

ЗАСТОСУВАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ РАНЬОГО ВИЯВЛЕННЯ ПРОБЛЕМНИХ КРЕДИТІВ

Кушнірова Катерина, студентка
Національний авіаційний університет
Київ, Україна
ORCID ID 0000-0000-0000-0000
e-mail: 6253675@stud.nau.edu.ua

Науковий керівник: Касьянова Наталія, доктор економічних наук, професор
Національний авіаційний університет
Київ, Україна
ORCID ID 0000-0000-0000-0000
e-mail: nataliia.kasianova@npp.nau.edu.ua

У умовах сучасного банківського сектору раннє виявлення потенційно проблемних кредитів є критично важливим для мінімізації кредитних ризиків. Застосування методів машинного навчання дозволяє автоматизувати цей процес: розробка та оцінка ефективності моделей машинного навчання для раннього виявлення проблемних кредитів з використанням сучасних аналітичних інструментів.

Актуальність даного дослідження обумовлена декількома ключовими факторами:

1. Зростання конкуренції на ринку банківських послуг, що вимагає вдосконалення механізмів оцінки кредитоспроможності, підвищення якості обслуговування клієнтів, мінімізації кредитних ризиків, оптимізації резервування коштів.
2. Необхідність автоматизації процесів, оскільки аналіз кредитоспроможності є найбільш трудомістким процесом, традиційні методи оцінки часто не встигають за динамікою ринку, зростає обсяг даних, що потребують обробки, підвищуються вимоги до швидкості прийняття рішень
3. Важливість раннього виявлення проблемних кредитів для зменшення потенційних збитків банку, підвищення якості кредитного портфеля, оптимізації роботи з проблемною заборгованістю, покращення фінансової стійкості банку
4. Розвиток технологій машинного навчання, що надає нові можливості для автоматизації процесів, підвищення точності прогнозування, можливість обробки великих масивів даних, зниження впливу людського фактору.
5. Економічна ефективність впровадження, що забезпечує скорочення операційних витрат, оптимізацію використання людських ресурсів, підвищення якості кредитних рішень, зменшення втрат від проблемних кредитів.

У сучасних умовах математичні та статистичні моделі вважаються найбільш ефективним інструментом визначення ймовірності дефолту позичальника [1]. Це обумовлено їх здатністю:

- обробляти великі обсяги даних;
- враховувати множину факторів;
- виявляти неочевидні закономірності;
- надавати об'єктивні оцінки;

Використання сучасних аналітичних інструментів, таких як SPSS Statistics, дозволяє реалізувати комплексний підхід до аналізу кредитних ризиків та впровадити ефективну систему раннього виявлення проблемних кредитів. Таким чином, дослідження застосування методів машинного навчання для раннього виявлення проблемних кредитів є актуальним науково-практичним завданням, вирішення якого сприятиме підвищенню ефективності банківської діяльності та стабільності фінансового сектору в цілому.

Кластеризація, або кластерний аналіз — це статистична процедура, задача якої полягає в розбитті вибірки об'єктів на підмножини, що не перетинаються і називаються кластерами. Кожен кластер має складатися зі схожих об'єктів, а об'єкти різних кластерів мають істотно відрізнятися один від одного. Задача кластеризації відноситься до статистичної обробки, а також до широкого класу задач навчання без вчителя. Ще її можна описати через задачу класифікації.

Для забезпечення ефективності використання кластерного аналізу в скорингу необхідно забезпечити регулярний моніторинг та актуалізацію кластерів, проводити аналіз міграції клієнтів між сегментами, здійснювати валідацію скорингових моделей та навчання персоналу щодо інтерпретації результатів. При цьому основними викликами при впровадженні кластерного аналізу є необхідність значного обсягу якісних історичних даних, складність інтерпретації результатів, ризики перенавчання моделей та технічні обмеження інформаційних систем.

