній великий