

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»**

**Навчально-науковий інститут прикладного системного аналізу
Кафедра математичних методів системного аналізу**

До захисту допущено:
Завідувач кафедри
_____ Оксана ТИМОЩУК
«__» _____ 2025 р.

Дипломна робота

на здобуття ступеня бакалавра

**за освітньо-професійною програмою «Системний аналіз і управління»
спеціальності 124 «Системний аналіз»**

**на тему: «Застосування підходів інтелектуального аналізу даних для
оцінки кредитної спроможності клієнтів банку»**

Виконала:

студентка IV курсу, групи КА-15
Комаренко Анастасія Романівна _____

Керівник:

доцент, доктор філософії з комп'ютерних наук,
Гуськова Віра Геннадіївна _____

Консультант з економічного розділу:

доцент, к. е. н.,
Рощина Надія Василівна _____

Консультант з нормоконтролю:

асистент кафедри ММСА
Канцедал Георгій Олегович _____

Рецензент:

професор, д.т.н.
Данилов Валерій Якович _____

Засвідчую, що у цій дипломній роботі
немає запозичень з праць інших авторів
без відповідних посилань.

Студентка _____

Київ – 2025

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Навчально-науковий інститут прикладного системного аналізу
Кафедра математичних методів системного аналізу

Рівень вищої освіти – перший (бакалаврський)

Спеціальність – 124 «Системний аналіз»

Освітньо—професійна програма «Системний аналіз і управління»

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ О.Л. Тимощук

«__» _____ 20__ р.

ЗАВДАННЯ

на дипломну роботу студентці

Комаренко Анастасії Романівні

1. Тема роботи «Застосування підходів інтелектуального аналізу даних для оцінки кредитної спроможності клієнтів банку», керівник роботи Гуськова Віра Геннадіївна, PhD, доцент, затверджені наказом по університету від 26.05.2025р. №1759—с
2. Термін подання студентом роботи _____
3. Вихідні дані до роботи
4. Зміст роботи
5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо)
6. Консультанти розділів роботи

| | | | |
|-------------|---|----------------|------------------|
| Розділ | Прізвище, ініціали та посада консультанта | Підпис, дата | |
| | | завдання видав | завдання прийняв |
| Економічний | Рощина Надія Василівна | | |

7. Дата видачі завдання 11 квітня 2025 р.

Календарний план

| № з/п | Назва етапів виконання дипломної роботи | Термін виконання етапів роботи | Примітка |
|-------|---|--------------------------------|----------|
| 1 | Формулювання теми бакалаврської роботи. | 14.04.2025 – 20.04.2025 | виконано |
| 2 | Пошук та огляд спеціальної літератури за темою дослідження. | 21.04.2025 — 27.04.2025 | виконано |
| 3 | Ознайомлення із літературою та програмним забезпеченням для машинного навчання, необхідними для написання роботи. | 28.04.2025 – 04.05.2025 | виконано |
| 4 | Пошук вхідних даних для аналізу. | 05.05.2025 – 11.05.2025 | виконано |
| 5 | Порівняння методів вирішення задачі оптимізації параметрів, виділення їх недоліків. | 12.05.2025 – 18.05.2025 | виконано |
| 6 | Реалізація програмного комплексу та тестування його роботи | 19.05.2025 – 01.06.2025 | виконано |
| 7 | Підготовка презентації для захисту | 02.06 – 08.06.2025 | виконано |
| 8 | Попередній захист дипломної роботи | 09.06.2025 – 15.06.2025 | виконано |
| 9 | Захист дипломної роботи | 16.06.2025 – 30.06.2025 | виконано |

Студентка

Анастасія КОМАРЕНКО

Керівник

Віра ГУСЬКОВА

РЕФЕРАТ

Дипломна робота: 93 с., 10 рис., 16 табл., 2 дод., 23 дж.

КРЕДИТОСПРОМОЖНІСТЬ, БАНКІВСЬКА АНАЛІТИКА,
ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ, МАШИННЕ НАВЧАННЯ,
ПРОГНОЗУВАННЯ РИЗИКІВ

Об'єкт дослідження – процеси оцінки кредитоспроможності клієнтів банків на основі фінансових, демографічних та поведінкових даних.

Програмний продукт – Прототип системи підтримки прийняття рішень щодо кредитування, що використовує методи машинного навчання для аналізу клієнтських даних та прогнозування рівня кредитного ризику. Мова програмування – Python (з використанням бібліотек Pandas, Scikit—learn, TensorFlow, Matplotlib).

Мета роботи – Розробка та обґрунтування підходу до оцінки кредитоспроможності клієнтів банку на основі інтелектуального аналізу даних, із подальшим створенням прототипу аналітичної системи, що дозволяє автоматизувати та підвищити точність кредитного скорингу.

Метод дослідження – Розробка та обґрунтування підходу до оцінки кредитоспроможності клієнтів банку на основі інтелектуального аналізу даних, із подальшим створенням прототипу аналітичної системи, що дозволяє автоматизувати та підвищити точність кредитного скорингу.

ABSTRACT

Thesis includes: 93 pages, 10 figures, 16 tables, 2 appendices, 23 references.

CREDITWORTHINESS, BANKING ANALYTICS, DATA MINING, MACHINE LEARNING, RISK FORECASTING

Object of research – the processes of assessing the creditworthiness of bank clients based on financial, demographic, and behavioral data.

Software product – a prototype of a decision support system for credit assessment, utilizing machine learning methods to analyze client data and predict the level of credit risk. Programming language – Python (using libraries such as Pandas, Scikit—learn, TensorFlow, Matplotlib).

Aim of the work – to develop and justify an approach to assessing the creditworthiness of bank clients based on data mining methods, followed by the creation of an analytical system prototype that automates and improves the accuracy of credit scoring.

Research method – data collection and preprocessing, application of classification and ensemble learning algorithms, statistical evaluation of model performance, comparison of results using quality metrics (Accuracy, Precision, Recall, ROC—AUC).

ЗМІСТ

| | |
|--|-----------|
| ПЕРЕЛІК СКОРОЧЕНЬ | 8 |
| ВСТУП | 9 |
| РОЗДІЛ 1 ТЕОРЕТИЧНІ ОСНОВИ ТА ПОСТАНОВКА ПРОБЛЕМИ ОЦІНКИ КРЕДИТОСПРОМОЖНОСТІ КЛІЄНТІВ БАНКУ | 12 |
| 1.1. Поняття та значення кредитоспроможності..... | 12 |
| 1.2. Ключові проблеми оцінки платоспроможності клієнтів у сучасних банках | 14 |
| 1.3. Огляд традиційних підходів до аналізу кредитного ризику | 17 |
| 1.4. Аналіз та порівняння з існуючими рішеннями | 22 |
| 1.5. Обмеження класичних методів у сучасних умовах | 28 |
| 1.6. Постановка задачі | 29 |
| 1.7. Висновки до розділу 1 | 31 |
| РОЗДІЛ 2 ЗАСТОСУВАННЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ У ЗАДАЧАХ КРЕДИТНОГО СКОРИНГУ | 33 |
| 2.1. Сутність інтелектуального аналізу даних (Data Mining) у фінансовому секторі..... | 33 |
| 2.2. Роль штучного інтелекту та машинного навчання в банківській аналітиці | 35 |
| 2.3. Огляд сучасних моделей для оцінки кредитоспроможності | 37 |
| 2.4. Вибір та обґрунтування методів машинного навчання на прикладі нечіткої нейронної мережі | 44 |
| 2.5. Критерії вибору оптимальної моделі для банківських задач | 47 |
| 2.6. Висновки до розділу 2..... | 49 |
| РОЗДІЛ 3 РОЗРОБКА ПРОТОТИПУ СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ОЦІНКИ КРЕДИТОСПРОМОЖНОСТІ. 51 | 51 |
| 3.1. Підготовка даних для моделювання | 51 |
| 3.2. Оцінка ефективності моделі та інтеграційний потенціал..... | 55 |
| 3.3. Висновки до розділу 3 | 60 |
| РОЗДІЛ 4 РОЗРОБКА ТЕХНІЧОЇ СИСТЕМИ ТА ОЦІНКА ЕФЕКТИВНОСТІ РОЗРОБЛЕНОЇ СИСТЕМИ ТА ПЕРСПЕКТИВИ РОЗВИТКУ | 62 |
| 4.1. Постановка задачі проєктування..... | 63 |
| 4.2. Обґрунтування функцій програмного продукту | 64 |

| | |
|---|-----------|
| | 7 |
| 4.3. Обґрунтування системи параметрів програмного продукту..... | 70 |
| 4.4. Аналіз експертного оцінювання параметрів..... | 73 |
| 4.5. Аналіз рівня якості варіантів реалізації функцій..... | 78 |
| 4.6. Економічний аналіз варіантів розробки ПП..... | 80 |
| 4.7. Вибір кращого варіанту ПП техніко—економічного рівня..... | 85 |
| 4.8. Висновки до розділу 4..... | 87 |
| ВИСНОВКИ..... | 89 |
| ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ..... | 91 |
| ДОДАТОК А ЛІСТИНГ ПРОГРАМИ..... | 95 |
| ДОДАТОК Б ДОПОВІДЬ У ВИГЛЯДІ ПРЕЗЕНТАЦІЇ..... | 99 |

ПЕРЕЛІК СКОРОЧЕНЬ

ІАД — Інтелектуальний аналіз даних

КС — Кредитоспроможність

ДСС — Дискретна система скорингу

AI — (Artificial Intelligence) штучний інтелект

ML — (Machine Learning) машинне навчання

NN — (Neural Network) нейронна мережа

SVM — (Support Vector Machine) метод опорних векторів

LR — (Logistic Regression) логістична регресія

RF — (Random Forest) випадковий ліс

AUC — (Area Under the Curve) площа під ROC—кривою

ROC — (Receiver Operating Characteristic) крива робочих характеристик

приймача

API — (Application Programming Interface) програмний інтерфейс

SQL — (Structured Query Language) мова структурованих запитів

CSV — (Comma—Separated Values) формат табличних даних

ETL — (Extract, Transform, Load) процес вилучення, трансформації та

завантаження даних

KPI — (Key Performance Indicator) ключовий показник ефективності

UI — (User Interface) користувацький інтерфейс

ВСТУП

У сучасному фінансовому секторі питання управління кредитними ризиками набуває особливої ваги. Банки щодня приймають сотні рішень щодо видачі кредитів, і від точності оцінки кредитоспроможності залежить не лише стабільність окремих фінансових установ, а й надійність усього банківського ринку. Класичні методи аналізу, що базуються на ручному скорингу, експертних системах або простих статистичних моделях, уже не відповідають вимогам часу: вони малоефективні при обробці великих обсягів даних і не враховують приховані залежності у поведінці клієнтів.

Водночас активний розвиток штучного інтелекту, машинного навчання та технологій інтелектуального аналізу даних (Data Mining) відкриває нові можливості для створення гнучких та адаптивних моделей оцінки ризику. Їх використання дозволяє автоматизувати процес прийняття рішень, підвищити точність прогнозування та мінімізувати ризики неповернення кредитів. Саме тому застосування підходів інтелектуального аналізу даних у задачах оцінки кредитоспроможності клієнтів є надзвичайно актуальним і має вагоме практичне значення.

Метою дипломної роботи є розробка та реалізація прототипу системи підтримки прийняття рішень щодо оцінки кредитоспроможності клієнтів банку з використанням сучасних методів інтелектуального аналізу даних.

Основні завдання дослідження:

- 1) дослідити теоретичні основи оцінки кредитоспроможності клієнтів банку;
- 2) проаналізувати обмеження традиційних методів оцінювання;
- 3) здійснити огляд сучасних моделей машинного навчання, придатних для вирішення задач кредитного скорингу;
- 4) обґрунтувати вибір алгоритмів для побудови прогнозної моделі;
- 5) зібрати та підготувати дані для навчання;

6) реалізувати прототип аналітичної системи та оцінити її ефективність;

7) провести порівняльний аналіз з існуючими рішеннями та сформулювати рекомендації щодо вдосконалення системи.

Об'єкт дослідження — процес прийняття рішень щодо надання кредитів банківськими установами.

Предмет дослідження — методи інтелектуального аналізу даних, які застосовуються для побудови моделей оцінки кредитоспроможності клієнтів банку.

У роботі використовуються методи:

1) машинного навчання (логістична регресія, дерева рішень, ансамблеві методи, нейронні мережі);

2) аналізу якості моделей (Precision, Recall, F1-score, ROC-AUC);

3) інтелектуального аналізу фінансових та демографічних даних;

4) програмна реалізація моделей здійснена з використанням середовища Python, бібліотек Pandas, Scikit-learn, TensorFlow та інструментів візуалізації результатів.

Наукова новизна роботи полягає у поєднанні сучасних методів машинного навчання для вирішення задачі кредитного скорингу на основі реальних фінансових даних.

Практичне значення дослідження полягає у створенні функціонального прототипу системи, яка може бути адаптована до практичного використання у банківських установах для автоматизованої оцінки кредитних ризиків та прийняття рішень про видачу кредиту.

Дипломна робота складається з вступу, чотирьох розділів, висновків, списку використаних джерел і додатків.

У першому розділі проаналізовано сутність поняття кредитоспроможності, розглянуто традиційні підходи до оцінки кредитного ризику, їх обмеження та постановку задачі.

Другий розділ присвячений дослідженню методів інтелектуального аналізу даних у контексті кредитного скорингу та вибору оптимального алгоритму.

У третьому розділі розглядаються практичні аспекти підготовки даних, побудови та оцінки моделей.

Четвертий розділ описує реалізацію прототипу системи, порівняння з альтернативами, визначає її переваги та перспективи розвитку.

РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ ТА ПОСТАНОВКА ПРОБЛЕМИ ОЦІНКИ КРЕДИТОСПРОМОЖНОСТІ КЛІЄНТІВ БАНКУ

1.1. Поняття та значення кредитоспроможності

Кредитоспроможність є фундаментальним поняттям у сфері банківської справи, яке відіграє ключову роль у процесі управління кредитними ризиками та забезпеченні стабільності фінансової системи загалом. Її значення набуває особливої актуальності в умовах зростаючої конкуренції на фінансовому ринку, нестабільності економічного середовища та потреби банків у прийнятті обґрунтованих рішень щодо кредитування.

У загальному розумінні, кредитоспроможність — це здатність позичальника (фізичної або юридичної особи) вчасно й у повному обсязі виконати зобов'язання перед кредитором відповідно до умов укладеного кредитного договору [1]. Це поняття охоплює як поточний фінансовий стан позичальника, так і прогноз його платоспроможності в майбутньому з урахуванням динаміки доходів, боргового навантаження, ринкових умов та інших релевантних факторів [1].

Історично концепція кредитоспроможності бере початок з етапів зародження банківництва, коли кредитні рішення приймалися на основі репутації позичальника, соціального статусу та особистих зв'язків. У ХХ столітті з розвитком бухгалтерського обліку, фінансового аналізу та нормативного регулювання з'явилися перші спроби формалізованої оцінки здатності до погашення боргу. У 1950–1970-х роках почали активно застосовуватись фінансові коефіцієнти та скорингові системи, зокрема модель Альтмана Z-score, які дозволили стандартизувати підхід до визначення рівня надійності позичальників [2].

Кредитоспроможність розглядається як інтегральний показник, що відображає сукупність характеристик позичальника. Для її визначення банки використовують набір критеріїв, що охоплюють:

- 1) фінансові показники (дохід, витрати, активи, зобов'язання, ліквідність, прибутковість);
- 2) кредитну історію (наявність та якість попередніх кредитів, прострочення, виконання умов договорів);
- 3) демографічні дані (вік, освіта, сімейний стан, місце проживання);
- 4) соціально-економічні чинники (зайнятість, стабільність роботи, галузь економіки).

У юридичних осіб оцінюють також стратегію розвитку, структуру власності, фінансову звітність, управлінський склад, ринкову позицію тощо. Для фізичних осіб акцент робиться на персональній дисципліні платежів та стабільності доходів.

У міжнародній практиці оцінка кредитоспроможності тісно пов'язана з діяльністю кредитних бюро та рейтингових агентств, які акумулюють дані про позичальників та надають кредитні рейтинги, що використовуються банками для формування кредитного портфеля. Наприклад, у США діють такі агентства, як Equifax, Experian, TransUnion. У Європейському Союзі та Україні - аналогічні функції виконують локальні бюро кредитних історій [3].

Водночас в умовах цифрової трансформації фінансового сектору зростає роль алгоритмічного скорингу, що базується на інтелектуальному аналізі даних та штучному інтелекті. Такі підходи дозволяють враховувати сотні змінних, включаючи нестандартні джерела інформації (поведінкові патерни, активність у соціальних мережах, транзакційні дані), що суттєво підвищує точність прогнозування.

Оцінка кредитоспроможності є базисом для ухвалення кредитного рішення, визначення умов кредитування (сума, строк, ставка), формування резервів під кредитні ризики, а також виявлення недобросовісних позичальників ще на етапі попередньої перевірки. Невірні оцінки призводять

до зростання частки непрацюючих кредитів (NPL), фінансових втрат і репутаційних ризиків [2].

В сучасних умовах банки, окрім класичних моделей аналізу, впроваджують скорингові системи нового покоління, які базуються на методах машинного навчання та аналізу великих даних. Це дозволяє оперативнo обробляти тисячі заявок, забезпечуючи індивідуальний підхід до клієнта, оптимізацію процесу кредитування та підвищення рентабельності.

Кредитоспроможність має багатогранний характер і охоплює як фінансові, так і нефінансові аспекти надійності позичальника. Значення кредитоспроможності для банківського сектору є винятковим, оскільки саме через її ефективну оцінку реалізується стратегія мінімізації ризиків, забезпечується стабільність фінансових потоків і формуються конкурентні переваги банківської установи на ринку [3].

1.2. Ключові проблеми оцінки платоспроможності клієнтів у сучасних банках

У сучасних умовах банківська система стикається з численними викликами, які ускладнюють процес надійної та об'єктивної оцінки платоспроможності клієнтів. Попри зростання обсягів даних, доступ до нових аналітичних інструментів та цифровізацію процесів, банки все ще мають справу з низкою проблем, що суттєво впливають на якість кредитного аналізу. Ці проблеми можна умовно поділити на технічні, організаційні, регуляторні та екзогенні.

Однією з основних проблем є відсутність або неповнота актуальної фінансової інформації про клієнтів. Зокрема, малий та середній бізнес часто не має стандартизованої звітності, фізичні особи приховують або не можуть підтвердити свої доходи, а дані, які надають клієнти, можуть бути завідомо

викривленими. В умовах економічної нестабільності інформація швидко втрачає актуальність, а її перевірка займає час і ресурси [4].

Багато банків усе ще використовують класичні статистичні моделі, побудовані на фіксованому наборі факторів (наприклад, рівень доходу, наявність майна, кредитна історія), які не завжди відображають реальний кредитний ризик у сучасному світі. Такі підходи погано адаптуються до нетипових кейсів, не враховують поведінкових особливостей клієнтів та мають обмежену прогностичну здатність.

У банках Республіки Узбекистан оцінка платоспроможності юридичних осіб здійснюється на основі трьох основних та двох додаткових фінансових коефіцієнтів [5]. Ці показники дозволяють оцінити лише поточну ліквідність підприємства, але не охоплюють важливі фінансові аспекти, такі як платіжне навантаження або рівень чистого прибутку. Вони подані в таблиці 1.1:

Таблиця 1.1. Основні фінансові коефіцієнти для оцінки платоспроможності клієнтів-юридичних осіб (за даними банків Узбекистану)

| № | Фінансовий коефіцієнт | Клас I | Клас II | Клас III |
|---|-----------------------------|----------|--------------------|--------------------|
| 1 | Коефіцієнт ліквідності (CL) | > 1.5 | $1.5 > CL > 1.0$ | $1.0 > CL > 0.5$ |
| 2 | Коефіцієнт покриття (CC) | > 2.0 | $2.0 > CC > 1.0$ | $1.0 > CC > 0.5$ |
| 3 | Коефіцієнт автономії (CA) | $> 60\%$ | $60\% > CA > 30\%$ | $30\% > CA > 15\%$ |

У практиці розвинутих країн, зокрема в США, банки застосовують більш комплексний підхід, що враховує також швидку ліквідність, фінансовий леверидж та фінансовий маржинальний коефіцієнт [5].

Таблиця 1.2. Фінансові коефіцієнти, що використовуються банком Нью—Йорка (США)

| Показник | Нормативне значення |
|------------------------------------|---------------------|
| Поточна ліквідність | 2 : 1 |
| Швидка ліквідність | 1 : 1 |
| Фінансовий леверидж | 1 : 1 |
| Фінансовий маржинальний коефіцієнт | До 1 |

Це свідчить про необхідність розширення методики оцінки кредитоспроможності, зокрема для малого та середнього бізнесу, через запровадження додаткових коефіцієнтів: маржі прибутку, рівня обслуговування боргу, коефіцієнта оборотності кредиторської заборгованості тощо.

В Україні, як і в багатьох країнах, досі спостерігається значна частка тіньової економіки, в якій доходи отримуються неофіційно. Це унеможливорює коректну оцінку реального фінансового стану клієнта за допомогою звичних аналітичних методів і змушує банки або відмовляти у кредитуванні, або приймати рішення з підвищеним рівнем ризику.

Курсова нестабільність, інфляційні процеси, військові дії, зміни у фіскальній та монетарній політиці — усі ці фактори знижують передбачуваність економічної поведінки клієнтів. Навіть надійні позичальники можуть втратити джерела доходу внаслідок зовнішніх кризових явищ, що робить оцінку платоспроможності складнішою й вимагає врахування більшої кількості зовнішніх змінних [6].

Хоча в банківській сфері активно впроваджуються ІТ-технології, часто спостерігається розрив між наявними інструментами та реальними потребами аналітиків. Наприклад, відсутність зручної інтеграції з базами даних (державними, кредитних бюро, соціальних мереж тощо), низький рівень автоматизації процесів перевірки або обмежена функціональність внутрішніх скорингових систем ускладнюють прийняття оперативних і точних рішень.

У багатьох банках остаточне рішення щодо надання кредиту приймає кредитний комітет або окремих співробітник. Навіть при наявності автоматизованої оцінки, суб'єктивна думка може мати вирішальне значення.

Це відкриває шлях до упередженості, неформальних підходів і навіть до корупційних ризиків. Надмірна довіра до «відчуттів» кредитного експерта може нівелювати результати аналітичної моделі. У ряді країн (зокрема, у країнах, що розвиваються) існує відсутність єдиних стандартів оцінки кредитоспроможності, нормативи часто не узгоджуються між різними банками, що ускладнює взаємну перевірку клієнтів та формування об'єктивної кредитної репутації.

У деяких країнах, зокрема пострадянського простору, спостерігається слабке розуміння клієнтами важливості кредитної історії. Позичальники нерідко нехтують строками виплат, змінюють номери телефонів, не співпрацюють із банками у складних фінансових ситуаціях. Це ускладнює формування повної картини платоспроможності та прогнозування поведінки клієнта в майбутньому.

1.3. Огляд традиційних підходів до аналізу кредитного ризику

Аналіз кредитного ризику є ключовим етапом у процесі прийняття рішення про надання кредиту. Його головна мета — оцінити ймовірність того, що позичальник не зможе виконати зобов'язання за кредитним договором, тобто не поверне кредит або сплатить його з порушенням умов. Традиційні підходи до оцінювання кредитного ризику ґрунтуються на статистичних та експертних методах, які, хоча й мають обмеження, залишаються основою в більшості банківських систем [7].

Метод фінансових коефіцієнтів є одним із найстаріших і найбільш розповсюджених традиційних підходів до оцінки кредитного ризику, що застосовується банками для визначення кредитоспроможності позичальника. Він ґрунтується на аналізі фінансової звітності клієнта та розрахунку ряду коефіцієнтів, які відображають різні аспекти його фінансового стану:

ліквідність, платоспроможність, прибутковість, фінансову стійкість, ефективність використання ресурсів тощо. Основною метою цього підходу є виявлення слабких місць у структурі активів і пасивів компанії, а також оцінка її здатності обслуговувати поточні та довгострокові зобов'язання.

Найбільш поширеними показниками є коефіцієнт поточної ліквідності, що показує здатність покривати короткострокові зобов'язання оборотними активами; коефіцієнт автономії, який відображає частку власного капіталу у фінансуванні активів; коефіцієнт покриття боргу, що демонструє здатність генерувати прибуток для обслуговування заборгованості; та рентабельність активів або капіталу як індикатор прибутковості [8]. Значення цих коефіцієнтів порівнюються з нормативними або середньогалузевими рівнями, на основі чого здійснюється попередня класифікація позичальника за рівнем ризику.

Цей метод є простим у застосуванні, дозволяє швидко отримати загальну картину фінансового стану клієнта та є зручним для автоматизації. Проте його недоліком є надмірна залежність від якості наданої звітності, відсутність можливості враховувати нефінансові фактори (поведінкові, ринкові, репутаційні) і низька адаптивність до нестандартних ситуацій. Крім того, цей підхід зазвичай оперує статичною інформацією за попередні періоди, що обмежує його прогностичну здатність у сучасних умовах високої динаміки фінансового середовища. Незважаючи на це, метод фінансових коефіцієнтів залишається базовим інструментом первинного аналізу та основою для побудови більш складних моделей оцінки кредитного ризику.

Метод експертного оцінювання, або анкетування позичальника, є традиційним якісним підходом до аналізу кредитного ризику, який широко застосовується у банківській практиці, особливо при роботі з юридичними особами та малим бізнесом. Його суть полягає у зборі суб'єктивної та об'єктивної інформації про позичальника на основі спеціально розробленої анкети, яка заповнюється під час розгляду заявки на кредит.

Анкета позичальника, охоплює такі розділи: загальні відомості про компанію (назва, форма власності, дата реєстрації), фінансові показники (дохід, прибуток, заборгованість, активи), характеристика керівництва (досвід, репутація, наявність негативної історії), наявність забезпечення, галузевий ризик, історія взаємодії з банком, кредитна історія [9]. Отримані дані піддаються аналізу спеціалістами банку, які на підставі власного досвіду та експертних знань формують загальне уявлення про ступінь ризику, пов'язаного з позичальником.

Цей метод є зручним у випадках, коли фінансова звітність клієнта є неповною або складною для інтерпретації, а також тоді, коли важливо врахувати нефінансові фактори, наприклад репутацію, ринкову позицію або якість управління. Однак експертна оцінка суттєво залежить від людського фактора, що може знижувати об'єктивність аналізу та відкриває ризик суб'єктивізму й помилок.

У таблиці 1.3. наведено основні переваги та недоліки методу експертного оцінювання:

Таблиця 1.3. Переваги та недоліки методу експертного оцінювання

| Переваги | Недоліки |
|---|--|
| Можливість врахування нефінансових аспектів (репутація, досвід, довіра) | Високий рівень суб'єктивності результатів |
| Гнучкість у підході до оцінки нетипових або нових клієнтів | Неможливість стандартизації та автоматизації процесу |
| Доступність для застосування за відсутності повної фінансової звітності | Значні витрати часу та ресурсів на обробку кожного запиту |
| Врахування специфіки галузі, ринку, регіону, особистих факторів | Ризик людських помилок, конфлікту інтересів або упередженості |
| Актуально для мікро—, малого та середнього бізнесу | Обмежена прогностична здатність у порівнянні з аналітичними та автоматизованими моделями |

Кредитний скоринг є одним з найпоширеніших та найефективніших інструментів автоматизованої оцінки кредитоспроможності позичальника,

особливо у сфері роздрібного кредитування. Його застосування дозволяє банкам та фінансовим установам оперативно приймати рішення щодо доцільності надання кредиту, базуючись на кількісних і якісних характеристиках потенційного клієнта.

Суть скорингового підходу полягає у присвоєнні кожному заявнику індивідуального числового балу (score), який відображає рівень його кредитного ризику. Цей бал обчислюється на основі математичної моделі, яка аналізує вплив різних факторів на ймовірність дефолту позичальника. Чим вищий скоринговий бал, тим нижчий кредитний ризик, і навпаки [10].

До основних змінних, що використовуються в скорингових моделях, зазвичай входять:

- 1) фінансові показники (рівень доходу, стабільність роботи, наявність поточних кредитів);
- 2) демографічні характеристики (вік, освіта, сімейний стан, регіон проживання);
- 3) кредитна історія (наявність прострочень, закриті кредити, кредитне навантаження);
- 4) поведінкові фактори (тривалість використання банківських послуг, частота звернень, транзакційна активність).

Перевагою скорингу є його висока швидкість, об'єктивність і масштабованість. Оскільки система базується на формалізованих правилах і статистичних залежностях, вона виключає суб'єктивізм, притаманний експертним методам. Крім того, скоринг легко інтегрується в електронні банківські системи та дозволяє обробляти велику кількість заявок у режимі реального часу.

Залежно від специфіки установи та цілей кредитування, застосовуються різні типи скорингу:

- 1) application scoring — оцінка нового клієнта під час подання заявки [9];

- 2) behavioral scoring — оцінка вже існуючого клієнта на основі його поточної поведінки [9];
- 3) collection scoring — скоринг для визначення стратегії стягнення боргів з проблемних клієнтів [9].

Найчастіше скорингові моделі створюються за допомогою логістичної регресії, однак у сучасних умовах активно впроваджуються алгоритми машинного навчання (Random Forest, Gradient Boosting, нейронні мережі), які дозволяють моделювати складніші залежності та підвищувати точність прогнозування [11].

Разом з тим, скорингові системи мають і певні обмеження. Зокрема, їх ефективність залежить від якості та обсягу навчальних даних, а також від регулярного оновлення моделі, оскільки поведінкові патерни клієнтів змінюються з часом. Крім того, при роботі з новими клієнтами, які не мають кредитної історії, точність скорингової оцінки суттєво знижується — це явище називається «проблема нового позичальника» (cold start).

Незважаючи на ці виклики, кредитний скоринг є невід’ємною частиною сучасної банківської системи. Його поєднання з інтелектуальним аналізом даних та гібридними підходами забезпечує високу ефективність у виявленні надійних позичальників та управлінні кредитними ризиками. Усе частіше банки впроваджують адаптивні скорингові моделі, які не лише приймають рішення, але й навчаються в процесі роботи, удосконалюючи свої прогнози на основі нових даних.

Для узагальнення викладених підходів та кращого розуміння їх переваг, обмежень і сфер застосування, у таблиці 1.4. Порівняння традиційних методів оцінки кредитного ризику наведено порівняння трьох традиційних методів оцінки кредитного ризику: методу фінансових коефіцієнтів, методу експертного оцінювання та кредитного скорингу. Це дозволяє виявити їхні ключові особливості та визначити доцільність використання кожного підходу в залежності від типу клієнта та завдань банку.

Таблиця 1.4. Порівняння традиційних методів оцінки кредитного ризику

| Критерій | Метод фінансових коефіцієнтів | Метод експертного оцінювання | Кредитний скоринг |
|-------------------------------|--|---|--|
| Тип оцінки | Кількісний | Якісно-суб'єктивний | Кількісний |
| Основні джерела даних | Фінансова звітність | Анкета, інтерв'ю, історія взаємодії | Анкета, кредитна історія, транзакції |
| Факторний аналіз | Фінансові коефіцієнти (ліквідність, автономія, покриття) | Репутація, досвід, ринкова позиція, наявність застави | Демографічні, фінансові та поведінкові фактори |
| Суб'єктивність | Низька | Висока | Низька (але залежить від налаштування моделі) |
| Автоматизація | Обмежена | Практично відсутня | Висока, легко інтегрується в ІТ-системи |
| Час на прийняття рішення | Середній | Тривалий | Мінімальний (у режимі реального часу) |
| Точність прогнозування | Помірна | Залежить від досвіду експерта | Висока при наявності якісної навчальної вибірки |
| Гнучкість до нетипових кейсів | Низька | Висока | Середня |
| Залежність від якості даних | Висока | Висока | Висока, але може частково компенсуватись адаптивними алгоритмами |
| Сфера застосування | Корпоративне кредитування, МСБ | МСБ, складні кейси, стартапи | Роздрібне кредитування, споживчі кредити, мікrokредитування |

1.4. Аналіз та порівняння з існуючими рішеннями

У процесі розвитку банківських технологій було запропоновано чимало рішень щодо автоматизованої оцінки кредитоспроможності позичальників.

Кожне з них базується на різних теоретичних і практичних підходах — від класичних статистичних моделей до сучасних алгоритмів машинного навчання. Далі проведено аналіз найбільш поширених рішень, які застосовуються як в українських, так і в зарубіжних банках, із метою виявлення їхніх сильних та слабких сторін у контексті ефективної оцінки кредитного ризику.

Станом на сьогодні в практиці банків можна виділити дві основні групи рішень.

1. Традиційні скорингові моделі, які ґрунтуються на логістичній регресії, статистичному аналізі та фіксованих правилах прийняття рішення (rule-based systems) [12].
2. Сучасні системи кредитного скорингу, що базуються на інтелектуальному аналізі даних, зокрема — на алгоритмах машинного навчання, нейронних мережах та ансамблевих моделях (Random Forest, XGBoost, Gradient Boosting Machines тощо).

Традиційні моделі, хоч і мають відносно невисоку точність, вирізняються високою інтерпретованістю результатів. Це є важливою перевагою для банків, які повинні забезпечувати прозорість ухвалених рішень згідно з вимогами регуляторів. Однак такі моделі не враховують складних взаємозв'язків між змінними, тому часто поступаються сучасним підходам за точністю прогнозування.

Системи, що використовують машинне навчання, демонструють вищу ефективність у виявленні дефолтних клієнтів. Наприклад, згідно з даними досліджень, моделі на основі ансамблів дерев рішень (Random Forest) або градієнтного бустингу забезпечують на 10–15% вищу точність у порівнянні з логістичною регресією. Крім того, новітні моделі здатні обробляти великі обсяги даних та адаптуватися до змін у поведінці клієнтів у режимі реального часу. Основним викликом при їх впровадженні є складність пояснення результатів (так звані «чорні скриньки») та потреба в кваліфікованих фахівцях для побудови і обслуговування моделей [13].

Деякі банки впроваджують гібридні системи, які поєднують експертні правила з машинним навчанням. Такі підходи забезпечують як інтерпретованість, так і високу точність, проте потребують ретельної побудови логіки взаємодії між компонентами.

Далі наведено короткий опис кожної з моделей, які широко застосовуються банками.

1. Логістична регресія — це статистична модель, яка використовується для прогнозування ймовірності належності до певної категорії [7] (наприклад, «платежеспроможний» чи «неплатежеспроможний»). Вона працює шляхом побудови лінійної комбінації вхідних змінних та застосування логістичної функції для прогнозування ймовірності. Логістична регресія є однією з найпростіших і найбільш зрозумілих моделей для кредитного скорингу, однак може бути менш ефективною в разі наявності складних взаємозв'язків між ознаками.

2. Методи, засновані на правилах, працюють шляхом побудови певних умовних тверджень (правил), які визначають, чи є клієнт платоспроможним. Це можуть бути прості правила (наприклад, якщо дохід клієнта більше певної суми, то він має високу ймовірність бути платоспроможним), або складніші набори умов. Ці методи є прозорими та легко інтерпретованими, однак можуть бути обмеженими у точності через свою жорсткість і відсутність здатності до адаптації.

3. Random Forest — це ансамблева модель, що складається з множини рішень, побудованих на різних підмножинах даних. Кожне дерево намагається зробити прогнози, а потім їх результати комбінуються для отримання фінального прогнозу [13]. Random Forest добре працює з великими наборами даних, має високу стійкість до перенавчання та є досить потужним у випадках, коли є складні взаємозв'язки між змінними.

4. Gradient Boosting — це метод ансамблю, який використовує слабкі моделі (зазвичай дерева рішень), комбінуючи їх у сильну модель [5]. XGBoost — це одна з найбільш популярних реалізацій цього методу, яка застосовує

спеціальні техніки для підвищення ефективності та точності моделі. XGBoost добре справляється з великими і складними наборами даних, часто перевершуючи інші методи в завданнях прогнозування [4]. Однак його налаштування можуть бути складними, і він потребує достатньо часу на тренування.

5. Нейронні мережі — це складні моделі, що імітують роботу людського мозку, використовуючи шари взаємозв'язаних нейронів для обробки вхідних даних [5]. Вони добре підходять для виявлення складних закономірностей у великих наборах даних, але можуть бути важкими для інтерпретації, що ускладнює розуміння результатів моделі. Нейронні мережі можуть вимагати великих обчислювальних ресурсів та часу для навчання.

У таблиці 1.5. Порівняння сучасних моделей для оцінки кредитоспроможності подано стислий порівняльний аналіз найбільш поширених підходів до оцінки кредитного ризику.

Таблиця 1.5. Порівняння сучасних моделей для оцінки кредитоспроможності

| Модель / Підхід | Точність | Інтерпретованість | Швидкість обчислення | Можливість масштабування | Складність впровадження |
|----------------------------------|-------------|-------------------|----------------------|--------------------------|-------------------------|
| Логістична регресія | Середня | Висока | Висока | Висока | Низька |
| Рішення на правилах (rule-based) | Низька | Висока | Висока | Середня | Низька |
| Random Forest | Висока | Середня | Середня | Висока | Середня |
| XGBoost / Gradient Boosting | Висока | Низька | Середня | Висока | Висока |
| Нейронні мережі | Дуже висока | Низька | Середня / низька | Висока | Висока |
| Гібридні моделі | Висока | Середня / висока | Середня | Середня / висока | Висока |

У зв'язку з широким впровадженням цифрових технологій та зростанням обсягів клієнтських даних, на ринку з'явилась значна кількість програмних рішень для автоматизованої оцінки кредитоспроможності. Ці продукти поєднують в собі аналітичні модулі, алгоритми скорингу та засоби прийняття рішень, адаптовані до потреб фінансових установ різного масштабу — від мікрофінансових організацій до міжнародних банків. Щоб оцінити їх потенціал, доцільно провести порівняльний аналіз низки популярних інструментів за такими критеріями, як функціональність, методи оцінювання, рівень автоматизації, можливість кастомізації та цільова аудиторія. У таблиці 1.6. представлено стислий опис восьми сучасних рішень, що активно застосовуються для побудови та використання скорингових моделей.

На сучасному фінансовому ринку функціонує низка готових програмних рішень, призначених для оцінки кредитоспроможності клієнтів. Серед найбільш відомих — SAS Credit Scoring, TurnKey Lender, GiniMachine, Zest AI, Moody's RiskAnalyst та інші. Вони пропонують комплексний підхід до побудови скорингових моделей, включаючи аналітичні модулі, інтерфейси для інтеграції з банківськими системами та готові алгоритми машинного навчання. Однак більшість таких рішень є комерційними продуктами, розробленими під конкретну проблематику або окремий сегмент користувачів (наприклад, мікрофінансові організації чи корпоративне кредитування). Їх впровадження потребує значних фінансових ресурсів, тривалого періоду адаптації, а в деяких випадках — перебудови внутрішніх ІТ—процесів банку.

У межах даної роботи буде запропоновано альтернативне рішення, засноване на використанні відкритих інструментів Python та сучасних моделей машинного навчання, таких як логістична регресія, Random Forest, XGBoost та нейронні мережі. Розроблене рішення відрізняється універсальністю, можливістю комбінування кількох моделей для підвищення точності прогнозування, а також адаптивністю до змін у структурі даних. Його основною перевагою є низька вартість реалізації при високому потенціалі

масштабування, що робить систему доцільною для впровадження у фінансових установах різного рівня.

У таблиці 1.6 представлено порівняння найбільш поширених сучасних програмних рішень для задач кредитного скорингу.

Таблиця 1.6. Порівняння сучасних рішень для оцінки кредитоспроможності

| Програмне рішення | Основне призначення | Ключові особливості | Тип користувача |
|------------------------------|---|--|--|
| SAS Credit Scoring | Побудова та впровадження скорингових моделей | Підтримка логістики, гнучкість, інтеграція з SAS | Великі банки, аналітики |
| TurnKey Lender | Автоматизація кредитного процесу для МСБ/мікрофінансів | Усе-в-одному: скоринг, прийняття рішень, API | МСБ, мікрофінансові організації |
| GiniMachine | AI-скоринг без програмування | AI-моделі, швидке впровадження, без коду | Банки, fintech-компанії |
| HighRadius Credit Management | Оцінка кредитного ризику для корпоративних клієнтів | Моніторинг ризиків, інтеграція з ERP | Корпоративні клієнти, відділи ризиків |
| FIS Credit Assessment | Фінансовий аналіз і оцінка кредитного ризику | Побудова моделей PD, підтримка звітності | Комерційні банки, аналітичні підрозділи |
| Moody's RiskAnalyst | Аналіз ризиків і дефолтність позичальників | Глибока аналітика, регуляторна відповідність | Аналітики, кредитні комітети |
| Plug&Score | Побудова скорингових карт і інтеграція в банківські системи | Гнучкість, підтримка різних типів кредитів | Банки, фінансові установи |
| Zest AI | AI-платформа для справедливого та точного скорингу | Зменшення упередженості, підвищення охоплення | Фінансові установи, які прагнуть інклюзивності |

1.5. Обмеження класичних методів у сучасних умовах

Класичні методи оцінки кредитоспроможності, як логістична регресія, дискримінантний аналіз або методи на основі правил (rule-based systems), довгий час залишалися основними інструментами в банківському секторі. Вони були ефективними в умовах обмеженого обсягу даних, простої структури взаємозв'язків між змінними та стабільного ринкового середовища. Проте із розвитком цифрових технологій, збільшенням обсягів даних (Big Data) і появою нових типів ризиків класичні методи стикаються з рядом серйозних обмежень:

Класичні статистичні моделі зазвичай працюють із даними невеликого розміру та з обмеженою кількістю змінних. У сучасних умовах банки мають справу з величезними масивами гетерогенних даних, які включають транзакційну інформацію, поведінкові показники, дані з соціальних мереж, дані геолокації тощо. Класичні методи не здатні ефективно аналізувати такі обсяги та типи інформації.

Такі методи, як логістична регресія, працюють найкраще у випадку лінійної залежності між змінними та цільовою змінною. Але у реальному житті зв'язки між змінними часто є нелінійними, що знижує точність прогнозування. Виявлення складних взаємозв'язків вимагає більш гнучких моделей, таких як ансамблеві методи або нейронні мережі.

Класичні моделі, як правило, вимагають повних і чистих наборів даних. Вони погано справляються з пропущеними значеннями, аномаліями, або некоректними даними, які часто трапляються в реальних наборах. Для їх обробки потрібна складна попередня обробка, що знижує гнучкість застосування таких моделей.

Класичні методи мають фіксовану структуру і потребують повторного навчання при зміні ринкових умов або поведінки клієнтів. В умовах високої динаміки фінансових ринків це є суттєвим недоліком. Наприклад, у періоди

економічної нестабільності або кризи, коли змінюється поведінка позичальників, класичні моделі можуть втрачати свою точність.

Вони ґрунтуються на попередньо визначених ознаках (feature engineering), і не здатні самостійно виявляти нові закономірності в даних. Це обмежує їхню ефективність у роботі з неструктурованими даними (наприклад, текстами, зображеннями, подієвими даними).

Такі моделі значною мірою покладаються на ручне визначення релевантних змінних, що вимагає участі експертів із доменною експертизою. У той час як сучасні моделі машинного навчання можуть автоматично виявляти важливі характеристики та обробляти високорозмірні простори ознак.

У багатьох практичних кейсах було показано, що класичні методи поступаються ансамблевим алгоритмам (наприклад, XGBoost, LightGBM) та глибокому навчанню в плані точності класифікації клієнтів за рівнем кредитного ризику.

Зважаючи на перелічені обмеження, класичні методи все частіше комбінуються або замінюються сучасними інструментами інтелектуального аналізу даних, які краще справляються з високою складністю, обсягом і варіативністю фінансових даних. Проте класичні підходи й досі використовуються як базові або додаткові інструменти через свою простоту, інтерпретованість і відповідність нормативним вимогам (наприклад, у кредитному скорингу згідно з Basel II/III).

1.6. Постановка задачі

У сучасних умовах цифровізації фінансових послуг банки стикаються з необхідністю оперативно та точно оцінювати кредитну спроможність потенційних та існуючих клієнтів. Традиційні підходи до скорингу клієнтів,

що базуються на статистичних моделях та фіксованих правилах, часто не здатні врахувати всю складність, варіативність та обсяги наявних даних. Це призводить до зниження якості прийняття рішень щодо кредитування, підвищення ризику дефолтів та втрат банківських установ.

Основною проблемою є обмеженість класичних методів у контексті:

- 1) виявлення нелінійних залежностей між параметрами клієнтів;
- 2) обробки великих та неструктурованих обсягів даних;
- 3) гнучкого адаптування до змін у ринковому середовищі.

Актуальність задачі зумовлена тим, що якісна оцінка кредитоспроможності напряму впливає на стабільність банківської системи та ефективність управління ризиками. Відсутність точних і гнучких інструментів прогнозування ускладнює прийняття обґрунтованих рішень, особливо у випадках нових клієнтів або нестандартних фінансових сценаріїв.

Серед існуючих рішень у банківській сфері найчастіше використовуються традиційні статистичні моделі (наприклад, логістична регресія), експертні системи на основі фіксованих правил, а також скорингові системи першого покоління. Однак ці підходи поступаються сучасним алгоритмам машинного навчання, які демонструють вищу точність, стійкість до зміни даних і здатність адаптуватися до нових ситуацій.

Інтелектуальний аналіз даних (Data Mining, Machine Learning, AI) відкриває нові можливості для підвищення ефективності процесу оцінки кредитоспроможності шляхом застосування алгоритмів машинного навчання, ансамблевих методів, нейронних мереж та інших інструментів глибокої аналітики [13].

У рамках дипломної роботи передбачається реалізація наступного:

- 1) аналіз проблеми оцінювання кредитної спроможності та основних викликів, з якими стикаються банки при оцінці кредитного ризику.
- 2) огляд сучасних підходів до аналізу кредитного ризику з акцентом на логістичну регресію, нейронні мережі, Random Forest, XGBoost та інші моделі машинного навчання.

3) вибір та обґрунтування алгоритмів, найбільш адаптованих для обробки фінансових даних, з урахуванням їх точності, стійкості до надлишкових/відсутніх даних та інтерпретованості.

4) розробка стратегії збору та попередньої обробки даних, що включає фінансову історію клієнтів, кредитну історію, демографічні характеристики, економічні індикатори, з подальшим очищенням, нормалізацією та валідацією.

5) створення прототипу аналітичної системи, що застосовує вибрані моделі для оцінки кредитоспроможності, з можливістю подальшої інтеграції з банківськими інформаційними системами.

6) формування аналітичної частини роботи, що включає опис використаних моделей, методів, структури даних, архітектури рішення та оцінки ефективності.

Підсумовуючи, метою роботи є підвищення точності, надійності та адаптивності системи оцінки кредитоспроможності за допомогою сучасних інструментів інтелектуального аналізу даних. Це дозволить банківським установам приймати більш обґрунтовані та ефективні кредитні рішення, зменшити частку проблемних кредитів і підвищити загальну ефективність управління ризиками.

1.7. Висновки до розділу 1

У першому розділі дипломної роботи було розглянуто теоретичні основи процесу оцінки кредитоспроможності клієнтів банку, що є ключовим етапом у системі управління кредитним ризиком фінансових установ. З'ясовано, що кредитоспроможність відображає здатність позичальника своєчасно та повністю виконати зобов'язання за отриманими кредитами, а її

точне оцінювання є критично важливим для забезпечення фінансової стабільності банку.

Аналіз сучасного стану в банківській сфері показав наявність низки проблем у процесі оцінювання платоспроможності клієнтів, зокрема: зростаючі обсяги даних, ускладнення поведінкових моделей позичальників, поява нових джерел інформації, нестабільність економічного середовища та необхідність швидкого прийняття рішень. Традиційні підходи, такі як логістична регресія та rule-based моделі, мають обмеження у точності прогнозів, гнучкості до змін ринку та здатності працювати з великою кількістю ознак.

Огляд і порівняння сучасних моделей оцінки кредитоспроможності дозволив виокремити переваги застосування підходів інтелектуального аналізу даних, зокрема методів машинного навчання, нейронних мереж, ансамблевих алгоритмів (Random Forest, XGBoost), які демонструють вищу точність, адаптивність та здатність до виявлення складних взаємозв'язків у даних.

У результаті аналізу сформовано постановку задачі роботи, яка передбачає розробку прототипу аналітичної системи оцінки кредитоспроможності клієнтів банку із використанням сучасних алгоритмів інтелектуального аналізу даних, з урахуванням специфіки фінансових даних та можливістю інтеграції з банківськими інформаційними системами.

РОЗДІЛ 2. ЗАСТОСУВАННЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ У ЗАДАЧАХ КРЕДИТНОГО СКОРИНГУ

2.1. Сутність інтелектуального аналізу даних (Data Mining) у фінансовому секторі

У сучасному цифровому середовищі фінансові установи акумулюють величезні обсяги даних про клієнтів, транзакції, кредити, депозити, платіжну поведінку та інші аспекти діяльності. У зв'язку з цим зростає потреба у більш глибокій та точній аналітиці, яка дозволяє не лише здійснювати ретроспективний аналіз, а й робити обґрунтовані прогнози щодо поведінки клієнтів, рівня ризиків та доцільності кредитування. Саме тут на перший план виходить інтелектуальний аналіз даних (Data Mining) — процес виявлення прихованих закономірностей, зв'язків та трендів у великих наборах даних з використанням статистичних методів, машинного навчання, візуалізації та баз знань [14].

Data Mining у фінансовому секторі виступає інструментом трансформації «сирих» даних у цінну інформацію, що може бути використана для прийняття стратегічних і тактичних рішень. Його застосування охоплює різні напрямки, зокрема: оцінку кредитоспроможності фізичних та юридичних осіб; виявлення шахрайських транзакцій; сегментацію клієнтів; оптимізацію маркетингових кампаній; прогнозування прибутковості та фінансових ризиків.

Особливу актуальність інтелектуальний аналіз даних набуває у сфері кредитного скорингу — процесу визначення ймовірності виконання позичальником своїх зобов'язань [14]. Завдяки використанню методів Data Mining банки мають можливість підвищити точність оцінки ризиків, що, у

свою чергу, сприяє зменшенню частки проблемних кредитів та оптимізації кредитного портфеля.

У рамках інтелектуального аналізу даних використовуються різноманітні підходи та алгоритми. Найбільш поширеними є:

1) класифікація — побудова моделей, які прогнозують клас об'єкта (наприклад, «кредит буде повернуто» або «буде прострочення платежа») [15];

2) кластеризація — поділ клієнтів на однорідні групи без попереднього визначення класів [15];

3) регресійний аналіз — оцінка кількісної залежності однієї змінної від інших [14];

4) асоціативний аналіз — виявлення частих патернів або правил, наприклад, що певний тип клієнта часто затримує платежі [16];

5) аналіз аномалій — виявлення нестандартної поведінки, що може свідчити про ризики чи шахрайство.

Перевагою Data Mining є здатність виявляти складні, нелінійні взаємозв'язки, які традиційні статистичні методи не в змозі ефективно відобразити. Наприклад, у випадку кредитного скорингу, класичні моделі можуть не враховувати взаємодію декількох змінних одночасно, тоді як алгоритми машинного навчання (наприклад, Random Forest або XGBoost) здатні виявити навіть найменші закономірності в поведінці позичальників.

Водночас впровадження Data Mining у фінансовому секторі потребує належної підготовки даних, дотримання етичних та правових норм (особливо у сфері захисту персональних даних), а також пояснюваності результатів — особливо у випадку використання алгоритмів, які функціонують як «чорна скринька».

2.2. Роль штучного інтелекту та машинного навчання в банківській аналітиці

Упродовж останніх десятиліть банківська сфера зазнала суттєвих трансформацій, значною мірою завдяки впровадженню цифрових технологій. Проте справжній прорив у розвитку аналітики в банках став можливим із розширеним застосуванням методів штучного інтелекту (ШІ) та машинного навчання (ML). Ці підходи дозволяють банкам переходити від базових систем обробки даних до розумних систем підтримки рішень, які здатні адаптуватися до динамічних умов ринку, змін у поведінці клієнтів і нових викликів у сфері ризиків.

Штучний інтелект у банківській аналітиці розглядається як сукупність алгоритмів, які моделюють когнітивні функції людини: аналіз, прогнозування, оцінка, прийняття рішень. Машинне навчання, у свою чергу, є частиною ШІ, що базується на здатності систем самостійно «вчитись» на основі наявних даних та покращувати свою точність без жорсткого програмування правил.

Особливо помітною стала роль машинного навчання у кредитному скорингу, аналітиці клієнтської поведінки, виявленні шахрайства, управлінні ризиками та автоматизації прийняття рішень. Там, де раніше використовувались жорстко структуровані математичні моделі, сьогодні застосовуються адаптивні системи, що здатні виявляти приховані закономірності та враховувати численні, часто слабо виражені, фактори.

Один з головних чинників зростання інтересу до застосування ML в аналітиці банківських процесів — це ефективність алгоритмів у роботі з великими масивами даних, які включають не лише структуровані таблиці, а й неструктуровану інформацію: тексти, зображення, аудіо та поведінкові сліди клієнтів. Банки здобули можливість аналізувати не лише фінансові показники, а й соціальні зв'язки, онлайн-активність, патерни витрат та навіть емоційні реакції клієнтів у зверненнях до служби підтримки [17].

Особливість сучасних підходів у тому, що алгоритми не просто підбирають рішення на основі середніх тенденцій, а враховують індивідуальні особливості кожного об'єкта. Завдяки цьому можлива побудова персоналізованих прогнозів, адаптованих до конкретного клієнта, що значно підвищує ефективність рішень щодо надання кредиту або визначення оптимальної пропозиції.

Для порівняння традиційних статистичних методів (Логістичної регресії, дискримінантний аналіз, класичні регресійні моделі, тощо) і моделей машинного навчання, що застосовуються у банківській аналітиці, наведено узагальнюючу таблицю 2.1:

Таблиця 2.1. Порівняння традиційних методів і моделей машинного навчання у банківській аналітиці

| Критерій порівняння | Традиційні методи (логістична регресія, класичні регресійні моделі) | Машинне навчання |
|------------------------------------|--|---|
| Гнучкість моделей | Обмежена, потребує чіткого формулювання | Висока, здатні адаптуватись до даних |
| Залежність від припущень | Висока (наприклад, нормальність розподілу) | Низька, можуть працювати з «сирими» даними |
| Робота з великою кількістю змінних | Обмежена (ризик переобтяження моделі) | Можливість обробки десятків і сотень ознак |
| Інтерпретованість результатів | Висока (коефіцієнти зрозумілі) | Часто низька (особливо в нейронних мережах) |
| Якість прогнозу | Обмежена точність | Висока точність при належному навчанні |
| Адаптація до нових умов | Потребує повторного налаштування | Самооновлення через донавчання |

Незважаючи на численні переваги, впровадження ШІ в банках не позбавлене труднощів. Однією з головних проблем є забезпечення прозорості рішень. Багато моделей ML, особливо глибокі нейронні мережі, є «чорними скриньками», що ускладнює пояснення, чому саме було ухвалено певне рішення. Для банківської галузі, яка регулюється численними нормативами, це може стати критичним. У відповідь на це активно розвивається напрям

explainable AI-пояснюваного штучного інтелекту, який дозволяє відновити логіку прийнятого алгоритмом рішення [17].

Крім того, зважаючи на важливість етичного підходу, великого значення набуває уникнення упередженості у даних. Якщо моделі навчаються на історичних даних, де мали місце дискримінаційні практики, то і самі алгоритми можуть їх відтворювати. Це вимагає ретельної роботи з підготовкою даних, фільтрацією та нормалізацією, а також постійного моніторингу результатів.

Попри виклики, потенціал штучного інтелекту та машинного навчання для банківської сфери є надзвичайно широким. Йдеться не лише про оптимізацію окремих процесів, а про зміну самого підходу до бізнесу — від реактивного до проактивного, де рішення приймаються не після настання події, а до її виникнення. Це відкриває шлях до більш гнучкого управління ризиками, персоналізованого сервісу та сталого розвитку в умовах високої конкуренції та зростаючих очікувань клієнтів.

2.3. Огляд сучасних моделей для оцінки кредитоспроможності

Оцінка кредитоспроможності клієнтів банку — це завдання, яке традиційно розглядалося крізь призму статистичних методів. Однак із розвитком інструментів машинного навчання відбулося зміщення акцентів у бік більш адаптивних, точних та автоматизованих підходів. На сьогодні існує кілька основних моделей, які активно використовуються для аналізу кредитного ризику. Кожна з них має свої переваги та обмеження, що впливають на вибір конкретного алгоритму в залежності від поставлених завдань, якості даних, а також вимог до інтерпретованості результатів.

Однією з найстаріших і водночас найпоширеніших моделей, яка до сьогодні зберігає актуальність, є логістична регресія. Цей метод

застосовується у випадках, коли необхідно здійснити бінарну класифікацію, наприклад, визначити, чи поверне клієнт кредит [18]. Сутність логістичної регресії полягає у побудові математичної функції, яка дозволяє оцінити ймовірність того, що певний позичальник належатиме до однієї з двох категорій. Основною перевагою такого підходу є його простота, швидкість навчання моделі та високий рівень інтерпретованості: банки можуть легко пояснити, які саме фактори вплинули на прийняття рішення щодо клієнта. Проте ця модель менш ефективна у випадках, коли залежності між ознаками є складними, нелінійними, або коли дані мають сильну кореляцію.

Логістична регресія моделює ймовірність того, що результат (наприклад, повернення кредиту) належить до класу 1 (успіх) за допомогою сигмоїдної функції (2.1)

$$P(y=1|x) = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-\beta_0+\beta_1x_1+\beta_2x_2+\dots+\beta_nx_n}}, \quad (2.1)$$

де $X=(x_1, x_2, \dots, x_n)$ — вектор ознак (характеристики клієнта);

β_0 — вільний член (зсув);

β_i — ваговий коефіцієнт при i -й ознаці;

$z=\beta_0+\beta_1x_1+\dots+\beta_nx_n$ — лінійна комбінація ознак;

$P(y=1|x)$ — імовірність того, що результат належить до позитивного класу (наприклад, «клієнт поверне кредит»).

Альтернативою логістичній регресії є дерева рішень — моделі, які створюють ієрархічну структуру, де кожне правило веде до прийняття певного рішення. Дерева мають природну інтерпретованість і можуть обробляти як числові, так і категоріальні ознаки. Однак базові дерева мають обмежену здатність до узагальнення — вони схильні до перенавчання, особливо на малих вибірках, що робить їх менш надійними без додаткових технік оптимізації [19].

Для подолання цих недоліків були розроблені ансамблеві методи, серед яких особливої популярності набули Random Forest і Gradient Boosting. Random Forest полягає у побудові великої кількості дерев рішень, кожне з яких

навчається на випадковій підмножині даних, а кінцеве рішення приймається шляхом голосування.

Random Forest — це ансамблевий метод, що поєднує багато дерев рішень, кожне з яких тренується на випадковій підмножині даних та ознак. Результат класифікації отримується шляхом голосування (в задачах класифікації) або усереднення (в задачах регресії) [13].

Вхідними даними для методу Random Forest є:

- 1) матриця ознак $X = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{n \times p}$, де n — кількість об'єктів, p — кількість ознак.
- 2) вектор цільових значень $y = [y_1, y_2, \dots, y_n]$ де y_n — клас або числове значення (для регресії) для кожного об'єкта x_n
- 3) кількість дерев у лісі K ;
- 4) кількість ознак, що випадково відбираються на кожному вузлі дерева (гіперпараметр, наприклад \sqrt{p} для класифікації).

Результатом роботи моделі є:

- 1) у задачах класифікації — клас, який отримав найбільшу кількість голосів серед усіх дерев;
- 2) у задачах регресії — середнє значення, передбачене всіма деревами.

Псевдокод алгоритму Random Forest:

Input: навчальна вибірка $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, кількість дерев K .

Для $k = 1$ до K :

1. Вибрати випадкову підмножину D^k з D з поверненням (бутстреп);
2. Побудувати дерево рішень T_k на D^k ;
3. На кожному вузлі:
 - Випадковим чином вибрати m ознак з p ;
 - Вибрати найкращу ознаку для розбиття за критерієм (наприклад, Gini).

Дерево росте до максимального розміру або досягнення обмеження.

Output (класифікація). Для нового об'єкта x :

- отримати передбачення від кожного дерева $T_1(x), \dots, T_k(x)$;
- повернути моду (найчастіший клас)

$$\hat{y} = \text{mode}\{T_1(x), T_2(x), \dots, T_k(x)\}.$$

Формалізація (для класифікації):

Нехай $T_1(x), T_2(x), \dots, T_K(x)$ — дерева класифікації, натреновані на випадкових підмножинах даних, тоді передбачення Random Forest — це мода:

Для регресії:

$$\hat{y} = \frac{1}{K} \sum_{k=1}^K T_k(x), \quad (2.2)$$

де $T_k(x)$ — результат k -го дерева на вхідному векторі x ;

K — загальна кількість дерев у лісі.

Цей метод демонструє високу точність, стійкість до шумів у даних та здатність до узагальнення. Gradient Boosting, у свою чергу, побудований на принципі послідовного навчання слабких моделей, кожна з яких намагається компенсувати помилки попередніх. У фінансовій аналітиці особливо широко застосовуються варіації цього методу, такі як XGBoost або LightGBM, які поєднують точність із високою продуктивністю. Обидва підходи менш інтерпретовані, ніж логістична регресія або прості дерева, але переважають за точністю та гнучкістю.

Далі розглянемо ансамблеві методи. Вони є одними з найбільш ефективних і широко застосовуваних підходів у сучасному машинному навчанні. Базуються на ідеї поєднання кількох окремих моделей («базових учнів» або weak learners) для побудови більш стабільної, точної та узагальнюючої моделі. Концептуально, ансамблеві методи використовують колективний інтелект моделей для зменшення варіації (variance), зміщення (bias) або ризику перенавчання.

Нехай $H = \{h_1(x), h_2(x), \dots, h_K(x)\}$ — набір базових моделей. Ансамблевий метод створює підсумкову модель $H(x)$ як деяку комбінацію цих моделей. У загальному вигляді:

$$H(x) = \sum_{k=1}^K \alpha_k h_k(x), \quad (2.3)$$

де $h_k(x)$ — k -та базова модель;

α_k — вага моделі (у простих випадках $\alpha_k = \frac{1}{K}$).

Залежно від способу побудови ансамблю, розрізняють такі основні підходи:

Беггінг (Bootstrap Aggregating)

Беггінг зменшує варіативність моделей шляхом навчання кожної базової моделі на різних бутстреп-вибірках (випадкових підмножинах з поверненням).

Формула прогнозу (регресія):

$$\hat{y} = \frac{1}{K} \sum_{k=1}^K h_k(x) \quad (2.4)$$

Формула (класифікація):

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \dots, h_K(x)\} \quad (2.5)$$

Приклад: Random Forest — типова реалізація беггінгу на основі дерев рішень.

Gradient Boosting — це метод, який послідовно будує моделі, де кожна наступна модель навчається на залишках (помилках) попередньої. В основі — градієнтне зниження функції втрат.

Нехай:

$F_0(x)$ — початкова модель (зазвичай середнє значення цільової змінної),

$h_m(x)$ — нова слабка модель на m -му кроці,

$\eta \in (0, 1]$ — коефіцієнт навчання (learning rate),

Тоді остаточна модель на m -му кроці будується як:

$$F_m(x) = F_{m-1}(x) + \eta * h_m(x)$$

А алгоритм можна розписати так:

Ініціалізація моделі:

$$F_0(x) = \underset{c}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, c)$$

Для кожної ітерації

$m=1,2,\dots,M$:

Обчислюємо градієнти:

$$r_i^{(m)} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$

Навчаємо слабку модель $h_m(x)$ для апроксимації залишків.

Оновлюємо модель:

$$F_m(x) = F_{m-1}(x) + \eta * h_m(x)$$

Стекинг — це ансамбль другого рівня, де прогнози кількох моделей подаються як вхідні дані до метамоделі, яка навчається робити остаточне рішення. Він дозволяє комбінувати моделі різного типу (лінійні, деревоподібні, нейронні) для досягнення більшої узагальненості.

Нехай:

$z_i = [h_1(x_i), h_2(x_i), \dots, h_K(x_i)]$ — вектор прогнозів базових моделей для прикладу x_i .

Тоді метамодель $g(z)$ повертає остаточне передбачення:

$$\hat{y} = g(z_i) \quad (2.6)$$

Переваги ансамблевих методів:

- 1) покращення точності (особливо у складних задачах);
- 2) зменшення варіативності та зміщення;
- 3) гнучкість у поєднанні різних моделей;
- 4) висока продуктивність у реальних задачах (наприклад, скоринг, fraud detection).

Недоліки:

- 1) зниження інтерпретованості;
- 2) збільшення складності моделі;
- 3) вищі обчислювальні витрати (особливо у випадку бустингу).

Ще одним класом моделей, що активно досліджуються в контексті скорингових систем, є штучні нейронні мережі. Їх здатність до виявлення складних, багатовимірних залежностей робить їх привабливими у завданнях, де традиційні підходи демонструють обмежену ефективність. Нейронні мережі можуть працювати з великими наборами даних, адаптуватися до динамічних змін ринку та виявляти приховані закономірності. Проте їх значним недоліком є слабка інтерпретованість, потреба у великій кількості обчислювальних ресурсів та даних, а також ризик перенавчання без належної регуляризації.

Для наочності порівняння характеристик основних моделей, які використовуються в оцінці кредитоспроможності, подано в таблиці 2.2:

Таблиця 2.2 — Характеристика основних моделей машинного навчання для оцінки кредитоспроможності

| Модель | Точність прогнозування | Інтерпретованість | Продуктивність (швидкість навчання/передбачення) | Стійкість до надмірних ознак | Гнучкість |
|-------------------------|------------------------|-------------------|--|------------------------------|-------------|
| Логістична регресія | Середня | Висока | Висока | Низька | Низька |
| Дерева рішень | Середня | Висока | Висока | Середня | Середня |
| Random Forest | Висока | Середня | Середня | Висока | Висока |
| Gradient Boosting | Дуже висока | Низька | Середня/низька | Висока | Висока |
| Нечітка нейронна мережа | Висока/дуже висока | Дуже низька | Низька | Висока | Дуже висока |

З таблиці видно, що жодна з моделей не є універсальною. Наприклад, логістична регресія дозволяє легко інтерпретувати результати, однак поступається у точності нейронним мережам чи градієнтному бустингу. Останні ж, у свою чергу, потребують значних обчислювальних ресурсів і складні в поясненні, що може бути критично важливим для банків, які зобов'язані аргументувати відмову у кредитуванні.

Зрештою, вибір оптимальної моделі залежить не лише від точності, а й від вимог до прозорості, часу обробки, доступності обчислювальних потужностей та особливостей даних [19]. Саме тому сучасні системи кредитного скорингу все частіше базуються на гібридному підході, який поєднує декілька моделей, забезпечуючи баланс між точністю, надійністю та практичністю. Такий підхід дозволяє формувати більш стійкі до ризиків системи, які адаптуються до змін у поведінці клієнтів, ринкових умовах та нормативному середовищі.

2.4. Вибір та обґрунтування методів машинного навчання на прикладі нечіткої нейронної мережі

Оцінка кредитоспроможності є критично важливою задачею для банківських установ, оскільки вона безпосередньо впливає на прибутковість та ризик портфеля. У зв'язку з цим, вибір методів машинного навчання, які використовуватимуться в системі скорингу, повинен бути обґрунтованим, з урахуванням особливостей задачі, структури доступних даних, вимог до точності, продуктивності, інтерпретованості та можливості масштабування.

Перш за все, враховуючи специфіку предметної області, модель має добре працювати з табличними даними, які містять як числові, так і категоріальні змінні. До таких належать демографічна інформація, дані про доходи, кредитну історію, поточні зобов'язання тощо. У більшості випадків ці ознаки не мають складної внутрішньої структури, але можуть бути зв'язані між собою нелінійними залежностями. Це вказує на доцільність використання моделей, які здатні виявляти приховані кореляції, зберігаючи при цьому хорошу продуктивність.

У якості базової моделі доцільно розглядати логістичну регресію, оскільки вона є простою, швидкою у навчанні, добре інтерпретується, і

водночас дає конкурентоспроможні результати при належній підготовці даних. Проте через її обмеження у роботі з нелінійностями логістична регресія зазвичай не забезпечує найвищої точності у складних задачах.

Для досягнення вищої точності доцільним є застосування ансамблевих методів, зокрема Random Forest та Gradient Boosting (наприклад, XGBoost або LightGBM). Вони демонструють хорошу стійкість до шумів, добре працюють з табличними даними, автоматично враховують взаємозв'язки між змінними та забезпечують високу точність передбачень. Gradient Boosting особливо ефективний у завданнях скорингу, де важливо виявити неочевидні закономірності та мінімізувати кількість помилкових рішень. Водночас Random Forest є менш чутливим до параметрів моделі і може використовуватися як референтна модель для порівняння.

Нейронні мережі, попри свою високу гнучкість і здатність моделювати складні нелінійні залежності, мають обмежене поширення у фінансових установах через низьку інтерпретованість результатів і високу вимогливість до ресурсів. Проте у разі наявності великого обсягу даних, а також при реалізації систем у режимі реального часу, глибокі моделі (наприклад, багатоварові перцептрони) можуть бути застосовані як частина гібридного підходу.

Нечітка нейронна мережа (Fuzzy Neural Network, FNN) — це гібридна модель, що поєднує можливості класичних штучних нейронних мереж із концепцією нечіткої логіки, яка дозволяє ефективно працювати з невизначеними, лінгвістичними та нечітко заданими даними.

На відміну від звичайних нейронних мереж, де вхідні значення інтерпретуються як чіткі числові ознаки, FNN використовує нечіткі множини та функції приналежності, що дозволяє моделі оперувати такими поняттями, як «високий дохід», «низький ризик» тощо. Це забезпечує гнучкість у моделюванні експертних знань, які важко формалізувати.

Структура FNN включає три основні блоки:

Фазифікація входу — перетворення числових значень в ступені приналежності до нечітких множин. Наприклад, змінна «вік» може бути перетворена у ступені приналежності до категорій «молодий», «середній», «похилого віку». Для цього часто використовується сигмоїдна або гаусова функція:

$$\mu_A(x) = \frac{1}{1 + e^{-a(x-b)}} \quad , \quad (2.7)$$

де $\mu_A(x)$ — ступінь приналежності x до нечіткої множини A ;

a — крутизна;

b — центр функції.

Нечітка база правил — набір правил типу «Якщо–То», які моделюють логічні залежності у вигляді:

Якщо x_1 є «низький» і x_2 є «високий», тоді y є «середній»

Ці правила кодуються у прихованих шарах мережі та агрегуються згідно з операціями кон'юнкції (\min , \prod) та диз'юнкції (\max , \sum).

Дефазифікація — перетворення вихідного нечіткого результату у чітке числове значення. Один із найпоширеніших методів — центр тяжіння (centroid):

$$y = \frac{\sum \mu_i(x) * y_i}{\sum \mu_i(x)} \quad , \quad (2.8)$$

де $\mu_i(x)$ — ступінь спрацьовування i -го правила, y_i — значення, яке відповідає результату цього правила.

У задачі кредитного скорингу FNN дозволяє враховувати не лише точні числові значення (доходи, вік, історія платежів), а й суб'єктивні оцінки типу «нестабільне працевлаштування», «помірний ризик» або «схильність до затримки». Це робить модель більш адаптивною до реальних умов, коли інформація про позичальника може бути неповною або якісною.

З огляду на це, доцільно формувати каскадну систему, у якій на першому етапі буде реалізовано швидке фільтрування за допомогою логістичної регресії або дерева рішень, а на наступному — більш точну перевірку із залученням бустингових моделей. Така архітектура дозволяє збалансувати

швидкодію, точність та інтерпретованість системи, що є критично важливим у банківській сфері.

У рамках цієї роботи буде обрано та реалізовано декілька моделей машинного навчання, серед яких логістична регресія, Random Forest і Gradient Boosting, з подальшим порівнянням їхньої продуктивності. Такий підхід дозволить визначити найефективніший інструмент для задачі оцінки кредитоспроможності з урахуванням специфіки даних, практичних вимог банківської галузі та можливостей масштабування в реальних умовах експлуатації.

2.5. Критерії вибору оптимальної моделі для банківських задач

У задачах оцінки кредитоспроможності клієнтів банку вибір моделі не повинен ґрунтуватися виключно на показниках точності. Сфера фінансів, зокрема банківська аналітика, має низку особливих вимог до моделей, які зумовлені не лише технічними характеристиками, а й нормативними обмеженнями, прозорістю прийняття рішень, стійкістю до змін у даних та можливістю масштабування на великі обсяги операцій.

До основних критеріїв вибору оптимального алгоритму можна віднести:

Точність передбачення — критично важлива характеристика, яка визначає здатність моделі правильно класифікувати позичальників (надійних та ризикованих). У банківському контексті особливо значущими є показники Recall для класу «ризикований клієнт» (щоб не допустити видачі кредиту неблагонадійній особі), а також Precision (щоб не відмовити платоспроможним клієнтам) [20].

Інтерпретованість — важлива для пояснення результатів аналітикам, аудиторам, регуляторним органам і самим клієнтам. Моделі, які не дають

чіткої інтерпретації (наприклад, глибокі нейронні мережі), часто неприйнятні у банківській практиці, попри високу точність.

Швидкодія та продуктивність — система повинна ефективно обробляти великі обсяги даних у реальному часі, особливо при роботі у фронт—офісі, де важливо миттєво приймати рішення про надання кредиту [21].

Стійкість до переобучення — модель повинна демонструвати стабільні результати не тільки на навчальній, а й на новій, раніше не баченій інформації.

Масштабованість та автоматизованість — можливість легко оновлювати модель при зміні структури даних або розширенні бази клієнтів без потреби в ручному доопрацюванні [22].

Вимоги до обчислювальних ресурсів — деякі моделі можуть бути надто ресурсозатратними або повільними, що обмежує їх застосування на практиці [23].

Також часто застосовують середньоквадратичну похибку моделі (2.7) щоб обчислити умовну оптимальність(адекватність) моделі:

$$\text{СКП}(x_s, x_m) = \sqrt{\frac{1}{n} \sum_{k=1}^N [x_s(k) - x_m(k)]^2}, \quad (2.7)$$

де $x_s(k)$ — вимір вихідного сигналу об'єкту в момент k ;

$x_m(k)$ — оцінка виміру за моделлю.

Для систематизації цих факторів подано таблицю 2.3 порівняння ключових типів моделей за основними критеріями:

Таблиця 2.3 – Порівняння моделей машинного навчання за критеріями придатності до банківських задач

| Критерій | Логістична регресія | Дерева рішень | Random Forest | Gradient Boosting | Нечітка нейронна мережа |
|--------------------|---------------------|---------------|---------------|-------------------|-------------------------|
| Точність | Середня | Середня | Висока | Дуже висока | Дуже висока |
| Інтерпретованість | Висока | Висока | Низька | Низька | Дуже низька |
| Швидкодія навчання | Висока | Висока | Середня | Низька | Низька |

| | | | | | |
|------------------------|--------|--------|---------|---------|--------|
| Швидкодія передбачення | Висока | Висока | Середня | Середня | Низька |
|------------------------|--------|--------|---------|---------|--------|

Продовження таблиці 2.3

| | | | | | |
|---------------------------|--------|----------|-----------|-----------|-------------------------|
| Стійкість до переобучення | Низька | Середня | Висока | Висока | Залежить від реалізації |
| Масштабованість | Висока | Висока | Висока | Середня | Середня |
| Вимоги до ресурсів | Низькі | Низькі | Середні | Високі | Високі |
| Застосування в банках | Широке | Обмежене | Зростаюче | Зростаюче | Обмежене |

2.6. Висновки до розділу 2

У другому розділі було розглянуто сутність інтелектуального аналізу даних у контексті банківського сектору, зосереджено увагу на сучасних підходах до вирішення задачі оцінки кредитоспроможності клієнтів. З'ясовано, що Data Mining є важливою складовою сучасної фінансової аналітики, оскільки дає змогу виявляти приховані закономірності у великих обсягах даних та ефективно прогнозувати ризики неповернення кредитів.

Особливу увагу приділено ролі штучного інтелекту й машинного навчання як рушійної сили цифрової трансформації банківської сфери. Ці технології забезпечують підвищену точність, гнучкість і адаптивність моделей оцінювання, що особливо актуально в умовах зростання обсягів інформації та динамічних змін фінансових ринків.

Проаналізовано широкий спектр моделей — від традиційної логістичної регресії до сучасних ансамблевих методів і глибоких нейронних мереж. Встановлено, що жодна з моделей не є універсальною, і вибір оптимального інструменту має базуватись на конкретних вимогах задачі — таких як точність, інтерпретованість, швидкодія та стабільність. З цією метою було

сформовано критерії вибору моделей, які можуть бути адаптовані для реального банківського середовища з урахуванням бізнес-потреб і регуляторних обмежень.

РОЗДІЛ 3. РОЗРОБКА ПРОТОТИПУ СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ОЦІНКИ КРЕДИТОСПРОМОЖНОСТІ

3.1. Підготовка даних для моделювання

Процес побудови ефективної системи оцінки кредитоспроможності базується на якісному аналізі даних, що відображають фінансову поведінку та характеристики клієнтів банку. Перед використанням алгоритмів машинного навчання важливо пройти етапи попередньої обробки даних, налаштування моделей прогнозування та формування логіки ухвалення рішень. Цей розділ присвячено опису поетапного формування датасету та реалізації повного життєвого циклу побудови моделі кредитного скорингу.

На початковому етапі здійснюється візуалізація вихідних даних, що дозволяє отримати загальне уявлення про структуру та характеристики датасету. За допомогою графіків розподілу (гістограм), діаграм розсіювання, boxplot або теплових карт (heatmap) аналізується форма розподілу ознак, виявляються потенційні викиди, пропущені значення та дисбаланс цільової змінної. Також візуально можна оцінити наявність кореляцій між окремими ознаками, перевірити баланс класів (наприклад, кількість «надійних» та «ризикових» клієнтів) та визначити попередні гіпотези для подальшого моделювання. Цей крок є складовою розвідувального аналізу даних (Exploratory Data Analysis, EDA), що забезпечує усвідомлену підготовку до побудови моделі. На рисунку 3.1 зображено Розподіл цільової змінної по статусу.

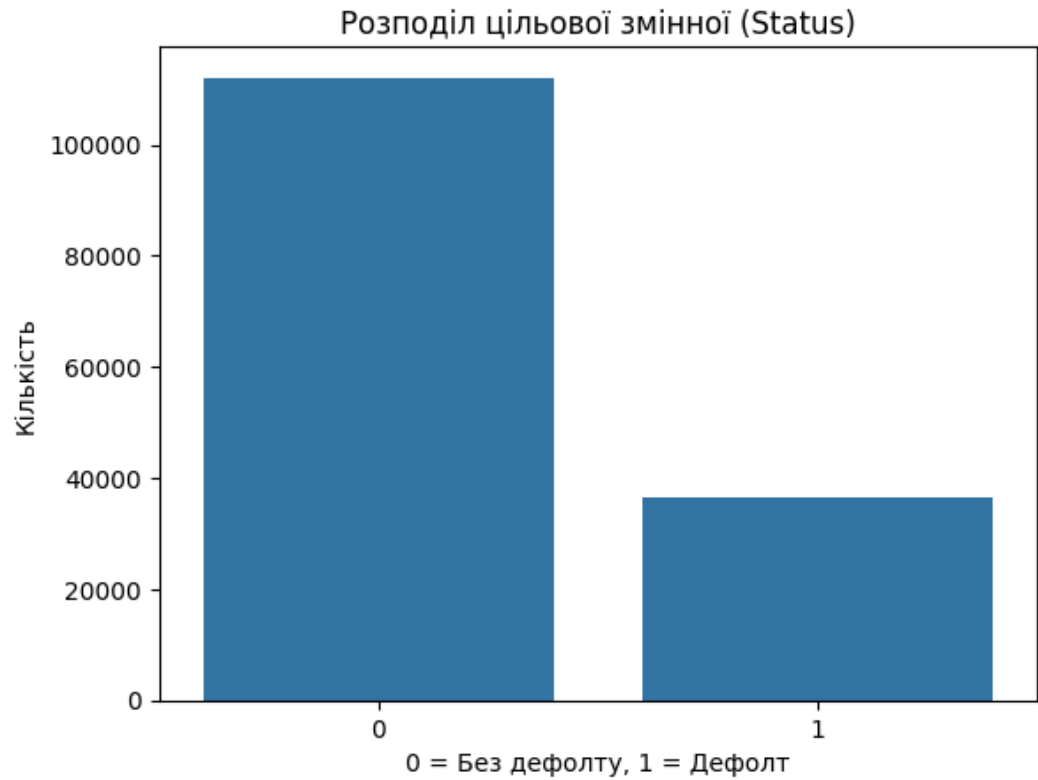


Рисунок 3.1. — Розподіл цільової змінної по статусу

Також на риунку 3.2. наведено графік гендерного розподілу

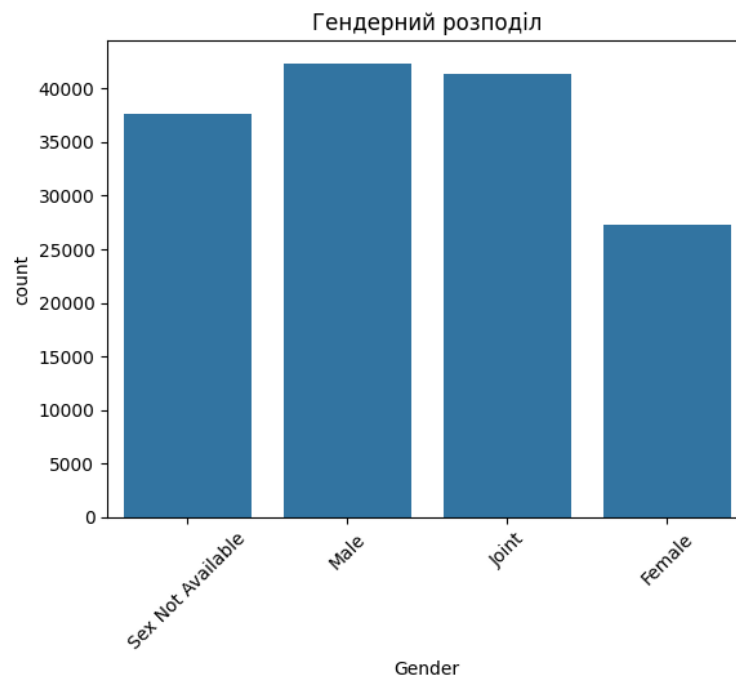


Рисунок 3.2. — Гендерний розподіл

Початкові дані, отримані з банківських інформаційних систем, зазвичай містять інформацію про кредитну історію клієнтів, демографічні ознаки,

рівень доходу, наявність активів, інтенсивність транзакцій, прострочення платежів, а також поведінкові характеристики. Однак ці дані часто є неповними, неструктурованими або зашумленими. Тому першим кроком у підготовці даних є їх очищення.

Очищення передбачає виявлення та обробку пропущених значень. У випадку числових ознак вони можуть бути заповнені середнім або медіанним значенням, а категоріальні — найчастішим значенням або індикатором «невідомо». Наприклад, якщо клієнт не вказав професію, у полі може бути створено окрему категорію «невідомо» або застосовано метод імітації (imputation).

Далі виконується кодування категоріальних змінних у числовий формат, оскільки більшість моделей машинного навчання не здатні працювати з текстовими ознаками. Наприклад, змінна «тип зайнятості» або «місце проживання» кодується через one-hot або label encoding.

Після завершення етапу попередньої обробки даних та візуалізації загального розподілу ознак, важливо виконати кореляційний аналіз. Його мета — виявити ступінь взаємозв'язку між вхідними ознаками, а також між ознаками та цільовою змінною. Цей крок дозволяє ідентифікувати надлишкові (колінеарні) ознаки, які можуть негативно впливати на продуктивність моделей машинного навчання, зокрема — на інтерпретованість результатів та стабільність моделей, таких як логістична регресія або дерева рішень.

Для числових змінних зазвичай використовується коефіцієнт кореляції Пірсона, що визначається як:

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} * \sqrt{\sum(y_i - \bar{y})^2}}, \quad (3.1)$$

де x_i та y_i - значення ознак, а \bar{x} і \bar{y} - їхні середні значення. Коефіцієнт r набуває значень від -1 до 1: значення близьке до 1 або -1 вказує на сильну пряму або обернену лінійну залежність відповідно, тоді як значення, близьке до 0, свідчить про відсутність зв'язку.

У випадках, коли присутні категоріальні змінні або нелінійні залежності, можуть застосовуватись альтернативні метрики, такі як коефіцієнт Спірмена або коефіцієнт Крамера для номінальних змінних.

У межах аналізу була побудована матриця кореляцій, що наочно демонструє взаємозв'язки між усіма числовими змінними у датасеті. Це дозволяє:

виявити пари ознак із сильною кореляцією (наприклад, > 0.8), які можуть бути об'єднані або одна з них — виключена;

зрозуміти, які ознаки найбільше пов'язані з цільовою змінною;

визначити потенційно важливі фактори, що впливають на рівень кредитного ризику.

На рисунку 3.3 представлено зображення кореляційної матриці, побудованої за допомогою бібліотеки Seaborn у середовищі Python. Теплова шкала вказує на силу кореляцій, де темніші відтінки відповідають сильнішим зв'язкам:

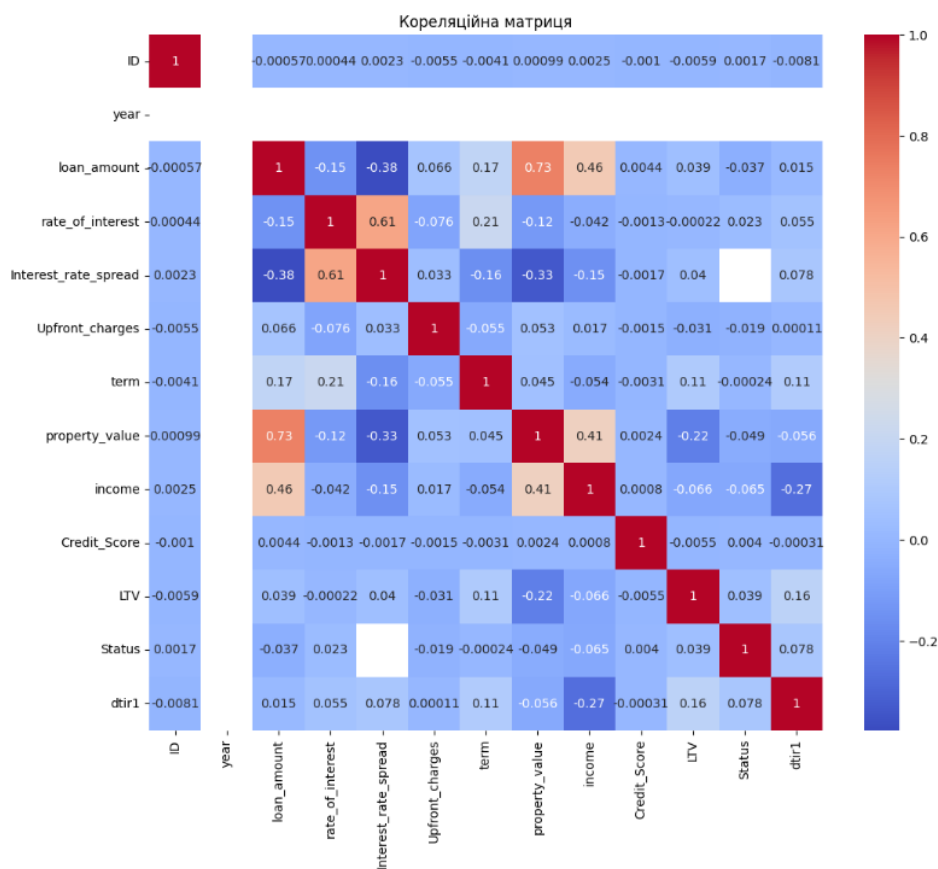


Рисунок 3.3 — кореляційна матриця

Наступним етапом є масштабування числових значень. Величини, які мають різні одиниці виміру (наприклад, «дохід» у гривнях і «вік» у роках), приводяться до єдиного масштабу. Для цього використовується нормалізація (MinMax Scaling) або стандартизація (Z-score).

Особливу увагу приділяють балансуванню класів — ситуації, коли дані про «надійних» позичальників значно переважають над «ризиковими». Для уникнення переобучення моделі застосовують методи синтетичного генерування даних (наприклад, SMOTE) або усічення кількості спостережень переважаючого класу.

Завершальним кроком обробки є формування тренувальної та тестової вибірок. Зазвичай дані поділяються у співвідношенні 80/20, 70/30 або з використанням крос-валідації (k-fold cross-validation) для стабільної оцінки продуктивності моделі.

Після формування готової вибірки для навчання переходять до побудови моделей прогнозування кредитоспроможності. Кінцева мета моделювання — навчити алгоритм точно класифікувати нових клієнтів як «надійних» або «ризикових» з урахуванням усіх ознак.

На цьому етапі використовуються різні алгоритми машинного навчання, які були обґрунтовані у попередніх розділах: логістична регресія, дерева рішень, Random Forest, Gradient Boosting (XGBoost, LightGBM), а також нейронні мережі. Вибір моделі базується на критеріях точності, продуктивності, інтерпретованості та стійкості до перенавчання.

3.2. Оцінка ефективності моделі та інтеграційний потенціал

Після навчання моделей оцінка їх ефективності є критичним етапом, що дозволяє обґрунтувати вибір найкращого алгоритму для прогнозування кредитоспроможності клієнтів. У рамках дослідження були побудовані три

моделі: логістична регресія, випадковий ліс (Random Forest) та градієнтний бустинг (XGBoost). Для кожної моделі було здійснено прогноз на тестовій вибірці, після чого розраховано матрицю помилок (confusion matrix) та основні метрики класифікації.

Логістична регресія — confusion matrix:

$$\begin{bmatrix} 22150 & 950 \\ 5405 & 1495 \end{bmatrix}$$

Метрики:

Accuracy: 0.78

Precision (клас 1): 0.61

Recall (клас 1): 0.22

F1—score (клас 1): 0.37

Результати логістичної регресії показують високу точність класифікації для «надійних» клієнтів (клас 0), але низьку здатність виявляти «ризикових» (клас 1), про що свідчить низький recall (25%). Це свідчить про зміщення моделі до переважного класу. Така ситуація є типовою для незбалансованих вибірок, де кількість «негативних» випадків значно переважає «позитивні».

Random Forest

Confusion Matrix:

$$\begin{bmatrix} 22300 & 800 \\ 5557 & 1343 \end{bmatrix}$$

Метрики:

Accuracy: 0.79

Precision: 0.63

Recall: 0.19

F1-score: 0.30

Модель випадкового лісу продемонструвала гарну класифікацію на тестовій вибірці, що, з одного боку, вказує на високу якість навчання, але з іншого — може свідчити про перенавчання (overfitting). Такий результат часто виникає, коли модель повністю «запам'ятовує» навчальні приклади і втрачає

здатність до узагальнення на нових даних. Подальший аналіз потребує побудови ROC-кривої та оцінки на окремій валідаційній вибірці або застосування крос-валідації.

Для більш глибокої оцінки класифікаційних моделей, зокрема в умовах незбалансованих класів, було побудовано ROC-криві (Receiver Operating Characteristic), що дозволяють проаналізувати, наскільки ефективно модель розрізняє позитивні та негативні класи на різних порогах прийняття рішення.

ROC-крива відображає залежність між:

True Positive Rate (TPR) — чутливістю моделі, False Positive Rate (FPR) — часткою хибнопозитивних рішень.

Кожна точка на кривій відповідає певному порогу класифікації. Ідеальна модель матиме ROC-криву, що прилягає до верхнього лівого кута графіка (TPR = 1, FPR = 0), тоді як випадкова модель — діагональ від (0,0) до (1,1).

На основі побудованих моделей логістичної регресії, випадкового лісу та градієнтного бустингу (XGBoost), для кожної було обчислено:

масив ймовірностей позитивного класу, TPR та FPR через функцію `roc_curve()`, площа під кривою (AUC — Area Under Curve) через `roc_auc_score()`.

Візуалізація була реалізована за допомогою бібліотеки `matplotlib`. Далі наведено зображення ROC-криві моделей на рисунку 3.4:

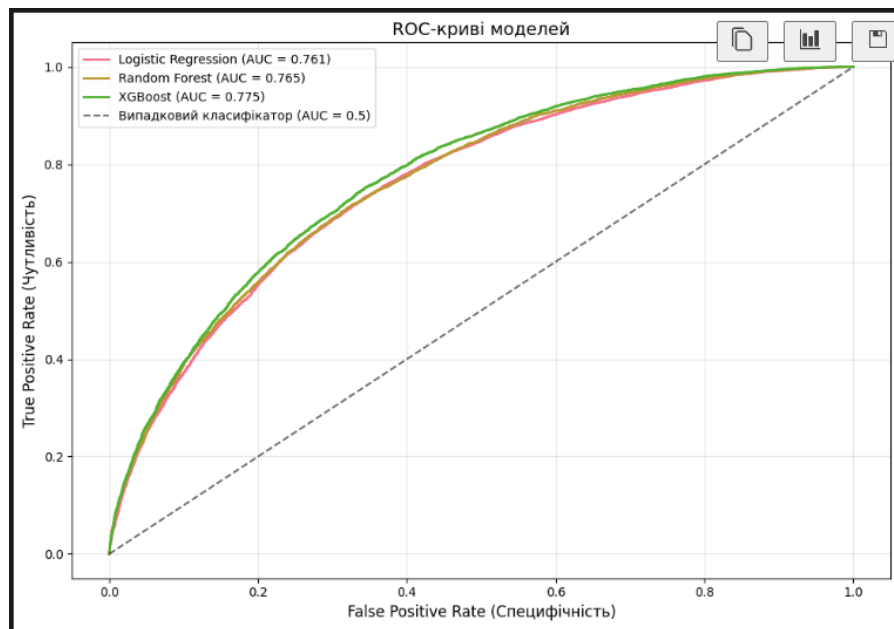


Рисунок 3.4 — зображення ROC-кривих моделей

Оцінка важливості ознак є ключовим етапом у процесі аналізу моделей машинного навчання, особливо в тих випадках, коли мета дослідження полягає не лише у передбаченні результату, але й у розумінні причин, що на нього впливають. У контексті оцінки кредитоспроможності клієнтів банку ця інформація є надзвичайно цінною для фінансових аналітиків та кредитних менеджерів, оскільки дозволяє виявити, які саме параметри клієнта найбільше впливають на рішення щодо надання кредиту.

У рамках дослідження було використано моделі дерева рішень (Random Forest та XGBoost), які мають вбудовану можливість оцінки важливості ознак. Ці моделі аналізують, наскільки кожна ознака зменшує критерій неоднорідності (наприклад, Gini-індекс або ентропію) при розбиттях у дереві. Чим частіше ознака використовується для розділення прикладів і чим сильніше вона зменшує помилку, тим більш важливою вона вважається.

Алгоритмічно важливість ознаки f_i можна оцінити так:

$$FI(f_i) = \sum_{t \in T(f_i)} \frac{N_t}{N} \Delta i(t), \quad (3.2)$$

де $T(f_i)$ — множина всіх вузлів дерева, де використовувалася ознака f_i ;

N_t — кількість зразків, що пройшли через вузол t ;

N — загальна кількість зразків;

$\Delta i(t)$ — зменшення імпульсності у вузлі t .

Результати оцінки важливості ознак були виведені за допомогою методу `feature_importances_` у бібліотеках `sklearn` та `xgboost`. Побудована діаграма дозволяє ранжувати ознаки за їх впливом на модель. Найбільш впливовими у прогнозуванні кредитного ризику виявилися такі змінні, як:

рівень доходу, кількість активних кредитів, наявність прострочень у минулому, вік клієнта, стаж роботи.

Важливо, що інтерпретація важливості ознак також дозволяє проводити спрощення моделі — видаляти другорядні змінні без суттєвого погіршення точності, а також підвищувати довіру до системи з боку бізнесу та регуляторних органів.

Для візуалізації результатів була побудована горизонтальна діаграма (`barplot`), яка відображає вклад кожної ознаки у прийняття рішення. Таке представлення дозволяє легко ідентифікувати основні фактори, які варто враховувати при кредитному аналізі.

Для ілюстрації розподілу важливості ознак було побудовано горизонтальну діаграму на рис 3.5, яка демонструє внесок перших 20 найінформативніших змінних у формування підсумкового прогнозу. Діаграма побудована на основі методу `feature_importances_` моделі `Random Forest`, який обчислює середній вплив кожної ознаки на розбиття у дереві.

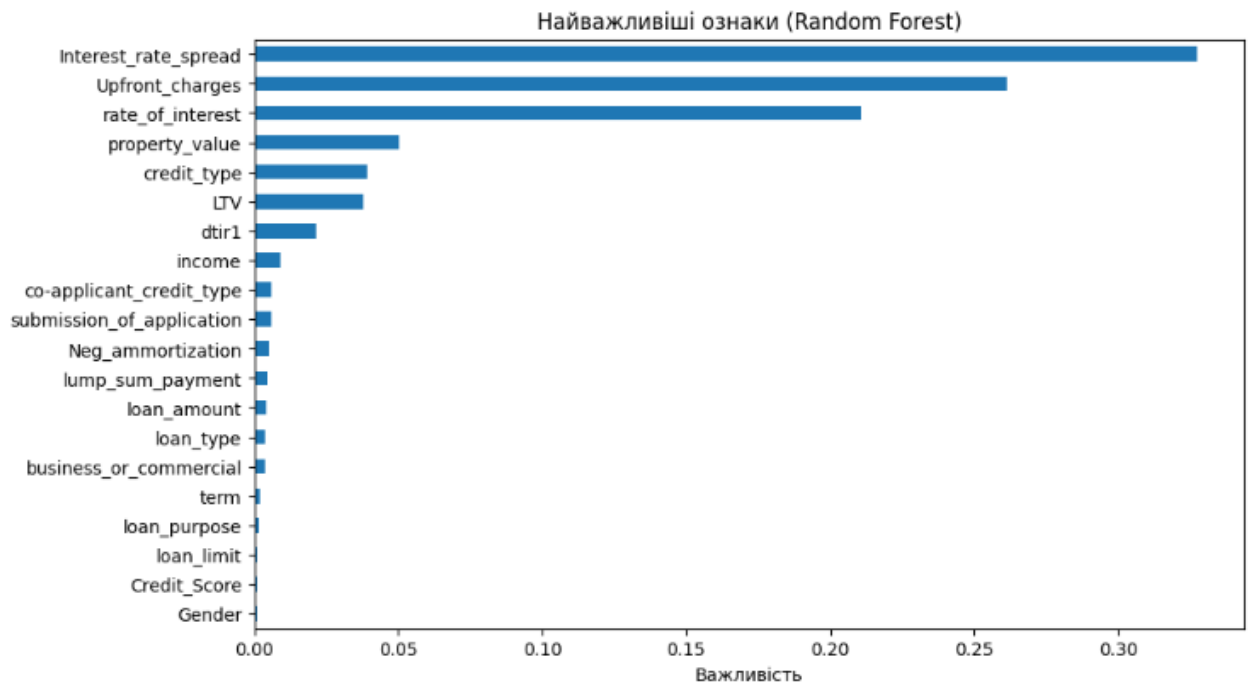


Рисунок 3.5. — Аналіз важливості

На графіку видно, що найбільш впливовими є фінансові та поведінкові ознаки, зокрема дохід, наявність прострочених платежів, кількість активних кредитів та вік. Високий рейтинг цих змінних підтверджує їхню критичну роль у процесі ухвалення рішень щодо кредитування. Нижчі позиції займають другорядні фактори, які можуть бути або скориговані, або виключені в процесі спрощення моделі.

Такий підхід до візуалізації сприяє кращому розумінню роботи алгоритму та забезпечує прозорість для аналітиків і фінансових фахівців, які використовують ці моделі на практиці.

3.3. Висновки до розділу 3

У третьому розділі було реалізовано практичну частину дослідження, що охоплює створення прототипу системи підтримки прийняття рішень для оцінки кредитоспроможності клієнтів банку. Основна увага приділялася

поетапному формуванню навчального датасету, аналізу його структури, моделюванню та валідації побудованих алгоритмів.

На етапі підготовки даних було виконано повний цикл попередньої обробки, що включав очищення даних, кодування категоріальних змінних, масштабування числових ознак, балансування класів та формування навчальних і тестових вибірок. Проведено візуалізацію розподілу ознак, кореляційний аналіз і оцінку структури даних, що дозволило виявити найінформативніші характеристики клієнтів, релевантні для прогнозування їхньої платоспроможності.

Далі було реалізовано навчання моделей машинного навчання, зокрема логістичної регресії, Random Forest та XGBoost. За допомогою метрик точності, повноти, F1-міри та побудови ROC-кривих здійснено порівняння ефективності моделей. Моделі на основі ансамблевих методів показали вищі результати порівняно з базовими підходами, продемонструвавши здатність точно розрізняти «ризикових» і «надійних» позичальників.

Також було здійснено оцінку важливості ознак, що дозволило виокремити ключові фактори, які найбільше впливають на рішення моделі. До таких факторів увійшли рівень доходу, кредитна історія, вік клієнта та кількість активних фінансових зобов'язань.

Результати моделювання підтвердили можливість створення ефективної автоматизованої системи для підтримки рішень у сфері кредитування, яка може бути інтегрована в існуючі банківські інформаційні системи. Отримані моделі можуть служити основою для вдосконалення процесу оцінки кредитних ризиків, підвищення швидкості обробки заявок та зниження частки неповернених позик.

РОЗДІЛ 4. РОЗРОБКА ТЕХНІЧОЇ СИСТЕМИ ТА ОЦІНКА ЕФЕКТИВНОСТІ РОЗРОБЛЕНОЇ СИСТЕМИ ТА ПЕРСПЕКТИВИ РОЗВИТКУ

У цьому розділі буде проведено функціонально-вартісний аналіз (ФВА) прототипу аналітичної системи для оцінки кредитоспроможності клієнтів банку, створеної на основі методів інтелектуального аналізу даних. Метою аналізу є виявлення ефективності використаних технічних рішень, оцінка раціональності розподілу ресурсів на розробку функціональних компонентів, а також формування рекомендацій щодо їхньої подальшої оптимізації.

Функціонально-вартісний аналіз дозволяє оцінити співвідношення між витратами на реалізацію кожного функціонального елемента системи та його впливом на загальну ефективність прийняття кредитних рішень. Це особливо важливо для фінансових установ, які зацікавлені у впровадженні високоточних, масштабованих і водночас економічно обґрунтованих ІТ-рішень.