У дослідженні використано комплексний підхід, що поєднує:

- Метод K-means кластеризації для сегментації клієнтської бази
- Логістичну регресію для побудови скорингової моделі
- Інструментарій SPSS Statistics для аналізу та обробки даних

Вибірка дослідження включала 345 кредитних заявок, що містили наступні характеристики позичальників [1]:

- Соціально-демографічні показники
- Фінансові параметри
- Кредитну історію
- Поведінкові характеристики

У результаті проведеного аналізу виявлено 4 чітко диференційованих кластери позичальників, кожен з яких має власні характерні особливості та рівень кредитного ризику:

1. Кластер 1 (96 випадків) - "Стабільні позичальники середнього класу"
 - Середній дохід: 62,773 грн/міс
 - Середня сума кредиту: 275,355 грн
 - Тривалий стаж роботи (14 років)
2. Кластер 2 (39 випадків) - "Преміум позичальники"
 - Найвищий середній дохід серед масових кластерів: 82,765 грн/міс
 - Найбільші суми кредитів: 456,717 грн
 - Найдовший стаж роботи (16 років)
3. Кластер 3 (209 випадків) - "Масовий сегмент"
 - Найнижчий середній дохід: 46,868 грн/міс
 - Найменші суми кредитів: 106,838 грн
 - Середній стаж роботи (17 років)
4. Кластер 4 (1 випадок) - "VIP клієнт"
 - Аномально високий дохід: 675,000 грн/міс
 - Відносно невелика сума кредиту: 127,500 грн
 - Найменший стаж роботи (6 років)

Практичні рекомендації, надані відповідно для кожного кварталу, будуть наступні:

1. Диференційований підхід до моніторингу: щоквартальний моніторинг, піврічний моніторинг, щомісячний моніторинг, індивідуальний підхід.
2. Специфічні індикатори ризику для кожного кластера: співвідношення платежу до доходу, зміна рівня доходу, платіжна дисципліна, індивідуальні тригери.

За результатами кластерного аналізу 345 кредитних справ було виділено чотири основні кластери позичальників, які представлені у таблиці 1.

Проведена кластеризація дозволила виявити чіткі сегменти позичальників з різними профілями платоспроможності, що створює основу для розробки диференційованого підходу до оцінки та моніторингу кредитних ризиків.

Таблиця 1 - Розподіл платоспроможності клієнтів відповідно до кластерів

Кластер	Середньомісячний дохід (грн)	Середня сума кредиту (грн)	Стаж роботи (років)	Частка в загальній вибірці (%)	Рівень платоспроможності
1	62 773	275 355	13,5	27,8 (96 осіб)	Середній
2	82 765	456 717	15,6	11,3 (39 осіб)	Високий
3	46 868	106 838	14,6	60,6 (209 осіб)	Нижче середнього
4	675 000	127 500	6,0	0,3 (1 особа)	Дуже високий

Кожен кластер має власні характерні особливості, які необхідно враховувати при розробці кредитної політики та системи управління ризиками.

Кластерний аналіз дозволив виявити чіткі сегменти позичальників з різними профілями ризику, що дає можливість оптимізувати процес раннього виявлення проблемних кредитів через застосування специфічних стратегій моніторингу для кожного сегмента. Запропонована модель машинного навчання демонструє високу ефективність у ранньому виявленні проблемних кредитів. Використання SPSS Statistics дозволяє ефективно реалізувати весь процес від аналізу даних до впровадження моделі. Зокрема, встановлено чітку сегментацію клієнтської бази на чотири кластери з унікальними характеристиками, що вимагають диференційованого підходу до моніторингу. Демографічний аналіз виявив значне переважання бездітних позичальників, що може бути важливим фактором при розробці кредитних продуктів та оцінці ризиків.

Запропонована модель демонструє високу загальну точність (0,893) та помірну прогностичну здатність ($AUC = 0,688$), що свідчить про її практичну цінність для банківської установи. Запропонована система диференційованого моніторингу та специфічних індикаторів ризику для кожного кластера дозволяє оптимізувати процес управління кредитним портфелем та мінімізувати ризики виникнення проблемної заборгованості. Подальше вдосконалення моделі та впровадження рекомендованих заходів сприятиме підвищенню ефективності кредитної діяльності банку.

Література:

1. Семененко, Юрій. "ТЕХНОЛОГІЇ МАШИННОГО НАВЧАННЯ У КРЕДИТНОМУ СКОРИНГУ." Наукові перспективи, 46 (2024).
2. Сайт статистики і даних Kaggle. Режим доступу: <https://www.kaggle.com/>
3. Бегчекова, Анна Артемівна. "Аналіз кредитних операцій банківської установи та антикризове управління ними." (2023).
4. Hand, D. J., & Henley, W. E. (1997). Statistical Classification Methods in Consumer Credit Scoring: A Review. Journal of the Royal Statistical Society.
5. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning.
6. Про затвердження Положення про визначення банками України розміру кредитного ризику за активними банківськими операціями: Постанова НБУ від 30.06.2016 № 351. URL: <https://zakon.rada.gov.ua/laws/show/v0351500-16>
7. Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring.