

07. <https://ekmair.ukma.edu.ua/server/api/core/bitstreams/0e5f5294-ddf5-4458-9689-5cfbea7a4097/content>
7. Ghosh A. YOLO11: Faster Than You Can Imagine! 2024 Жов 08. URL: <https://learnopencv.com/yolo11/>
8. Ultralytics YOLO docs / YOLO Performance Metrics. 2023 Лис 12. URL: <https://docs.ultralytics.com/guides/yolo-performance-metrics/>
9. Building-faced-segmentation-original Computer Vision Project (dataset). URL: <https://universe.roboflow.com/building-facade/building-facade-segmentation-original>

DOI 10.70286/ISU-22.04.2026.010

ВИКОРИСТАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ В ІНФОРМАЦІЙНІЙ СИСТЕМІ КРЕДИТНОГО СКОРИНГУ БАНКУ

Кузьменко Олександра

к.е.н., доцент

Бабаджанова Карина

здобувач вищої освіти ОП «ІТ-Фінанси»

Національний університет «Полтавська політехніка
імені Юрія Кондратюка», м. Полтава, Україна

Вступ. У сучасних умовах цифрової трансформації економіки особливого значення набуває впровадження інтелектуальних інформаційних систем, здатних автоматизувати процеси аналізу даних, прогнозування та підтримки прийняття рішень. Однією з найбільш актуальних сфер використання таких технологій є банківський сектор, у якому швидкість обробки заявок, точність оцінювання ризиків та якість обслуговування клієнтів безпосередньо впливають на конкурентоспроможність фінансових установ.

Кредитні операції формують вагомую частку доходів банківських установ, проте супроводжуються ризиком неповернення коштів, виникненням простроченої заборгованості та погіршенням структури кредитного портфеля. Традиційні підходи до оцінювання кредитоспроможності ґрунтуються на аналізі доходів клієнта, кредитної історії, заставного забезпечення та експертних висновках працівників банку. Разом із тим зростання обсягів даних, збільшення кількості клієнтів і необхідність оперативного прийняття рішень зумовлюють потребу у використанні сучасних цифрових інструментів.

Методи машинного навчання дозволяють автоматично виявляти приховані закономірності у великих масивах інформації, прогнозувати кредитний ризик, адаптуватися до змін зовнішнього середовища та зменшувати вплив суб'єктивного людського фактора. Саме тому побудова інформаційних систем

кредитного скорингу на основі алгоритмів штучного інтелекту є перспективним напрямом розвитку сучасних фінансових технологій [1; 2].

Метою дослідження є аналіз можливостей застосування методів машинного навчання в інформаційній системі кредитного скорингу банку з використанням відкритого набору даних German Credit Data та визначення найбільш ефективного алгоритму класифікації кредитного ризику.

Матеріали та методи. Інформаційною базою дослідження використано відкритий набір даних German Credit Data [6], який широко застосовується у міжнародній практиці як тестове середовище для задач кредитного скорингу. Набір містить інформацію про одну тисячу позичальників банку та результати їх кредитної поведінки. Використання такого масиву даних є доцільним, оскільки він включає як числові, так і категоріальні параметри, що відображають типові характеристики клієнтів фінансових установ.

До переліку вхідних параметрів віднесено вік клієнта, стать, тип зайнятості, житловий статус, суму кредиту, строк кредитування, стан поточного рахунку, наявність заощаджень та мету кредитування. Цільова змінна характеризувала рівень кредитного ризику та визначала належність клієнта до категорії надійних або ризикових позичальників.

З позиції інформаційних технологій задача розглядається як побудова інтелектуальної системи класифікації, у якій вхідні дані перетворюються на автоматично сформоване рішення щодо можливості кредитування. Архітектурно система включає модуль введення даних, блок попередньої обробки, модуль машинного навчання, підсистему оцінювання результатів, інтерфейс користувача, засоби збереження інформації та механізми журналювання подій.

Перед навчанням моделей виконано перевірку якості даних, аналіз структури ознак, виявлення пропущених значень та попередню підготовку вибірки. Категоріальні змінні було перетворено у числовий формат методом One-Hot Encoding, а числові параметри стандартизовано для забезпечення порівняльності масштабів. Після цього набір даних поділено на навчальну та тестову частини у співвідношенні 80:20.

Реалізацію експериментальної частини виконано мовою Python із використанням бібліотек Pandas, NumPy, Scikit-learn, Matplotlib та Seaborn.

Для порівняльного аналізу використано такі алгоритми, як логістична регресія; дерево рішень; випадковий ліс; градієнтний бустинг.

Оцінювання якості моделей здійснювалося за показниками Accuracy, Precision, Recall, F1-score та ROC-AUC [3; 4].

Результати та обговорення. Проведене дослідження підтвердило високу ефективність застосування алгоритмів машинного навчання в інформаційних системах кредитного скорингу. Усі досліджувані моделі продемонстрували належний рівень класифікації, однак отримані результати відрізнялися залежно від математичної природи алгоритмів, здатності працювати з нелінійними залежностями та стійкості до варіативності даних. Для узагальнення результатів сформовано порівняльну таблицю ефективності моделей (табл. 1).

Таблиця 1. Порівняльні результати моделей класифікації кредитоспроможності

Модель	Точність класифікації	Прецизійність	Повнота	F1-міра	Площа під ROC-кривою
1	2	3	4	5	6
Логістична регресія	0,755	0,682	0,600	0,638	0,742
Дерево рішень	0,720	0,633	0,633	0,633	0,714
Випадковий ліс	0,790	0,729	0,683	0,705	0,834
Градiєнтний бустинг	0,805	0,756	0,700	0,727	0,852

Як свідчать результати таблиці 1, найвищі показники якості продемонстрував алгоритм градієнтного бустингу. Він забезпечив максимальні значення точності класифікації та площі під ROC-кривою, що вказує на високу здатність моделі коректно розмежовувати групи клієнтів за рівнем ризику.

Високу ефективність також показав випадковий ліс, який характеризується стійкістю до перенавчання та стабільністю прогнозування. Логістична регресія поступилася ансамблевим моделям за точністю, однак зберігає перевагу інтерпретованості результатів, що є важливим для пояснення прийнятих рішень.

Для наочного порівняння точності моделей доцільно використати графічну інтерпретацію результатів (рис. 1).

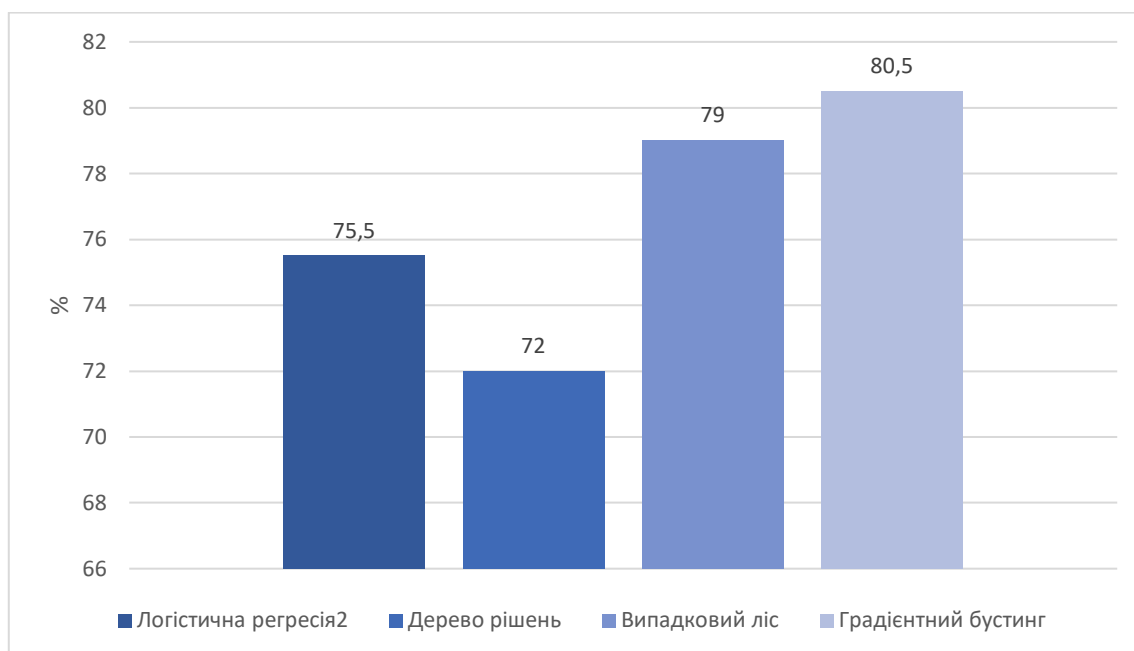


Рисунок 1. Порівняння точності моделей.

Наведена візуалізація підтверджує перевагу ансамблевих методів над класичними алгоритмами у задачах кредитного скорингу.

Додатковий аналіз важливості ознак дозволив визначити ключові фактори, що найбільше впливають на результати класифікації (табл. 2, рис. 2).

Таблиця 2. Найбільш значущі фактори оцінювання кредитоспроможності

Ранг	Фактор	Практичне значення
1	2	3
1	Стан поточного рахунку	Характеризує фінансову дисципліну клієнта
2	Строк кредитування	Визначає тривалість ризикового періоду
3	Сума кредиту	Відображає масштаб потенційного ризику
4	Вік клієнта	Опосередковано характеризує стабільність доходів
5	Наявність заощаджень	Свідчить про фінансову стійкість

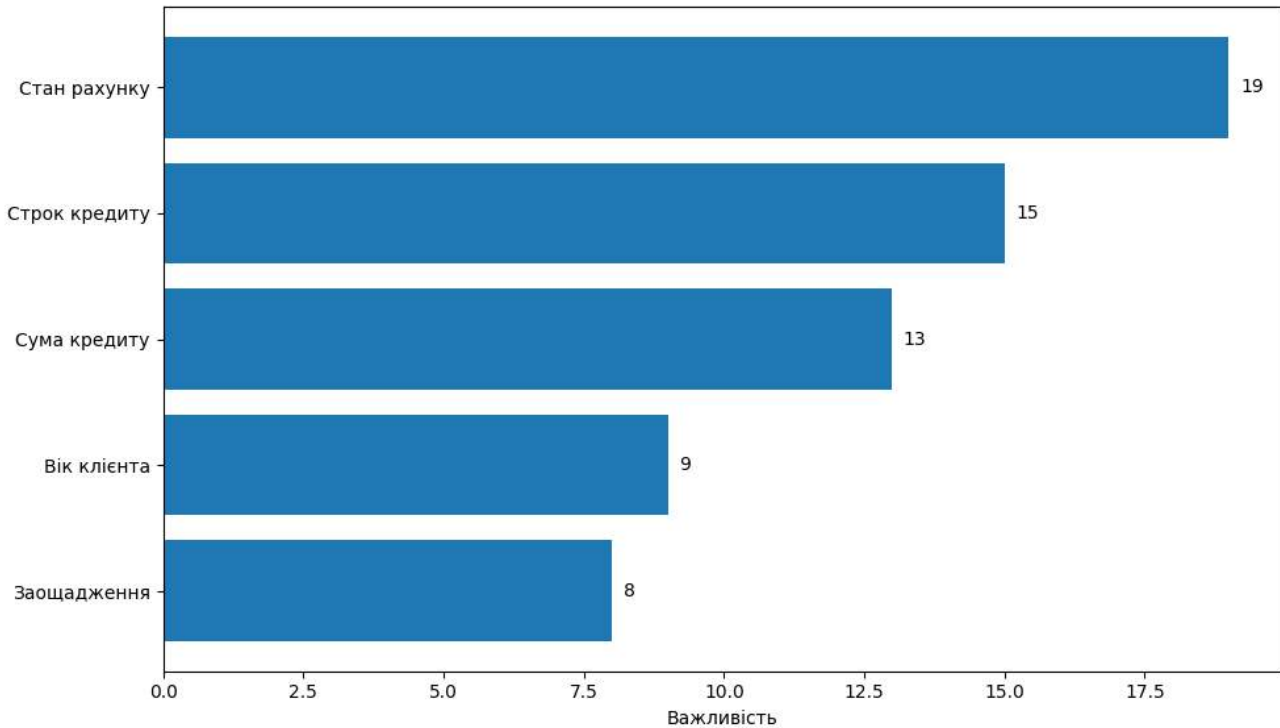


Рисунок 2. Значущість факторів

Отримані результати свідчать, що найбільшу прогностичну цінність мають фінансові та поведінкові характеристики клієнта. Це означає, що при побудові сучасних скорингових систем доцільно робити акцент на реальній платіжній спроможності клієнта, історії управління коштами та параметрах кредитної заявки, а не лише на формальних соціально-демографічних ознаках.

Практичне значення проведеного дослідження полягає у можливості впровадження автоматизованих систем оцінювання кредитоспроможності у діяльність банківських установ. Використання таких моделей дозволяє скоротити час розгляду заявки, зменшити навантаження на працівників кредитних підрозділів, підвищити об'єктивність прийняття рішень та знизити рівень кредитного ризику.

Окремою перевагою є можливість інтеграції алгоритмів машинного навчання у цифрові банківські платформи, веб-сервіси, мобільні застосунки та системи дистанційного обслуговування клієнтів.

Під час впровадження інтелектуальних скорингових систем особливу увагу необхідно приділяти питанням захисту персональних і фінансових даних користувачів. До ключових вимог належать автентифікація користувачів,

розмежування прав доступу, журналювання дій, резервне копіювання та шифрування інформації.

Окремої уваги потребує захист моделей машинного навчання від несанкціонованої модифікації, підміни вхідних даних та маніпулятивних впливів на результати прогнозування.

Висновки. У результаті проведеного дослідження доведено доцільність застосування методів машинного навчання для побудови інформаційних систем кредитного скорингу банку.

Встановлено, що найвищу ефективність серед досліджуваних алгоритмів продемонстрував градієнтний бустинг, який забезпечив найкращі показники якості класифікації.

Проведений аналіз важливості ознак показав, що найбільший вплив на результати моделі мають стан поточного рахунку клієнта, строк кредитування, сума кредиту, вік позичальника та наявність заощаджень.

Практична цінність роботи полягає у можливості створення програмного модуля для цифрових банківських платформ, веб-сервісів, мобільних застосунків та автоматизованих систем підтримки прийняття рішень.

Подальші дослідження доцільно спрямувати на використання глибокого навчання, пояснюваного штучного інтелекту, хмарних технологій обробки даних та вдосконалених механізмів кіберзахисту.

Список використаних джерел

1. Altman E. I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*. 1968. Vol. 23(4). P. 589–609.
2. Baesens B. et al. Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*. 2003. Vol. 54(6). P. 627–635.
3. Breiman L. Random Forests. *Machine Learning*. 2001. Vol. 45(1). P. 5–32.
4. Hand D. J., Henley W. E. Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society*. 1997. Vol. 160(3). P. 523–541.
5. Thomas L. C. A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International Journal of Forecasting*. 2000. Vol. 16(2). P. 149–172.
6. German Credit Data // UCI Machine Learning Repository. URL: [https://archive.ics.uci.edu/ml/datasets/statlog+\(german+credit+data\)](https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data)) (дата звернення: 21.04.2026).