ISSN 2078-6115. Вісник Львівського університету. Серія економічна. 2022. Випуск 63. С. 56–66
Visnyk of the Lviv University. Series Economics. 2023. Issue 63. P. 56–66

КІЛЬКІСНІ МЕТОДИ В ЕКОНОМІЦІ. ЦИФРОВА ЕКОНОМІКА

DOI: http://dx.doi.org/10.30970/ves.2022.63.0.6305

УДК 336.77.067:330.45
JEL C61; E41; G33

МОДЕЛЮВАННЯ КРЕДИТНОГО СКОРИНГУ МЕТОДАМИ МАШИННОГО НАВЧАННЯ

Мирослав Дацко[1], Назарій Друщак[2]

[1]Львівський національний університет імені Івана Франка,
79008, м. Львів, просп. Свободи, 18,
e-mail: myroslav.datsko@lnu.edu.ua; ORCID: 0000-0002-5181-4632

[2]Український Католицький Університет
e-mail: naz2001r@gmail.com; ORCID: 0000-0002-5056-3026

Анотація. *Кредитний скоринг можна вважати однією з найуспішніших сфер застосування і методологічною підмножиною інтелектуального аналізу даних. Проблему кредитного скорингу у рамках інтелектуального аналізу даних можна віднести до класу класифікаційних методів машинного навчання.*
Метою статті є виявлення та моделювання залежності між факторними змінними, які показують певні характеристики клієнта банку та на основі цього визначити його відповідність до певного класу позичальників.
У дослідженні використано методи моделювання кредитного скорингу з використанням машинного навчання.
Результати дослідження дозволили виділити, серед побудованих, дві моделі з найкращими характеристиками: модель «випадкових лісів» та TPOT модель.
За допомогою класифікаційного звіту та ROC кривої було сформульовано висновок про те, що обидві моделі досягають значення показника f1 рівного 0.71 для досліджуваної вибірки, що є цілком прийнятним результатом для задачі кредитного скорингу.
Ключові слова: *машинне навчання, кредитний скоринг, «випадкові ліси», TPOT модель, ROC крива.*

Постановка проблеми. Вплив банківської системи невпинно зростає разом із розвитком економіки та економічних відносин. Одним з основних факторів успішної діяльності банківських установ незаперечно є ефективне управління кредитним ризиком. З огляду на вище сказане оцінка потенційного повернення або неповернення кредиту особливо актуальна.

ISSN 2078-6115. Вісник Львівського університету. Серія економічна. 2022. Випуск 63

Паралельно банківській системі розвивається і ІТ сфера, зокрема можна виділити методи машинного навчання, котрі дозволяють із великої кількості даних, які утворюються у компаніях знаходити потрібну інформацію для вирішення певних проблем та оптимізації процесів. Методи машинного навчання набули надзвичайної популярності через зростання обчислювальних потужностей комп'ютерної техніки, що дозволило використовувати штучний інтелект у банківській сфері серед методів аналізу кредитного скорингу.

Скорингові системи є зручним інструментом оцінювання кредитоспроможності адже використання скорингу сприяє підвищенню швидкості та правильності прийняття рішень щодо видачі кредитів, що є важливим у реаліях сьогодення. Окрім того, скоринг може враховувати не тільки фінансові показники діяльності позичальника, але і якісні також.

Аналіз останніх досліджень і публікацій. Кредитна діяльність банків достатньо широко висвітлена у працях зарубіжних і вітчизняних науковців. Так оцінка та моделювання кредитних ризиків банку, зокрема, розглядаються у роботі Б. Кишакевича [4]. Кредитоспроможність фізичних осіб досліджується у праці Г. Гаврилюка [1], де автор звертає увагу на врахування вагомості критеріїв при оцінці клієнта. Дослідження кредитного ризик-менеджменту проводили зарубіжні та вітчизняні вчені Р. Андерсон [7], Л. Томас [13],тА. Камінський [2], К. Писанець [2]. Проблематику скорингового моделювання розглядають Ю. Клебан [3] та Н. Горошко [3] які пропонують використати методику бінінгу показників у поєднанні з методами машинного навчання. Використання методів машинного навчання досліджено також у роботі В.Філатова [6], у якій автор розглядає побудову алгоритму оцінки кредитного ризику позичальника.