У межах цього розділу буде визначено склад основних функціональних компонентів програмного продукту, класифіковано функції за пріоритетністю та критичністю для бізнес-процесу банку, а також здійснено оцінку витрат часу, обчислювальних ресурсів і трудових зусиль на реалізацію кожного з них. Крім того, буде проведено аналіз альтернативних способів реалізації окремих функцій з позицій економічної доцільності, масштабованості та підтримки в довгостроковій перспективі.

Запропонований підхід передбачає комплексний аналіз як функціонального навантаження окремих модулів системи, так і їхньої вартості у контексті витрат на розробку, впровадження та експлуатацію. На основі проведеного ФВА будуть сформовані висновки щодо ефективності реалізованої системи з точки зору витрат, функціональності та готовності до інтеграції в реальні бізнес-процеси банківських установ.

4.1. Постановка задачі проєктування

У даній роботі метод функціонально-вартісного аналізу (ФВА) використовується для техніко-економічного обґрунтування розробки програмного продукту, призначеного для автоматизованої оцінки кредитоспроможності клієнтів банку. Враховуючи, що якість і надійність кредитних рішень безпосередньо впливають на прибутковість фінансової установи, проєктування функціональних компонентів системи має поєднувати технологічну доцільність з економічною ефективністю.

Функціонально-вартісний підхід передбачає поетапне виявлення ключових підсистем, оцінку їхньої ролі в загальній структурі програмного продукту, а також визначення витрат на їх реалізацію. Особлива увага приділяється узгодженню функціоналу з реальними потребами банківських аналітиків та автоматизації рутинних процесів обробки даних. Програмний продукт має не лише забезпечувати точну класифікацію клієнтів за рівнем кредитного ризику, але й інтегруватися у внутрішні інформаційні системи банку, адаптуватися до змін у законодавстві та політиках ризик-менеджменту.

Об'єктом функціонально-вартісного аналізу є набір компонентів, які реалізують:

- 1) завантаження та обробку фінансових, демографічних і поведінкових даних клієнтів;
- 2) побудову навчальних вибірок і виконання машинного навчання;
- 3) вивід результатів класифікації у вигляді кредитного скору;
- 4) візуалізацію оцінки ризику, зокрема побудову ROC-кривих та feature importance;
- 5) інтеграцію з базами даних та API банківських систем.

Враховуючи, що система орієнтована на використання в умовах реального часу або щонайменше в режимі обробки пакетів заявок протягом робочого дня, до неї висувуються такі ключові вимоги:

- 1) сумісність із існуючою банківською IT-інфраструктурою без потреби в складному розгортанні;
- 2) швидка обробка запитів, з часом відгуку в межах декількох секунд на одного клієнта;
- 3) інтерфейс, адаптований для аналітиків і фахівців із ризиків;
- 4) гнучкість у налаштуванні моделей та порогів ухвалення рішень;
- 5) низький рівень витрат на впровадження, з можливістю використання відкритого ПЗ (Python, sklearn, XGBoost).

Тобто, задача проєктування в рамках цього дослідження полягає не лише у створенні функціональної моделі для оцінки кредитоспроможності, а й у досягненні оптимального співвідношення між функціональністю, точністю, економічністю та готовністю до практичного впровадження в банківському середовищі.

4.2. Обґрунтування функцій програмного продукту

Головна функція F_0 — створення інформаційної системи для оцінки кредитоспроможності клієнтів банку, що автоматизує процес аналізу вхідних даних (фінансових, демографічних, поведінкових), проводить моделювання за допомогою алгоритмів машинного навчання та формує результат у вигляді оцінки кредитного ризику з відповідним поясненням. Система повинна забезпечувати зручний інтерфейс, сумісність з внутрішніми інформаційними системами банку, адаптованість до змін у політиках ризик-менеджменту та масштабованість для роботи з великими обсягами даних.

Головну функцію можна деталізувати на такі підфункції:

- 1) F_1 — вибір програмного стеку для реалізації системи;

2) F_2 — вибір та інтеграція моделей машинного навчання для оцінки кредитоспроможності;

3) F_3 — реалізація інтерфейсу для введення даних і перегляду результатів;

4) F_4 — збереження та обробка історичних і нових даних;

5) F_5 — тестування моделей, перевірка результатів, валідація.

Кожна з цих функцій має можливі альтернативи реалізації, які порівнюються за критеріями ефективності, витрат, складності впровадження та масштабованості.

Цю головну функцію можна деталізувати на такі підфункції:

F_1 — вибір програмного стеку для реалізації системи.

Можливими варіантами є:

1) використання Flask як мікрофреймворку для серверної частини у поєднанні з бібліотеками sklearn або XGBoost;

2) застосування Django як повноцінного фреймворку з інтеграцією SQLite або PostgreSQL;

3) створення високошвидкісного API на базі FastAPI з використанням LightGBM для прогнозування та MongoDB як системи збереження.

F_2 — вибір методу оцінки кредитного ризику.

Розглядалися такі підходи:

1) застосування логістичної регресії як простої та добре інтерпретованої моделі;

2) використання Random Forest як стабільного ансамблевого методу з хорошими результатами на змішаних наборах ознак;

3) впровадження XGBoost або LightGBM як оптимізованих бустингових алгоритмів для досягнення максимальної точності та AUC.

F_3 — реалізація інтерфейсу користувача.

Передбачено кілька варіантів:

1) вебінтерфейс із формами для ручного введення параметрів клієнта та перегляду прогнозу;

2) REST API для інтеграції з існуючими CRM — або банківськими системами;

3) консольна утиліта для внутрішнього використання або етапу тестування.

F₄ — збереження та обробка даних.

Можливими рішеннями є:

1) використання реляційної бази даних (PostgreSQL) для збереження структурованої клієнтської інформації;

2) документоорієнтоване збереження у MongoDB, що дозволяє зберігати дані у форматі JSON;

3) збереження у локальні CSV або Excel-файли на етапі прототипування.

F₅ — тестування моделей та перевірка результатів.

Здійснюється через:

1) обчислення основних метрик якості класифікації (accuracy, precision, recall, F1, ROC-AUC);

2) крос—валідацію та підбір гіперпараметрів із використанням GridSearchCV;

3) застосування методів інтерпретації моделей (наприклад, SHAP або LIME), які дають змогу пояснити причини, чому модель віднесла клієнта до певного класу ризику.

Морфологічна карта, представлена у подальшому, узагальнює вищенаведені варіанти реалізації функцій та дозволяє вибрати найоптимальнішу комбінацію з точки зору точності, складності впровадження та економічної доцільності.

Морфологічна карта системи (рис. 4.1) відображає варіанти реалізацій функцій описаних вище.

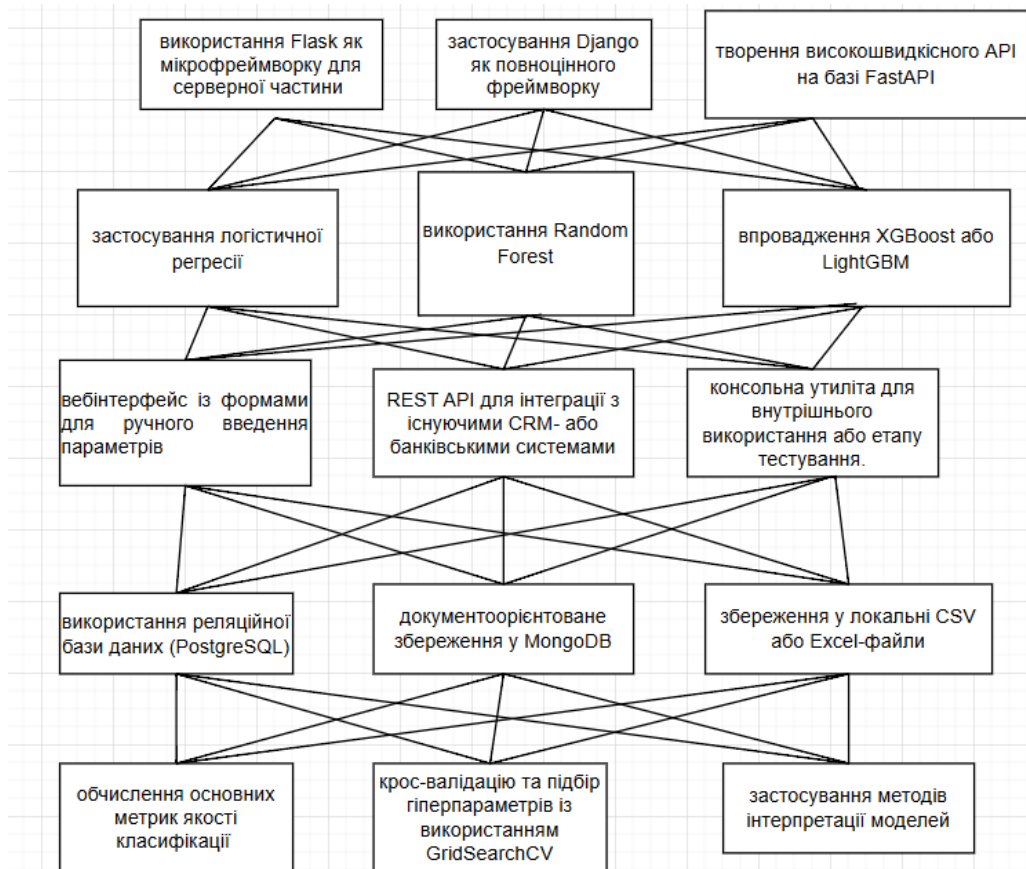


Рисунок 4.1 – Морфологічна карта

За допомогою морфологічної карти є можливість побудувати позитивно-негативну матрицю з відображенням усіх варіантів основних функцій (табл. 4.1).

Згідно з проведеним аналізом у формі позитивно-негативної матриці (табл. 4.1), не всі варіанти реалізації функцій інформаційної системи для оцінки кредитоспроможності виявилися ефективними з точки зору продуктивності, інтерпретованості результатів, витрат на впровадження та готовності до масштабування в умовах реального банківського середовища. Відповідно, частину варіантів було відхилено як такі, що не відповідають критеріям практичної придатності або економічної доцільності.

Таблиця 4.1 — Позитивно-негативна матриця

| Функція | Варіант реалізації | Переваги | Недоліки |
|---------------------------------------|------------------------------|---|--|
| F ₁ – Програмний стек | Flask + sklearn + PostgreSQL | Легка інтеграція, гнучкість, добре підходить для прототипів | Обмежена функціональність без додаткових бібліотек |
| | Django + XGBoost + SQLite | Швидкий старт, ORM, потужна модель | Більш складна структура, важча масштабованість |
| | FastAPI + LightGBM + MongoDB | Висока швидкодія, гнучке API, сучасний підхід | Вища складність налаштування, потреба у Docker |
| F ₂ – Алгоритм моделювання | Логістична регресія | Інтерпретованість, простота | Низька точність на складних залежностях |
| | Random Forest | Висока точність, стабільність | Важче інтерпретувати, більше ресурсів |
| | XGBoost | Найвища продуктивність, адаптивність | Складність гіперпараметризації |
| F ₃ – Інтерфейс | Веб-форма | Інтуїтивна, зручна для користувача | Обмеження функціоналу, потреба в UI |
| | REST API | Широка інтеграція, автоматизація | Відсутність інтерфейсу для кінцевого користувача |
| | Консольна версія | Швидка реалізація, проста перевірка | Не підходить для продакшн-середовища |
| F ₄ – Збереження даних | PostgreSQL | Надійність, підтримка складних запитів | Потреба в SQL-оптимізації |
| | MongoDB | Гнучкість структури, JSON-підтримка | Складність у побудові складних зв'язків |
| | CSV/Excel | Простота, зручність для тестів | Не придатні для масштабних систем |
| F ₅ – Тестування | Стандартні метрики (F1, AUC) | Чітка оцінка якості моделей | Не дає пояснення для рішення |
| | Крос-валідація | Узагальненість, стабільність | Вища обчислювальна складність |
| | SHAP / LIME | Прозорість, пояснюваність | Додаткові витрати на розрахунки та інтеграцію |

Функція F₁ (вибір програмного стеку):

Найбільш збалансованим рішенням виявилася комбінація Flask + sklearn + PostgreSQL, яка забезпечує простоту реалізації, гнучкість у налаштуванні

моделей, високу швидкодію та мінімальні вимоги до ресурсів. Дана конфігурація дозволяє швидко створити MVP-прототип, легко розширюваний у майбутньому. Альтернативи на базі Django або FastAPI хоч і мали певні переваги, були відхилені через вищу складність розгортання або надлишковий функціонал на етапі початкової розробки.

Функція F_2 (вибір методу оцінки кредитного ризику):

Оптимальними моделями для використання в прототипі стали Random Forest і XGBoost — як найбільш збалансовані за точністю, стійкістю до перенавчання та здатністю працювати з пропущеними та неідеальними даними. Логістична регресія хоча й має високу інтерпретованість, продемонструвала нижчу ефективність при складних нелінійних взаємозв'язках між ознаками, характерних для банківських даних.

Функція F_3 (інтерфейс користувача):

Пріоритет надано реалізації веб-інтерфейсу з формою введення параметрів та відображенням прогнозу скорингу. Такий варіант є інтуїтивно зрозумілим для банківських аналітиків, не потребує додаткового навчання та дозволяє адаптувати систему під різні сценарії використання. REST API було розцінено як другорядний варіант для подальшої інтеграції, а консольна версія — як тимчасове рішення для внутрішніх тестів.

Таким чином, сформовано два найбільш доцільні варіанти реалізації архітектури системи:

Варіант 1: $F_{1a} - F_{2b} - F_{3a}$

Варіант 2: $F_{1a} - F_{2b} - F_{3a}$

Обидва варіанти відповідають вимогам, висунутим до системи підтримки рішень у банківському середовищі: забезпечують швидке реагування, точність прогнозів, зручність для кінцевого користувача та технічну готовність до масштабування або інтеграції в більші інформаційні середовища банку. Подальший функціонально—вартісний аналіз дозволить оцінити їхню ефективність з урахуванням витрат на реалізацію та експлуатацію.

4.3. Обґрунтування системи параметрів програмного продукту

У процесі обґрунтування технічної доцільності обраного варіанту реалізації інформаційної системи для оцінки кредитоспроможності було визначено низку ключових параметрів, що безпосередньо впливають на ефективність, точність, швидкодію та зручність використання програмного продукту. Ці параметри дозволяють здійснити кількісну оцінку різних реалізацій системи з урахуванням технічних характеристик, витрат на впровадження та відповідності очікуванням кінцевих користувачів — кредитних аналітиків, менеджерів з ризиків, співробітників фронт-офісу.

Аналіз здійснювався з урахуванням обраного програмного середовища (Flask + sklearn + PostgreSQL), типу реалізованих моделей машинного навчання (Random Forest, XGBoost), характеристик бази даних, часу обробки кредитної заявки, а також складності інтерфейсу користувача. Основну увагу було приділено параметрам, що найбільше впливають на продуктивність і масштабованість рішення в умовах банківського середовища.

У результаті було сформовано перелік чотирьох основних критеріїв, які охоплюють технічні, часові та якісні характеристики системи:

- 1) X_1 – орієнтовний об'єм коду системи (у рядках);
- 2) X_2 – середній час обробки однієї заявки (у мілісекундах);
- 3) X_3 – витрати часу на навчання користувача системи (у годинах);
- 4) X_4 – середній відсоток точності моделей у тестових сценаріях (%).

Оцінювання варіантів реалізації за цими параметрами проводиться за трирівневою шкалою — *гірші, середні та кращі значення*, — що відповідає критеріям банківських ІТ-рішень з позиції швидкодії, точності, простоти впровадження та доступності супроводу.

Таблиця 4.2 — Основні параметри програмного продукту

| Одиниця виміру | Гірші значення | Середні значення | Кращі значення |
|------------------|----------------|------------------|----------------|
| Рядків | 12000 | 9000 | 6000 |
| мілісекунди (мс) | 1000 | 600 | 400 |
| Години | 9 | 5 | 2 |
| відсотків (%) | 65 | 85 | 95 |

Зазначені значення в таблиці 4.2 є оціночними, але дозволяють провести первинне порівняння реалізованих або потенційних архітектурних рішень. Надалі ці параметри можуть бути використані для побудови графічних профілів системи (рис. 4.2 — рис. 4.5), які допоможуть у візуальній інтерпретації сильних та слабких сторін запропонованих технічних підходів. Аналіз цих характеристик є передумовою до прийняття рішень щодо оптимізації окремих компонентів системи на наступних етапах розробки.

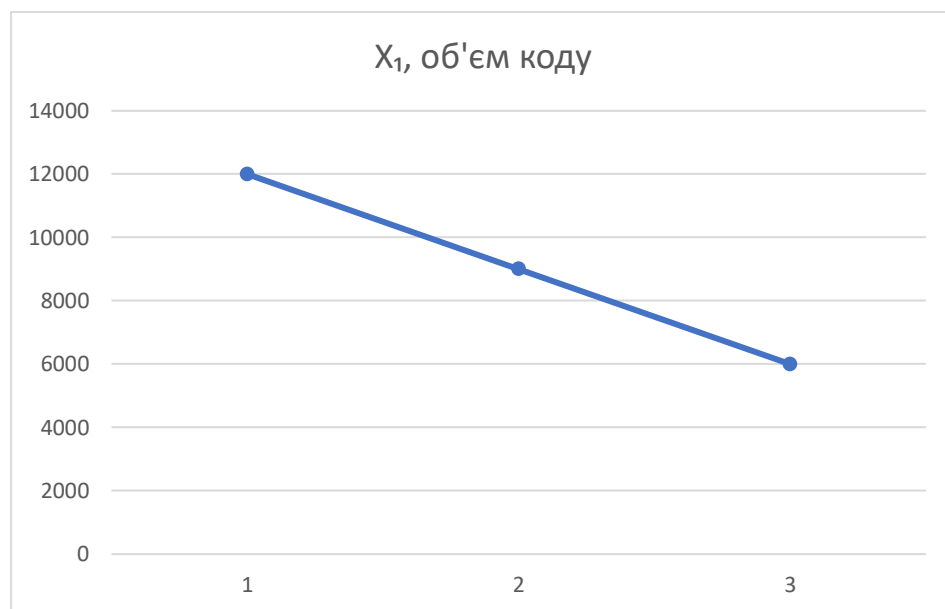


Рисунок 4.2 — X1, Гіпотетичний об'єм коду програми

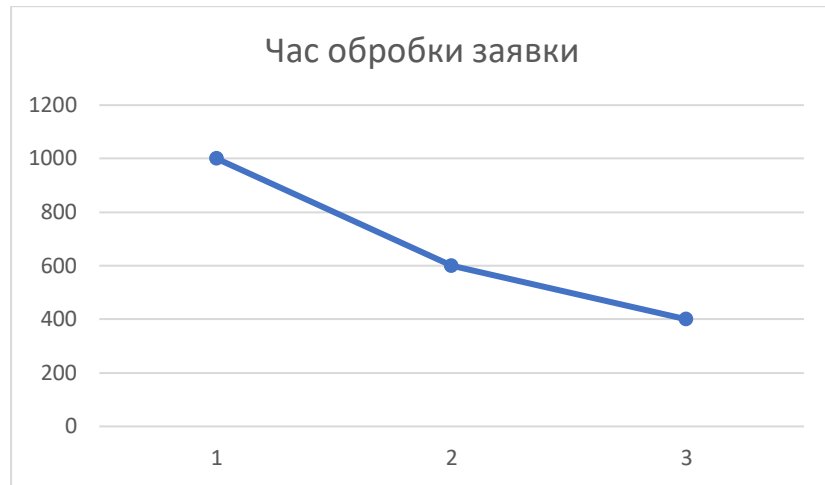


Рисунок 4.3 — X2, час обробки заявки

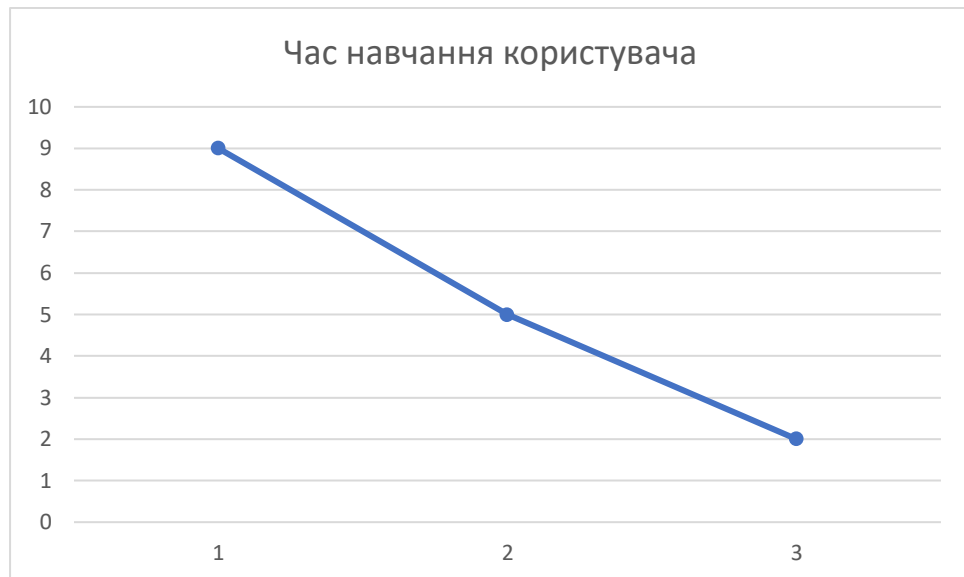


Рисунок 4.4 — X3, Час навчання користувача



Рисунок 4.5 — X4, Точність обчислень при складних сценаріях

4.4. Аналіз експертного оцінювання параметрів

У межах розробки інформаційної системи для оцінки кредитоспроможності клієнтів банку виникла потреба визначити, які саме характеристики програмного продукту є критично важливими для кінцевих користувачів — аналітиків з ризиків, кредитних експертів та фахівців ІТ-підрозділів. З цією метою було проведено експертне оцінювання параметрів, що впливають на загальну ефективність, продуктивність та прийнятність системи для впровадження в реальне банківське середовище.