Проте незважаючи на те, що теоретична та практична база для дослідження кредитного ризику у сфері моделювання скорингових систем та кредитного ризик-менеджменту достатньо добре розроблена можна констатувати також гетерогенність запропонованих вченими підходів. З огляду на сказане подальші пошуки у цій сфері і надалі є актуальними, а тому наукова дискусія стосовно питання розробки скорингової моделі оцінки позичальників та пов'язаних з цим кредитних ризиків потребує подальших пошуків, що є завданням цього дослідження.

Постановка завдання. Однією із передумов розвитку суспільства є можливість кредитування та функціонування кредитного ринку. Позика дає можливість вирішити питання браку фінансових ресурсів з одного боку, а з іншого дозволяє кредиторам вигідно розмістити надлишкові кошти або заощадження.

Метою роботи є виявлення та моделювання залежності між факторними змінними, які показують певні характеристики клієнта банку та на основі цього визначити його відповідність до певного класу позичальників. Поділ на класи дозволить покращити роботу банківських установ та зменшити ризики при видачі кредитів за допомогою математичних методів, які лежать в основі машинного навчання, та які набувають широкого поширення на сучасному етапі розвитку.

Виклад основного матеріалу дослідження. Кредит – категорія ринкової економіки, що відображає економічне життя суспільства. Кредит став незамінним атрибутом сучасного виробництва, що дозволяє забезпечити безперервність виробничого процесу та модернізувати виробничі потужності і впроваджувати інноваційні технології.

Банкам потрібно ефективно розприділяти свої грошові кошти, які вони використовують для видачі кредитів, тобто потрібно спрогнозувати, хто із потенційних позичальників поверне гроші без проблем, оскільки виникнення проблем із поверненням кредиту може призвести до додаткових витрат які понесе банк або втрати розміщених коштів. Можливим варіантом вирішення поставленої проблеми є класифікаційні моделі машинного навчання, котрі розглянуті у даній роботі.

Під скоринговою моделлю розуміють математичну або статистичну модель, яка використовуючи дані кредитної історії банку визначає імовірність того, що потенційний позичальник поверне кредит у встановлені банком терміни. У сучасних реаліях все частіше для проведення класифікації клієнтів використовують машинне навчання.

Для побудови моделей використано показники: стать, вік, сімейний стан, професія, зайнятість, річний дохід, бінарні показники, тощо. Для практичної реалізації клієнтську базу банку взято із сайту для проведення змагань із аналізу даних Kaggle [11]. Дані на сайт було викладено іноземним банком анонімно.

Відзначимо, що для подальшого калібрування побудованих моделей до вітчизняних реалій можна було б використати дані із Кредитного реєстру.

На думку В.Філатова використання Кредитного реєстру, дасть змогу з високою точністю прогнозувати вірогідність дефолту для позичальників-фізичних осіб з кредитним портфелем більше 400 000 грн. Цей сегмент повністю представлений в Кредитному реєстрі, тому є наявною інформація про всіх позичальників України. [6]

Для практичної реалізації моделей було обрано мову програмування Python та її інтелектуальну оболонку Jupyter Notebook [10], яка успішно використовується для проведення експериментальних досліджень у сфері аналізу даних.

За результатами проведених досліджень побудовано 6 моделей з використанням машинного навчання, для розв'язання задачі кредитного скорингу, зокрема логістична регресія, близьких сусідів, дерево прийняття рішень, «випадкові ліси», градієнтний бустинг та TPOT модель. Кожна із моделей має свої характеристики такі, як крива AUC-ROC та класифікаційний звіт, який включає показники точності (Precision), чутливості (Recall) і оцінку F1, які дають змогу зрозуміти на скільки модель є адекватною і чи можна її застосовувати у банківській діяльності при оцінці надання кредиту.

Точність ¾ це відношення правильно передбачених позитивних спостережень до загальної кількості передбачуваних позитивних спостережень.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

Чутливість ¾ це відношення правильно передбачених позитивних спостережень до всіх спостережень у фактичному класі ¾ так.

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

Оцінка F1 ¾ це середньозважене значення точності та відкликання.

ISSN 2078-6115. Вісник Львівського університету. Серія економічна. 2022. Випуск 63

$$F1\ score = 2 * (Recall * Precision)/(Recall + Precision) \quad (3)$$

Для дослідження моделей було обрано метод перебору за допомогою модуля Gridsearch для вибору оптимального набору гіперпараметрів моделей.

В результаті побудови та оцінки логістичної регресії, близьких сусідів, дерева прийняття рішень та градієнтного бустингу на обраному наборі даних зроблено висновок про недоцільне їх застосування через меншу точність отриманих результатів, оскільки недоцільно застосовувати модель із меншою точністю, навіть, якщо різниця складає всього 1%. Значення основних показників цих моделей наведено на рис. 1.

Назва моделі	accuracy	precision	recall	f1-score
Логічтична регресія	0.50	0.50	0.50	0.50
близькі сусіди	0.70	0.70	0.71	0.70
дерева прийняття рішень	0.69	0.69	0.70	0.69
градієнтний бустинг	0.70	0.71	0.71	0.71

Рис. 1. Показники оцінки моделей
Джерело: Власні розрахунки

На наступному кроці побудовано – модель «випадкових лісів». Це ансамбль дерев рішень, які, зазвичай, навчають методом «мішків». Загальна ідея методу мішків полягає в тому, що поєднання моделей навчання покращує загальний результат [8].

Однією з найбільших переваг випадкового лісу є його універсальність. Він може бути використаний як для регресії, так і для завдань класифікації, а також дозволяє переглянути коефіцієнти, які модель визначила для кожної з ознак.

Однією із проблем машинного навчання є перенавчання, але у більшості випадків це не відбувається завдяки класифікатору випадкових лісів.

Алгоритм «випадкових лісів» є надзвичайно популярним у таких сферах, як банківська справа та фондовий ринок. Наприклад, у фінансах він використовується для виявлення клієнтів, які частіше виплачують борг вчасно або частіше користуються послугами банку. У цьому домені він також використовується для виявлення шахраїв, які обманюють банк. У торгівлі алгоритм можна використовувати для визначення майбутньої поведінки акцій [9].

Класифікаційний звіт моделі «випадкових лісів», на якому представлено точність, повноту та зважений показник F1, котрі дозволяють описати адекватність моделі на тестовій та тренувальній вибірці даних наведено на рисунку 2.

Train data

	precision	recall	f1-score	support
Class-1	0.83	0.83	0.83	13487
Class-2	0.83	0.83	0.83	13438
accuracy			0.83	26925
macro avg	0.83	0.83	0.83	26925
weighted avg	0.83	0.83	0.83	26925

Test data

	precision	recall	f1-score	support
Class-1	0.71	0.70	0.71	4463
Class-2	0.71	0.72	0.72	4512
accuracy			0.71	8975
macro avg	0.71	0.71	0.71	8975
weighted avg	0.71	0.71	0.71	8975

Рис. 2. Класифікаційний звіт моделі «випадкових лісів»
Джерело: Власні розрахунки.

Порівнюючи результати даної моделі із усіма попередніми, слід зазначити, що модель «випадкових лісів» показує найкращі результати по усіх метрикам. Модель досягнула значення оцінки F1 на тестових даних на рівні 0.71, що можна вважати добрим результатом для даного класу задач.

Також на рис. 3 наведено ROC криву для моделі «випадкових лісів», котра показує здатність моделі розрізняти класи у бінарній класифікації.

Відмінна модель має AUC близько до 1, що означає, що вона володіє хорошим вмінням розпізнання класів. Якщо AUC близька до 0 то це означає, що вона обертає результат. Модель відносить представника першого класу до 2 другого класу і навпаки. А коли AUC = 0,5, це означає, що модель не вміє розрізняти класи і для кожного елемента вказує клас випадково.

У даному дослідженні значення площі AUC утвореної кривою ROC досягає значення 0,71, що вказує на хороший результат для задачі кредитного скорингу.

Для кращого сприйняття результатів на рис. 4 зображено матрицю неточностей для моделі «випадкових лісів», яка була розрахована для тестового набору даних. Матриця неточностей представляє відмінності між прогнозованими та фактичними значеннями.

Дослідимо ще модель TPOT, що працює на основі генетичного алгоритму [12].

TPOT модель на відміну від моделі «випадкових лісів» застосовує генетичний алгоритм, який дозволяє оптимізувати роботу моделі. TPOT модель автоматизує пошук алгоритмів машинного навчання, таких як дерева рішень, «випадкові ліси», логістична регресія та інші, а також за допомогою генетичного алгоритму оптимізує процес.

Генетичний алгоритм особливо вигідний, коли множина даних складна і неопукла, так що класичні методи оптимізації, такі як градієнтний спуск, є неефективним засобом для пошуку глобального рішення. Генетичні алгоритми часто називають евристичними алгоритмами пошуку, оскільки вони не гарантують знаходження оптимального рішення, але мають велику ймовірність знайти досить хороше рішення проблеми за короткий проміжок часу.

ISSN 2078-6115. Вісник Львівського університету. Серія економічна. 2022. Випуск 63

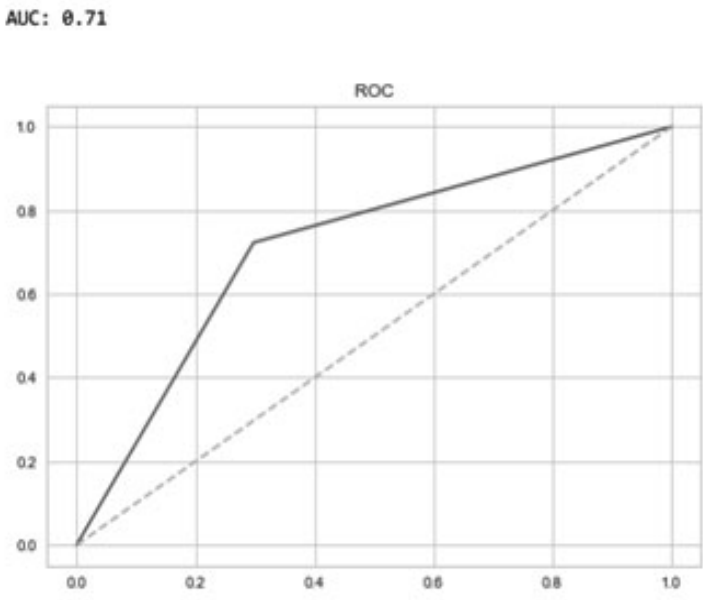

Рис. 3. ROC крива моделі «випадкових лісів»
Джерело: Власні розрахунки.

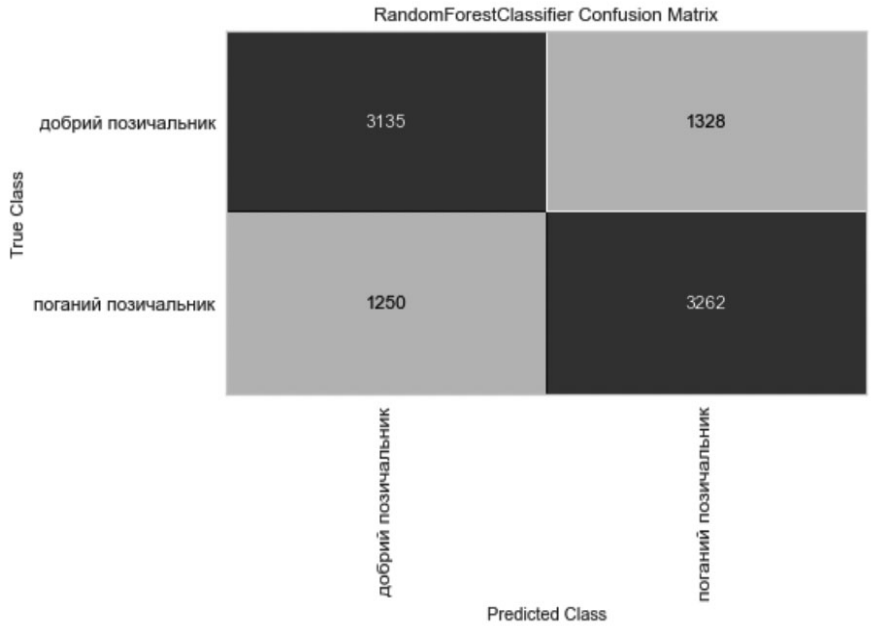

Рис. 4. Матриця неточностей моделі «випадкових лісів»
Джерело: Власні розрахунки.

Для вирішення проблем, генетичний алгоритм використовує такі еволюційні концепції, як виживання найбільш пристосованих, генетичний кросовер та генетична мутація. Замість того, щоб намагатися створити єдине рішення, ці еволюційні концепції застосовуються до сукупності різних рішень проблем, кожне з яких спочатку є випадковим. Популяція проходить через кілька поколінь, буквально розвиваючи рішення за допомогою таких механізмів, як розмноження (кросинговер) та мутація. Після ряду поколінь еволюції найкращим рішенням, знайденим серед усіх поколінь, вибирається остаточне рішення проблеми [14].