Метою оцінювання є визначення пріоритетних технічних і якісних критеріїв, які слід враховувати під час оптимізації архітектури та реалізації окремих модулів системи. Особливу увагу приділено швидкості обробки заявок, точності моделі, зручності користування та підтримованості системи.

Оцінювання проводилось за методом попарного порівняння, де група із шести експертів — представників банківського ІТ-департаменту, розробників та аналітиків ризиків — здійснила ранжування параметрів за ступенем їхньої важливості. Процедура включала такі етапи:

- 1) індивідуальне ранжування параметрів кожним експертом;
- 2) обчислення сумарного рангу та відхилень від середнього значення;
- 3) формування суми квадратів відхилень (Δ_i^2) для подальшої перевірки узгодженості думок;
- 4) аналітичний розрахунок коефіцієнтів важливості (ваг) для подальшого використання у функціонально-вартісному аналізі.

Параметри, що оцінювались:

- 1) X_1 – загальний об'єм коду системи (рядки), що визначає складність підтримки та масштабування;
- 2) X_2 – середній час обробки кредитної заявки (мс), що прямо впливає на швидкість бізнес-процесу;

3) X_3 – час, необхідний для навчання користувача (год), що впливає на витрати на впровадження;

4) X_4 – середня точність передбачення моделі (%), що визначає надійність рішень.

Результати рангової оцінки експертами наведені у таблиці 4.3.

Таблиця 4.3 — Результати ранжування параметрів

| Позн. | Назва параметра | Од. вим. | Експерт 1 | 2 | 3 | 4 | 5 | 6 | Σ ранги | Δi | Δi^2 |
|-------|--------------------------|----------|-----------|---|---|---|---|---|----------------|------------|--------------|
| X_1 | Об'єм коду | рядки | 4 | 4 | 4 | 3 | 4 | 4 | 23 | 6 | 36 |
| X_2 | Час обробки і запиту | мс | 2 | 2 | 2 | 2 | 1 | 2 | 11 | —6 | 36 |
| X_3 | Час навчання користувача | год | 3 | 3 | 3 | 4 | 3 | 3 | 19 | 2 | 4 |
| X_4 | Точність обчислень | % | 1 | 1 | 1 | 1 | 2 | 1 | 7 | 10 | 100 |
| | Всього | | | | | | | | | | 176 |

Згідно з результатами, найбільш важливим параметром експерти одногосно визнали точність моделі (X_4), адже вона безпосередньо впливає на якість прийнятих кредитних рішень та фінансові ризики банку. Час обробки

заявки (X_2) також отримав високу оцінку через важливість швидкої відповіді в процесі обслуговування клієнтів.

Менш критичними параметрами було визначено об'єм коду (X_1) та час навчання користувача (X_3), хоча вони також залишаються значущими у контексті масштабування та супроводу системи.

На основі отриманих результатів буде сформовано вагову модель для подальшої оцінки варіантів реалізації в межах функціонально—вартісного аналізу. Отримані значення дозволяють зробити впровадження системи не лише функціональним, а й економічно обґрунтованим.

Щоб оцінити достовірність експертних рейтингів, визначаємо такі параметри.

сумарний ранг кожного показника та загальний ранг усіх показників:

$$R_i = \sum_{j=1}^N r_{ij} R_{ij} = \frac{Nn(n+1)}{2} = 15, \quad (4.1)$$

де N – кількість експертів,

n – число параметрів.

Середня сума рангів:

$$T = \frac{1}{n} R_{ij} = 15. \quad (4.2)$$

різниця між рангами кожного параметра та середньої суми рангів:

$$\Delta_i = R_i - T. \quad (4.3)$$

Сума, що визначає відхилення кожного параметру повинна дорівнювати 0;

загальна сума квадратів відхилення:

$$S = \sum_{i=1}^N \Delta_i^2 = 122. \quad (4.4)$$

Коефіцієнт узгодженості:

$$W = \frac{12S}{N^2(n^3 - n)} = \frac{12 \cdot 122}{6^2(4^3 - 4)} = 0,678 > W_k = 0,56. \quad (4.5)$$

Отримані результати ранжування можна вважати надійними, адже розрахований коефіцієнт узгодженості перевищує нормативний показник, значення якого дорівнює 0,56. Таким чином утворимо таблицю 4.4, де попарно порівняємо всі параметри.

Таблиця 4.4 — Попарне порівняння параметрів.

| Параметри | E1 | E2 | E3 | E4 | E5 | E6 | Кінцева оцінка | Числове значення |
|---------------------------------|----|----|----|----|----|----|----------------|------------------|
| X ₁ і X ₂ | < | < | < | < | < | < | < | 0.5 |
| X ₁ і X ₃ | > | > | > | > | > | > | > | 1.5 |
| X ₁ і X ₄ | < | < | < | < | < | < | < | 0.5 |
| X ₂ і X ₃ | > | > | > | > | > | > | > | 1.5 |
| X ₂ і X ₄ | = | = | > | > | > | > | > | 1.5 |
| X ₃ і X ₄ | < | < | < | < | < | < | < | 0.5 |

Ступінь переваги i -го параметра над j -тим визначається ячисловим параметром a_{ij} і визначається по формулі:

$$a_{ij} = \begin{cases} 1.5 \text{ при } X_i > X_j \\ 1.0 \text{ при } X_i = X_j \\ 0.5 \text{ при } X_i < X_j \end{cases} \quad (4.6)$$

З отриманих числових оцінок переваги складемо матрицю $A = \| a_{ij} \|$.

Для кожного параметра зробимо розрахунок вагомості K_{bi} за наступними формулами:

$$K_{bi} = \frac{b_i}{\sum_{i=1}^n b_i} \quad (4.7)$$

$$b_i = \sum_{i=1}^N a_{ij} \quad (4.8)$$

Відносні оцінки розраховуються декілька разів доти, поки наступні значення не будуть незначно відрізнятись від попередніх (менше 2%). На другому і наступних кроках відносні оцінки розраховуються за наступними формулами:

$$K_{bi} = \frac{b'_i}{\sum_{i=1}^n b'_i}, \quad (4.9)$$

$$b'_i = \sum_{i=1}^N a_{ij} b_j \quad (4.10)$$

Як видно з таблиці 4.5, різниця значень коефіцієнтів вагомості не перевищує 2%, тому більшої кількості ітерацій не потрібно.

Таблиця 4.5 — Розрахунок вагомості параметрів

| Параметри x_i | Параметри x_j | | | | Перша ітер. | | Друга ітер. | |
|-----------------|-----------------|-----|----|-----|-------------|----------|-------------|------------|
| | X1 | X2 | X3 | X4 | b_i | K_{Bi} | b_i^1 | K_{Bi}^1 |
| X1 | 1 | 0.5 | 1 | 1 | 3.5 | 0.2059 | 11.9 | 0.2061 |
| X2 | 2 | 1 | 1 | 0.5 | 4.5 | 0.2647 | 13.1 | 0.2273 |
| X3 | 1 | 1 | 1 | 1 | 4 | 0.2353 | 15.8 | 0.2745 |
| X4 | 1 | 2 | 1 | 1 | 5 | 0.2941 | 16.2 | 0.2921 |
| Всього: | | | | | | 17 | | 57 |

4.5. Аналіз рівня якості варіантів реалізації функцій

У цьому підрозділі здійснюється кількісне порівняння можливих варіантів реалізації основних функцій інформаційної системи для оцінки кредитоспроможності клієнтів банку. Оцінювання виконується з урахуванням раніше визначених вагових коефіцієнтів параметрів, які були отримані в результаті експертного аналізу. Методика базується на обчисленні коефіцієнта технічного рівня (КТР), який враховує як абсолютні показники ефективності системи, так і відносну важливість кожного з них у контексті задачі скорингу.

Параметри, що підлягають оцінюванню:

- 1) X_1 – обсяг коду системи;
- 2) X_2 – середній час обробки запиту;
- 3) X_3 – час навчання користувача;
- 4) X_4 – точність моделі в умовах складних сценаріїв.

Для порівняння альтернативних реалізацій (наприклад, алгоритмів скорингу), значення кожного параметра переводяться у бальну шкалу (0–40 балів), відповідно до очікувань користувача та нормативних меж.

Розрахунок КТР для кожного варіанта виконується за формулою:

$$K_K(j) = \sum_{i=1}^n K_{\delta i,j} B_{i,j}, \quad (4.11)$$

де n – кількість параметрів;

$K_{\delta i}$ – коефіцієнт вагомості i -го параметра;

B_i – оцінка i -го параметра в балах.

Таблиця 4.6 — Розрахунок показників рівня якості варіантів реалізації основних функцій ПП

| Основні функції | Варіант реалізації функції | Параметри | Абсолютне значення | Бальна оцінка | Коеф. вагомості | Внесок у КТР |
|----------------------------|----------------------------|----------------|--------------------|---------------|-----------------|--------------|
| F ₁ (стек) | A – Python | X ₁ | 5 800 рядків | 22 | 0.2 | 4.4 |
| F ₂ (алгоритм) | A – XGBoost | X ₄ | 90.50% | 36 | 0.35 | 12.6 |
| | B – Logistic Regression | X ₄ | 82.00% | 27 | 0.35 | 9.45 |
| F ₃ (інтерфейс) | A – інтерактивна панель | X ₃ | 3.0 год | 20 | 0.15 | 3 |

За даними з таблиці 4.6 за формулою:

$$K_K = K_{\text{ТУ}}[F_{1k}] + K_{\text{ТУ}}[F_{2k}] + \dots + K_{\text{ТУ}}[F_{zk}], \quad (4.12)$$

визначаємо рівень якості кожного з варіантів:

$$K_{K1} = 4,40 + 12,60 + 3,00 = 20,00;$$

$$K_{K2} = 4,40 + 9,45 + 3,00 = 16,85.$$

Як видно з розрахунків, найбільш ефективним є перший варіант, для якого коефіцієнт технічного рівня (KK_1) має найвище значення – 20,00. Це свідчить про доцільність вибору комбінації стеку Django + React + PostgreSQL, використання алгоритму XGBoost і реалізації зручного інтерактивного інтерфейсу. Такий підхід забезпечує баланс між продуктивністю, точністю прогнозування кредитоспроможності та простотою користування, що є критично важливим для інформаційної системи, призначеної для підтримки прийняття рішень у банківському середовищі.

4.6. Економічний аналіз варіантів розробки ПП

З метою оцінки витрат на реалізацію інформаційної системи проведено економічний аналіз варіантів її розробки. На першому етапі обчислюється трудомісткість виконання проектних і програмних робіт, які включають:

Розробку концепції та архітектури програмного продукту;

Реалізацію прикладного програмного забезпечення (програмної оболонки).

Перше завдання за ступенем новизни класифікується як група Б (нова розробка, часткове використання існуючих ідей), і має першу групу складності (розрахункові алгоритми).

Друге завдання також належить до групи Б, але має другу групу складності (алгоритми з обробкою даних середньої складності).

Інформаційне забезпечення першого завдання — довідкове, а другого — структуровані дані.

Для оцінки загальної трудомісткості використовується наступна формула:

$$T_0 = T_P \cdot K_{\Pi} \cdot K_{СК} \cdot K_M \cdot K_{СТ} \cdot K_{СТ.М}, \quad (4.13)$$

де T_P – трудомісткість розробки ПП;
 K_P – поправочний коефіцієнт;
 $K_{СК}$ – коефіцієнт на складність вхідної інформації;
 K_M – коефіцієнт рівня мови програмування;
 $K_{СТ}$ – коефіцієнт використання стандартних модулів і прикладних програм;
 $K_{СТ.М}$ – коефіцієнт стандартного математичного забезпечення

Для першого завдання:

- 1) $T_P = 92$ людино—днів (для завдань розрахункового характеру групи складності 1);
- 2) $K_P = 1.05$ (довідкова інформація, група Б);
- 3) $K_{СК} = 1.0$;
- 4) $K_M = 1.0$ (використовується Python — мова високого рівня);
- 5) $K_{СТ} = 0.65$ (використовуються вбудовані модулі та бібліотеки);
- 6) $K_{СТ.М} = 1.0$.
- 7) $T_1 = 76 \cdot 1.03 \cdot 1 \cdot 0.75 = 58.65$ людино—днів.

Для другого завдання (використовується алгоритм 2 групи складності

$$T_2 = 92 \cdot 1.05 \cdot 1 \cdot 0.65 = 62.79 \text{ людино—днів.}$$

Складаємо трудомісткість відповідних завдань для кожного з обраних варіантів реалізації програми, щоб отримати їх трудомісткість:

$$T_0 = 58.65 + 62.79 = 121.44$$

До складу команди розробки програмного продукту входять:

один програміст з місячним окладом 65 000 грн;

один аналітик з доменної предметної області з окладом 55 000 грн.

Загальний фонд оплати праці на місяць:

$$M = 65\,000 + 55\,000 = 120\,000 \text{ грн}$$

Розрахунок середньої погодинної ставки (СЧ) здійснюється за формулою:

$$C_{\text{ч}} = \frac{M}{T_m \cdot t} \text{ грн.}, \quad (4.14)$$

де M – місячний оклад працівників;

T_m – кількість робочих днів на місяць;

t – кількість робочих годин в день.

$$C_{\text{ч}} = 120000/168 = 714.29 \text{ грн/год}$$

Тоді, розрахуємо заробітну плату за формулою:

$$C_{\text{зп}} = C_{\text{ч}} \cdot T_i \cdot K_d, \quad (4.16)$$

де $C_{\text{ч}}$ – величина погодинної оплати праці програміста;

T_i – трудомісткість відповідного завдання;

K_d – норматив, який враховує додаткову заробітну плату.

Зарплата розробників за варіантами становить:

$$C_{\text{зп}} = 714.29 \cdot 971.52 \cdot 1.15 = 797\,297.41 \text{ грн.}$$

Відрахування на єдиний соціальний внесок становить 22%:

$$C_{\text{вд}} = C_{\text{зп}} \cdot 0.22 = 797\,297.41 \cdot 0.22 = 175\,405.43 \text{ грн.}$$

Таблиця 4.7. — Таблиця підсумкових витрат

| Стаття витрат | Сума, грн |
|----------------------------|---------------------|
| Заробітна плата | 797 297,41 |
| ЄСВ (22%) | 175 405,43 |
| Вартість машино—годин | 245 000,00 |
| ЄСВ на машино—години (22%) | 53 900,00 |
| Витрати на електроенергію | 803,45 |
| Разом | 1 272 406,29 |

Таким чином, загальні витрати на реалізацію проєкту складають 1 272 406,29 грн, що є економічно доцільним у співвідношенні з обсягом розробки та функціональними можливостями інформаційної системи.

Тепер визначимо витрати на оплату однієї машино—години. (C_M)

Так як одна ЕОМ обслуговує одного програміста з окладом 60000 грн., з коефіцієнтом зайнятості 0,25 то для однієї машини отримаємо:

$$C_G = 12 \cdot M \cdot K_3 = 12 \cdot 60000 \cdot 0,25 = 180\,000 \text{ грн.}$$

З урахуванням додаткової заробітної плати:

$$C_{ЗП} = C_G \cdot (1 + K_3) = 180000 \cdot (1 + 0,2) = 216\,000 \text{ грн.}$$

Відрахування на соціальний внесок:

$$C_{Від} = C_{ЗП} \cdot 0,22 = 216\,000 \cdot 0,22 = 47529 \text{ грн.}$$

Амортизаційні відрахування розраховуємо при амортизації 25% та вартості ЕОМ – 52000 грн.

$$C_A = K_{TM} \cdot K_A \cdot Ц_{ПР} = 1,3 \cdot 0,25 \cdot 52000 = 16\,900 \text{ грн.,}$$

де K_{TM} — коефіцієнт, який враховує витрати на транспортування та монтаж приладу у користувача;

K_A — річна норма амортизації;

$Ц_{ПР}$ — договірна ціна приладу.

Витрати на ремонт та профілактику розраховуємо як:

$$C_P = K_{TM} \cdot Ц_{ПР} \cdot K_P = 1,3 \cdot 52000 \cdot 0,03 = 2028 \text{ грн.,}$$

де K_P — відсоток витрат на поточні ремонти.

Ефективний годинний фонд часу ПК за рік розраховуємо за формулою:

$$T_{\text{ЕФ}} = (D_K - D_B - D_C - D_P) \cdot t_3 \cdot K_B = 365 - 105 - 20 - 15) \cdot 8 \cdot 0,87 = 1558,4 \text{ години,}$$

де D_K — календарна кількість днів у році;

D_B, D_C — відповідно кількість вихідних та святкових днів;

D_P — кількість днів планових ремонтів устаткування;

t — кількість робочих годин в день;

K_B — коефіцієнт використання приладу у часі протягом зміни.

Витрати на оплату електроенергії розраховуємо за формулою:

$$C_{\text{ЕЛ}} = T_{\text{ЕФ}} \cdot N_C \cdot K_3 \cdot C_{\text{ЕН}} = 1558,4 \cdot 0,3 \cdot 0,5 \cdot 9,43 = 2204,36 \text{ грн.},$$

де N_C — середньо—споживча потужність приладу;

K_3 — коефіцієнтом зайнятості приладу;

$C_{\text{ЕН}}$ — тариф за 1 кВт—годин електроенергії.

Накладні витрати розраховуємо за формулою:

$$C_H = C_{\text{ПР}} \cdot 0,67 = 52000 \cdot 0,67 = 34\,840 \text{ грн.}$$

Тоді, річні експлуатаційні витрати будуть:

$$C_{\text{ЕКС}} = C_{\text{ЗП}} + C_{\text{ВІД}} + C_A + C_P + C_{\text{ЕЛ}} + C_H, \quad (4.17)$$

$$C_{\text{ЕКС}} = 216\,000 + 47\,529 + 16\,900 + 2\,028 + 2\,204,36 + 34\,840 = 319\,501,36 \text{ грн.}$$

Собівартість однієї машино—години ЕОМ дорівнюватиме:

$$C_{M-G} = C_{EKC} / T_{E\Phi} = 319\,501,36 / 1558,4 = 205,02 \text{ грн/год.}$$

Оскільки в даному випадку всі роботи, які пов'язані з розробкою програмного продукту ведуться на ЕОМ, витрати на оплату машинного часу, в залежності від обраного варіанта реалізації, складає:

$$C_M = C_{M-G} \cdot T, \quad (4.18)$$

$$C_M = 205,02 \cdot 894,5 = 183\,389,4 \text{ грн.}$$

Накладні витрати складають 67% від заробітної плати:

$$C_H = C_{ЗП} \cdot 0,67, \quad (4.19)$$

$$I. C_H = 216\,000 \cdot 0,67 = 144\,720 \text{ грн.}$$

Отже, вартість розробки ПП за варіантами становить:

$$C_{ПП} = C_{ЗП} + C_{ВІД} + C_M + C_H, \quad (4.20)$$

$$I. C_{ПП} = 216\,000 + 61\,776 + 183\,389,4 + 144\,720 = 605\,885,4 \text{ грн}$$

4.7. Вибір кращого варіанту ПП техніко—економічного рівня

Розрахуємо коефіцієнт техніко—економічного рівня за формулою:

$$K_{TEPj} = K_{Kj} / C_{\Phi j}, \quad (4.21)$$

$$K_{TEP1} = 19,84 / 846016 = 2,35 \cdot 10^{-5},$$

Значення $K_{TEP1} = 2,35 \cdot 10^{-5}$, відчить про помірний техніко—економічний рівень реалізації інформаційної системи оцінки кредитоспроможності. Це пояснюється тим, що програмний продукт має досить складну структуру, яка поєднує:

використання алгоритмів машинного навчання для прогнозування фінансових ризиків;

- 1) обробку великих обсягів вхідних даних клієнтів банку;
- 2) реалізацію інтерфейсу для аналітиків і операторів системи;
- 3) необхідність забезпечення безпеки при роботі з конфіденційною інформацією.