Оцінити TPOT модель можна за допомогою тих самих показників, якими було оцінено і модель «випадкових лісів». На рис. 5 наведено класифікаційний звіт моделі TPOT.

Показник F1 набуває значення 0.71, що є таким самим результатом, як для моделі «випадкових лісів» у попередній моделі. Усі показники збалансовані, але TPOT краще прогнозує представників класу, які будуть віддавати кредит (Class-1).

Train data

	precision	recall	f1-score	support
Class-1	0.82	0.85	0.83	13487
Class-2	0.85	0.81	0.83	13438
accuracy			0.83	26925
macro avg	0.83	0.83	0.83	26925
weighted avg	0.83	0.83	0.83	26925

Test data

	precision	recall	f1-score	support
Class-1	0.70	0.74	0.72	4463
Class-2	0.73	0.69	0.71	4512
accuracy			0.71	8975
macro avg	0.71	0.71	0.71	8975
weighted avg	0.71	0.71	0.71	8975

Рис. 5. Класифікацій ний звіт моделі TPOT
Джерело: Власні розрахунки.

Також, щоб побачити, як модель TPOT може розрізняти класи потрібно розглянути ROC криву, що зображена на рисунку 6.

Порівнюючи ROC криві двох розглянутих моделей також видно, що вони майже ідентичні та показують добрий результат.

Розглянемо матрицю неточності моделі TPOT, яка зображена на рис. 7.

Матриця неточності показує, що модель TPOT, краще визначає позичальників, котрі повернуть кошти ніж модель «випадкових лісів», але менше впізнає «поганих» позичальників.

ISSN 2078-6115. Вісник Львівського університету. Серія економічна. 2022. Випуск 63

```
AUC: 0.71
```

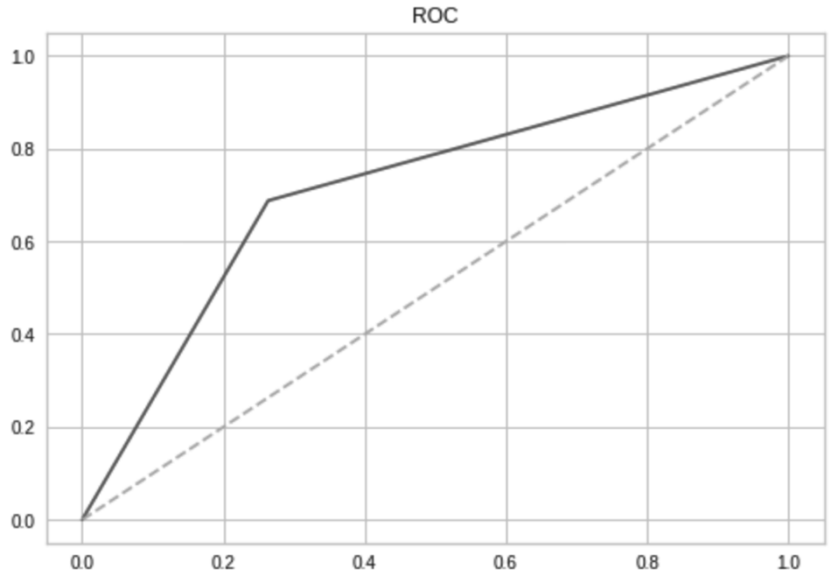

Рис. 6 ROC крива моделі TPOT
Джерело:Власні розрахунки.

	добрий позичальник	поганий позичальник
добрий позичальник	3288	1175
поганий позичальник	1409	3103

Рис. 7. Матриця неточностей моделі TPOT
Джерело: Власні розрахунки.

При навчанні моделі TPOT було проведено тільки 2 епохи навчання тому модель не змогла досягнути оптимальних значень, на нашу думку при значних обчислювальних потужностях TPOT модель дозволить отримати кращі результати та зекономити більше ресурсів банківським установам ніж модель «випадкових лісів», котра на даний момент виявилася кращою у проведеному дослідженні.

Висновки. Отже, обидві моделі показали себе надзвичайно добре. В результаті проведених розрахунків отримано значення показника F1 на рівні 0.71, що є прийнятним для даного класу задач. Також при значних обчислювальних потужностях, на нашу думку, можна отримати кращі результати моделі TPOT, оскільки при даному дослідженні

було реалізовано тільки генетичні цикли, що не дозволило розкрити увесь потенціал цієї моделі.

Побудовані моделі можна використовувати при оцінці кредитоспроможності клієнтів банку. Модель кредитного скорингу виключає суб'єктивні судження працівників банку та аналізує закономірності в даних, котрі людина зауважити не може, тому модель кредитного скорингу ефективніше визначає тих клієнтів, котрим можна надати кредит. Також використання методів машинного навчання дозволяє зекономити час на аналіз даних, котрі надав клієнт, для затвердження кредиту. Модель дозволяє візуалізувати результати, та відповісти на питання що вплинуло на відмову, якщо результат буде негативним для клієнта.

У подальших дослідженнях доцільно створити ансамблі розглянутих алгоритмів або розглянути Баєсові моделі чи певні види нейромереж. Серед різноманіття нейромереж потрібно випробувати сіамські моделі, які дозволяють утворювати кластери даних, згортаючі нейромережі тощо. Також доцільно провести детальніший інжиніринг ознак та можливо утворити нові ознаки.

Список використаних джерел

1. Гаврилюк Г.В. Аналіз вагомості критеріїв в оцінюванні кредитоспроможності фізичних осіб. Нейро-нечіткі технології моделювання в економіці. 2017. №6. С. 3–23.
2. Камінський А.Б. Скорингові технології в кредитному ризик-менеджменті / Камінський А.Б., Писанець К.К. // Бізнес-інформ. 2012. № 4. С. 197–201.
3. Клебан Ю., Горошко Н. Ідентифікація дефолтних клієнтів банку методами машинного навчання на основі біннінгу показників. Економічний аналіз. 2021. Том 31. № 1. С. 133–142.
4. Кишакевич, Б.Ю. Моделювання та оптимізація кредитних ризиків банку. Дрогобич : Коло, 2011. 412 с.
5. Рева Р.В. Система кредитного скорингу позичальників кредитів на основі інтелектуального аналізу даних: магістерська дисертація: Київ 2018. 115с. URL https://ela.kpi.ua/bitstream/123456789/26074/1/Revva_magistr.pdf
6. Філатов В.Ю. Алгоритм оцінки кредитного ризику позичальників – фізичних осіб з використанням методів машинного навчання. Матеріали XII міжнародної науково-практичної конференції: Сучасні проблеми моделювання соціально-економічних систем. 09-10 квітня 2020 року URL: https://mpsesm.org/book/2020/thesis03-904.html#thesis03-904 (дата звернення 20.04.2021)
7. Anderson, R. The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation, Oxford University Press. UK. 2007. 731p.
8. A Complete Guide To The Random Forest Algorithm URL: https://builtin.com/data-science/random-forest-algorithm (дата звернення 18.04.2021)
9. IBM Random Forest URL: https://www.ibm.com/cloud/learn/random-forest (дата звернення 18.04.2021)
10. Jupyter Notebook URL: https://jupyter.org (дата звернення 18.04.2021)
11. Kaggle URL: https://www.kaggle.com/ (дата звернення 23.04.2021)
12. Tpot documentation URL:http://epistasislab.github.io/tpot/ (дата звернення 23.04.2021)
13. Thomas. L.C. A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. International Journal of Forecasting, vol. 16, 2000. p.149–172.

14. Using genetic algorithms on AWS for optimization problems URL: https://aws.amazon.com/ru/blogs/machine-learning/using-genetic-algorithms-on-aws-for-optimization-problems/ (дата звернення 23.04.2021)

References

1. Havrylyuk, H. V. (2017), Analiz vahomosti kryteriiv v otsiniuvanni kredytospromozhnosti fizychnykh osib [Analysis of the significance of the criteria in assessing the creditworthiness of individuals], Nejro-nechitki tekhnolohii modeliuvannia v ekonomitsi, vol. 6, 3–23. [in Ukrainian]
2. Kamins'kij A.B. and Pisanec' K.K. (2012), Ckorynhovi tekhnolohii v kredytnomu ryzyk-menedzhmenti [Scoring technologies in credit risk-management], Biznes-inform, vol. 4, 197–201.[in Ukrainian]
3. Kleban Yu., Horoshko N. (2021) Identyfikatsiia defoltnykh kliientiv banku metodamy mashynnoho navchannia na osnovi binninhu pokaznykiv. [Identification of defaulted bank clients using machine learning methods based on indicator binning] Ekonomichnyi analiz. Tom 31. № 1, 133–142. [in Ukrainian]
4. Kyshakevych B.Yu. (2011). Modeliuvannia ta optymizatsiia kredytnykh ryzykiv banku [Modeling and optimization of bank credit risks]. Drohobych: Kolo. 412 [in Ukrainian]
5. Reva R.V. (2018) Systema kredytnoho skorynhu pozychalnykiv kredytiv na osnovi intelektualnoho analizu danykh [Credit scoring system of loan borrowers based on intelligent data analysis]: mahisterska dysertatsiia: Kyiv. 115. URL https://ela.kpi.ua/bitstream/123456789/26074/1/Revva_magistr.pdf [in Ukrainian]
6. Filatov V.Iu. (2020) Alhorytm otsinky kredytnoho ryzyku pozychalnykiv – fizychnykh osib z vykorystanniam metodiv mashynnoho navchannia. [Machine learning algorithms for measuring the credit risk of individual borrowers] Materialy XII mizhnarodnoi naukovo-praktychnoi konferentsii: Suchasni problemy modeliuvannia sotsialno-ekonomichnykh system. 09–10 kvitnia 2020 roku https://mpsesm.org/book/2020/thesis03-904.html#thesis03-904. [in Ukrainian]
7. Anderson, R. (2007) The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation, Oxford University Press. UK. 731.
8. A Complete Guide To The Random Forest Algorithm URL: https://builtin.com/data-science/random-forest-algorithm
9. IBM Random Forest URL: https://www.ibm.com/cloud/learn/random-forest
10. Jupyter Notebook URL: https://jupyter.org
11. Kaggle URL: https://www.kaggle.com/
12. Tpot documentation URL:http://epistasislab.github.io/tpot/
13. Thomas. L.C. (2000) A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. International Journal of Forecasting, vol. 16, 149–172.
14. Using genetic algorithms on AWS for optimization problems URL: https://aws.amazon.com/ru/blogs/machine-learning/using-genetic-algorithms-on-aws-for-optimization-problems/

CREDIT SCORING MODELLING USING MACHINE LEARNING METHODS

Myroslav Datsko[1], Nazarii Drushchak[2]

[1]Ivan Franko National University of Lviv,
18 Prospekt Svobody, Lviv, 79008,
e-mail: myroslav.datsko@lnu.edu.ua; ORCID: 0000-0002-5181-4632

[2]Ukrainian Catholic University,
2 Kozelnitskaya Str., Lviv,
e-mail: naz2001r@gmail.com; ORCID: 0000-0002-5056-3026

Abstract. Credit scoring is considered a successful field of application and a methodological subset of intelligent data analysis. The problem of credit scoring within the framework of thoughtful data analysis can be attributed to the class of classification methods of machine learning. A classification task is used to find different representatives of certain predefined classes.

The main area of application for scoring modeling is risk management, but in general, scoring models are used for various tasks of binary classification, diagnostics, forecasting of the probability of occurrence of a certain unexpected event, detection of hidden signs through the prism of observed characters with a certain probability, etc.

The article discusses credit scoring models built using machine learning methods. The article builds 6 models, among which two with the best characteristics are highlighted: the «random forest» model and the TPOT model. Using the classification report and the ROC curve, it was determined that both models achieve an F1-score -- 0.71 for the studied sample, which is quite acceptable for the credit scoring task.

The built models can be used when assessing the creditworthiness of a bank's clients. The credit scoring model eliminates subjective judgments of bank employees and analyzes data that a person may not notice, making it more effective for those clients who can be given credit. The use of machine learning methods also saves time in analyzing data provided by the client for credit approval.

In future research, it is advisable to create ensembles of the analyzed algorithms or to consider Bayesian models, and certain types of neural networks or boosters. It is also necessary to carry out detailed feature engineering and model interpretation which is important for final decision-making.

Keywords: machine learning, credit scoring, «random forests», TPOT model, classification report, ROC curve

Стаття надійшла до редакції 26.11.2022
Прийнята до друку 02.02.2023