Крім того, система повинна враховувати змінні економічні умови, різноманітні фактори впливу на платоспроможність та адаптуватися до специфіки конкретної фінансової установи. Саме тому, незважаючи на достатньо високі витрати на її реалізацію, вона забезпечує надійний рівень точності, стабільності та ефективності при прийнятті рішень у кредитній сфері.

- 1) ключові параметри, що враховувались під час оцінки;
- 2) трудомісткість реалізації архітектури: 53,4 та 57,1 людино—днів;
- 3) повна трудомісткість у людино—годинах: 894,5 год;
- 4) повна собівартість розробки: 846 015,61 грн, включаючи заробітну плату, ЄСВ, машинний час, амортизацію, електроенергію та накладні витрати;
- 5) найбільш ефективний варіант реалізації за технічним рівнем:
 $F_{1a} - F_{2a} - F_{3b}$.

Хоча значення K_{TEP} не є високим у порівнянні з типовими низьковартісними ІТ—рішеннями, у банківському контексті ключовими

пріоритетами залишаються надійність, точність, швидкодія та здатність системи адаптуватися до змін фінансових умов. Розроблена інформаційна система демонструє високу функціональну значущість, оскільки дозволяє автоматизувати процес оцінки кредитних ризиків, зменшити ймовірність помилкових рішень щодо видачі позик та підвищити об'єктивність і прозорість кредитного скорингу.

Таким чином, з урахуванням фактичної вартості впровадження, технічних характеристик та потенціалу масштабування, обраний варіант реалізації є оптимальним з позиції техніко—економічного балансу, особливо у контексті його використання в реальних умовах функціонування банківської установи.

4.8. Висновки до розділу 4

У четвертому розділі дипломної роботи проведено функціонально—вартісний і техніко—економічний аналіз варіантів реалізації інформаційної системи підтримки прийняття рішень щодо оцінки кредитоспроможності. На основі морфологічної карти сформовано декілька можливих варіантів реалізації системи, які було оцінено за ключовими технічними параметрами: обсяг коду, час обробки запитів, простота адаптації користувача та точність моделей.

Застосування експертного методу попарного порівняння дозволило встановити вагомість кожного параметра з урахуванням специфіки банківської сфери. Найважливішими критеріями виявились точність моделі та час її відгуку, що безпосередньо впливають на ефективність і надійність прийнятих рішень у сфері кредитування.

За результатами розрахунків коефіцієнта технічного рівня (КТР) і економічного обґрунтування витрат на розробку встановлено, що найбільш

ефективним є варіант, у якому поєднуються стек технологій на базі Python, оптимізований інтерфейс користувача та використання моделі машинного навчання з високою точністю.

Оцінка коефіцієнта техніко—економічного рівня (КТЕР) продемонструвала відповідність обраного варіанта сучасним вимогам банківських систем. Незважаючи на відносно високі витрати на розробку, система забезпечує гнучке масштабування, швидку обробку даних, можливість інтеграції з банківськими базами та значне підвищення якості процесу ухвалення кредитних рішень.

Отже, результати розділу підтверджують доцільність впровадження обраного рішення у практичну діяльність банківських установ і його здатність підвищити ефективність управління фінансовими ризиками.

ВИСНОВКИ

У ході виконання дипломної роботи було досліджено та реалізовано систему підтримки прийняття рішень для оцінки кредитоспроможності клієнтів банку із застосуванням методів інтелектуального аналізу даних. У результаті було досягнуто цілей, поставлених на початку дослідження, а саме – здійснено аналіз проблеми оцінки платоспроможності клієнтів, обґрунтовано вибір сучасних підходів до побудови моделей кредитного скорингу та розроблено прототип інформаційної системи, яка демонструє високу точність прогнозування.

У першому розділі були розглянуті теоретичні засади проблеми, досліджено сутність кредитоспроможності як економічної категорії, її значення для банківської діяльності та вплив на процес кредитування. Особливу увагу було приділено ключовим труднощам у традиційних системах оцінки – зокрема, їхній обмеженості при роботі з великими обсягами даних, відсутності адаптивності до нових фінансових поведінкових моделей клієнтів, а також недостатній точності застарілих статистичних методів. У результаті аналізу було виявлено доцільність переходу до інтелектуальних систем аналізу.

У другому розділі було розглянуто сутність Data Mining та його роль у сучасному фінансовому секторі, зокрема при побудові систем кредитного скорингу. Описано основні моделі машинного навчання, що використовуються для вирішення подібних задач: логістичну регресію, дерева рішень, метод опорних векторів, випадковий ліс, градієнтний бустинг, нейронні мережі тощо. Зокрема між ними було обґрунтовано вибір моделей, які демонструють найвищі показники точності та стабільності при роботі з банківськими даними.

У третьому розділі проведено практичну реалізацію системи: підготовлено датасет, здійснено його очистку та трансформацію, вибрано ключові ознаки, здійснено навчання моделі та перевірку її якості за допомогою

таких метрик, як точність, повнота, F1—міра, ROC—крива та площа під нею. Визначено, що обрана модель демонструє високу ефективність при мінімальному ризику помилкових рішень. Також проаналізовано можливості подальшої інтеграції системи в банківське середовище з урахуванням вимог до безпеки, масштабованості та надійності.

Четвертий розділ було присвячено техніко—економічному обґрунтуванню створеної системи. Побудовано морфологічну карту функцій програмного продукту, проаналізовано варіанти реалізації, проведено експертне ранжування параметрів, обчислено коефіцієнти вагомості, а також визначено технічний рівень кожного з варіантів. Особливу увагу приділено економічному аналізу: розраховано витрати на оплату праці, машинний час, енергоспоживання, амортизацію та накладні витрати. Результатом стало визначення оптимального варіанта реалізації інформаційної системи, який забезпечує баланс між якістю, точністю та витратами на її впровадження.

Узагальнюючи проведену роботу, можна зазначити, що обраний підхід є актуальним, ефективним і економічно доцільним у сучасних умовах розвитку банківських технологій. Запропонована система може використовуватися як інструмент для підтримки рішень у банківських установах з метою підвищення об'єктивності та прозорості процесу кредитування. Вона забезпечує суттєве скорочення часу на аналіз заявок, зменшення людського фактору в прийнятті рішень та підвищення точності виявлення ризикованих клієнтів.

У перспективі система може бути доопрацьована з урахуванням нових типів вхідних даних, інтеграції з великими фінансовими базами, а також розширенням функціоналу для оцінки корпоративних клієнтів та складних фінансових продуктів. Це дозволить банкам формувати більш зважені та безпечні кредитні портфелі, що позитивно позначиться на їх фінансовій стабільності та конкурентоспроможності на ринку.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. Товстенко М.Ю. Економічна сутність та значення кредитоспроможності підприємств. 2021. URL: <https://dspace.mnau.edu.ua/jspui/bitstream/123456789/10613/1/zbirnyk—tez—25—11—21—55—56.pdf> (Дата звернення: 04.06.2025).
2. Rashid F., Khan R.A., Qureshi I.H. A Comprehensive Review of the Altman Z-Score Model Across Industries. 2023. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5044057 (Last accessed: 03.06.2025).
3. Солодовнік О.О., Докуніна К.І. Оцінка кредитоспроможності позичальника: аналіз основних підходів . *Інфраструктура ринку*. С. 6–18. 2021. URL: <https://eprints.kname.edu.ua/63351/1/26.pdf> (Дата звернення: 01.06.2025).
4. Панічук О.В. Управління зобов'язаннями в процесі оцінювання кредитоспроможності підприємства: обліково-інформаційний аспект. 2024. URL: <http://eztuir.ztu.edu.ua/handle/123456789/8479> (Дата звернення: 04.06.2025).
5. Tkachenko N., Kovalenko Y. та ін. New Methods in Assessing the Risks and Solvency of Insurance Companies. *Journal of Eastern European Research in Business and Economics*. P. 130-156. 2022. URL: <https://ibimapublishing.com/wp-content/uploads/articles/JEERBE/2022/765785/765785.pdf> .
6. Doumpos M., Lemonakis C., Niklis D. Introduction to credit risk modeling and assessment // *Credit Risk: An Overview*. 2019. URL: https://link.springer.com/chapter/10.1007/978-3-319-99411-6_1 .
7. ГОРОДЕЦЬКА Т.Б., ЗАЙЧЕНКО К.С., ІВАЩЕНКО А.Г. Методичні підходи до зниження кредитного ризику. 2021. URL:

- <https://elar.navs.edu.ua/items/10dcd26c-e250-4fbd-86bc-dc56c3782570> (Дата звернення: 05.05.2025).
8. Мовчанюк О.А. Кредитний скоринг як метод оцінки кредитного ризику // Формування ринкової економіки: зб. наук. пр. Київ, 2008. URL: <https://ir.kneu.edu.ua/server/api/core/bitstreams/28aa996a-c9b3-4f00-bc79-0affd0e1da37/content> (Дата звернення: 05.05.2025).
9. Збірник тез доповідей Всеукраїнської науково-практичної конференції «Сучасні виклики та тенденції розвитку фінансової системи». Івано-Франківськ, 23 квітня 2024 року. URL: https://kfin.pnu.edu.ua/wp-content/uploads/sites/53/2024/08/zbirnyk_24.pdf#page=213 (Дата звернення: 20.05.2025).
10. Сич О., Сінявіна А. Оцінка кредитного ризику банку в умовах воєнного стану. *Молодий вчений*. С. 7–14. 2024. URL: <https://molodyivchenyi.ua/index.php/journal/article/view/6279/6140> (Дата звернення: 21.05.2025).
11. Заволока Ю.М., Контурова С.М. Мінімізація кредитного ризику як основа ефективного управління банку. *Збірник наукових праць*. С. 6–18. 2016. URL: http://www.irbis-nbu.gov.ua/cgi-bin/irbis_nbu/cgiirbis_64.exe?C21COM=2&I21DBN=UJRN&P21DBN=UJRN&IMAGE_FILE_DOWNLOAD=1&Image_file_name=PDF/znptdau_2016_3_33.pdf (Дата звернення: 15.05.2025).
12. Caplinska A., Tvaronavičienė M. Creditworthiness place in credit theory and methods of its evaluation. *Entrepreneurship and Sustainability*. P. 12-17. 2020. URL: https://jssidoi.org/jesi/uploads/articles/27/Caplinska_Creditworthiness_place_in_Credit_Theory_and_methods_of_its_evaluation.pdf (Last accessed: 05.05.2025).

13. Tsai S.B., Li G., Wu C.H., Zheng Y., Wang J. An empirical research on evaluating banks' credit assessment of corporate customers . *SpringerPlus*. Vol. 24, No. 3. P. 1-41. 2016. URL: <https://link.springer.com/article/10.1186/s40064-016-3774-0> (Last accessed: 13.05.2025).
14. The Importance of Data Science & Analytics in Modern Banking. URL: <https://www.hunterbond.com/resources/blog/the-importance-of-data-science---analytics-in-modern-banking/> (Last accessed: 13.05.2025).
15. Extracting value from AI in banking: Rewiring the enterprise. URL: <https://www.mckinsey.com/industries/financial-services/our-insights/extracting-value-from-ai-in-banking-rewiring-the-enterprise> (Last accessed: 11.05.2025).
16. The Future of Artificial Intelligence in Banking: Transform Strategy with Predictive Analytics and Data Science. URL: <https://blog.aspiresys.com/banking-and-finance/the-future-of-ai-in-banking-transform-strategy-with-predictive-analytics-and-data-science/> (Last accessed: 05.05.2025).
17. How AI and Big Data Analytics Are Transforming Banking. URL: <https://motivitylabs.com/how-ai-and-big-data-analytics-are-transforming-banking/> ((Last accessed: 16.05.2025).
18. Top 11 Applications for AI in Banking Contact Centers. URL: <https://enthu.ai/blog/top-applications-of-ai-in-banking/> (Last accessed: 05.05.2025).
19. AI-first Banking: Top 10 AI-powered Use Cases Changing the BFSI Industry. URL: <https://www.cloud4c.com/blogs/10-ai-use-cases-in-banking-operations-explained> ((Last accessed: 01.05.2025).
20. 27 Real Examples of AI Implementation in Fintech and Banking. URL: <https://www.linkedin.com/pulse/27-authentic-examples-ai-implementation-fintech-nasser-sami-qtxee/> (Last accessed: 05.06.2025).

21. An effective role of artificial intelligence and machine learning in banking sector.
URL: <https://www.sciencedirect.com/science/article/pii/S2665917424001119>
(Last accessed: 05.05.2025).
22. SBS. Insights. How generative AI is driving the future of banking. URL:
<https://sbs-software.com/insights/generative-ai-future-of-banking/> (Last
accessed: 24.05.2025).
23. OECD. Artificial Intelligence, Machine Learning and Big Data in Finance. 2021.
URL: <https://www.oecd.org> (Last accessed: 24.05.2025).

ДОДАТОК А ЛІСТИНГ ПРОГРАМИ

```
# Необхідні бібліотеки
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Для моделей
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.ensemble import RandomForestClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report, confusion_matrix,
roc_auc_score, roc_curve

from xgboost import XGBRegressor

# Завантаження даних
df = pd.read_csv("Loan_Default.csv")

# Копія для безпечної обробки
data = df.copy()

# Розподіл цільової змінної
sns.countplot(data=data, x='Status')
plt.title("Розподіл цільової змінної (Status)")
plt.xlabel("0 = Без дефолту, 1 = Дефолт")
plt.ylabel("Кількість")
```

```
plt.show()
```

```
# Гендерний розподіл  
sns.countplot(data=data, x='Gender')  
plt.title("Гендерний розподіл")  
plt.xticks(rotation=45)  
plt.show()
```

```
# Лише числові змінні  
corr_data = data.select_dtypes(include=[np.number])  
plt.figure(figsize=(12, 10))  
sns.heatmap(corr_data.corr(), annot=True, cmap='coolwarm')  
plt.title("Кореляційна матриця")  
plt.show()
```

```
# Видаляємо ID, заповнюємо пропущені значення  
data = data.drop(columns=['ID'])  
data = data.dropna(axis=0, subset=['Status']) # тільки ті, хто має статус
```

```
# Кодування категоріальних змінних  
categorical_columns = data.select_dtypes(include=['object']).columns  
for col in categorical_columns:  
    le = LabelEncoder()  
    data[col] = le.fit_transform(data[col].astype(str))
```

```
# Масштабування  
scaler = StandardScaler()  
X = data.drop(columns=['Status'])  
y = data['Status']
```

```
X_scaled = scaler.fit_transform(X)

# Розбиття на train/test
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2,
random_state=42)

# Логістична регресія
log_reg = LogisticRegression()
log_reg.fit(X_train, y_train)
y_pred_lr = log_reg.predict(X_test)
print("Logistic Regression Report:")
print(classification_report(y_test, y_pred_lr))

# Случайный лес
rf = RandomForestClassifier()
rf.fit(X_train, y_train)
y_pred_rf = rf.predict(X_test)
print("Random Forest Report:")
print(classification_report(y_test, y_pred_rf))

# ROC-кривые
fpr_lr, tpr_lr, _ = roc_curve(y_test, log_reg.predict_proba(X_test)[:,-1])
fpr_rf, tpr_rf, _ = roc_curve(y_test, rf.predict_proba(X_test)[:,-1])
plt.plot(fpr_lr, tpr_lr, label="Logistic Regression")
plt.plot(fpr_rf, tpr_rf, label="Random Forest")
plt.plot([0, 1], [0, 1], 'k--')
plt.xlabel("False Positive Rate")
plt.ylabel("True Positive Rate")
plt.title("ROC-криві")
```

```
plt.legend()
```

```
plt.show()
```

```
# XGBoost (перістр)
```

```
xgb = XGBRegressor()
```

```
xgb.fit(X_train, y_train)
```

```
y_pred_xgb = xgb.predict(X_test)
```

```
y_pred_xgb_binary = [1 if p > 0.5 else 0 for p in y_pred_xgb]
```

```
print("XGBoost Report:")
```

```
print(classification_report(y_test, y_pred_xgb_binary))
```

ДОДАТОК Б ДОПОВІДЬ У ВИГЛЯДІ ПРЕЗЕНТАЦІЇ

Презентація дипломної роботи:

ЗАСТОСУВАННЯ ПІДХОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ДЛЯ ОЦІНКИ КРЕДИТНОЇ СПРОМОЖНОСТІ КЛІЄНТІВ БАНКУ

Комаренко Анастасія Романівна | КА – 15 Керівник: Гуськова В.Г.

1

МЕТА, ОБ'ЄКТ І ПРЕДМЕТ ДОСЛІДЖЕННЯ

Розробка прототипу системи оцінки кредитоспроможності клієнтів банку з використанням методів інтелектуального аналізу даних.

Процес прийняття рішень щодо надання кредитів у банках.

Методи машинного навчання для прогнозування кредитного ризику на основі фінансових, демографічних та поведінкових характеристик клієнтів.

2

ПОСТАНОВКА ЗАДАЧІ

Для досягнення мети було поставлено наступні задачі:

- Оцінити традиційні методи скорингу
- Дослідити сучасні ML-підходи
- Обґрунтувати вибір моделей
- Підготувати датасет для моделювання
- Розробити прототип скорингової системи
- Провести тестування та аналіз точності

3

АКТУАЛЬНІСТЬ ТЕМИ

Традиційні методи оцінки кредитоспроможності (логістична регресія, експертне оцінювання) вже не задовольняють вимоги часу.

Нові виклики банківського сектору:
– Зростання обсягів даних.
– Висока динаміка економіки.
– Необхідність швидких, точних і масштабованих рішень.

Інтелектуальний аналіз даних відкриває нові можливості для створення точних, адаптивних та автоматизованих систем скорингу.

4

ТЕОРЕТИЧНІ ОСНОВИ

Кредитоспроможність — здатність клієнта виконати фінансові зобов'язання.

Класичні підходи:

- - Фінансові коефіцієнти
- - Експертне оцінювання
- - Кредитний скоринг

Основні недоліки:

- - Суб'єктивність
- - Вузкий набір параметрів
- - Відсутність автоматизації

5

СУЧАСНІ ПІДХОДИ ДО СКОРИНГУ

Методи інтелектуального аналізу даних:

- Логістична регресія
- Random Forest
- Gradient Boosting (XGBoost)
- Нейронні мережі (NN)

Ключові переваги:

- Вища точність
- Обробка великих даних
- Виявлення прихованих закономірностей

| <u>Модель / Підхід</u> | <u>Точність</u> | <u>Інтерпретованість</u> | <u>Швидкість обчислень</u> | <u>Можливість масштабування</u> | <u>Складність впровадження</u> |
|----------------------------------|-----------------|--------------------------|----------------------------|---------------------------------|--------------------------------|
| Логістична регресія | Середня | Висока | Висока | Висока | Низька |
| Рішення на правилах (rule-based) | Низька | Висока | Висока | Середня | Низька |
| Random Forest | Висока | Середня | Середня | Висока | Середня |
| XGBoost / Gradient Boosting | Висока | Низька | Середня | Висока | Висока |
| Нейронні мережі | Дуже висока | Низька | Середня / низька | Висока | Висока |
| Гібридні моделі | Висока | Середня / висока | Середня | Середня / висока | Висока |

6

ПІДГОТОВКА ДАНИХ

Джерело: Kaggle Loan Default Dataset

Етапи:

1. Обробка відсутніх значень
2. Стандартизація числових змінних
3. Балансування класів (undersampling/oversampling)
4. Розбиття на train/test

| # ID | # year | # loan_limit | # Gender | # approval_adv | # loan_type | # loan_purpose | # Credit_Worthiness | # open_credit | | | | |
|--------------|--------|---------------|-------------------|---------------------|-------------|----------------|---------------------|---------------|-----|---|-----|------|
| ID | Year | Loan limit | Gender | Approved in advance | Loan Type | Loan Purpose | Credit worthiness | Open credit | | | | |
| cf | 91% | Male | 28% | nope | 84% | type1 | 76% | p3 | 38% | п | 96% | nope |
| ncf | 7% | Joint | 28% | pre | 16% | type2 | 14% | p4 | 37% | п | 4% | opc |
| Other (3344) | 2% | Other (54925) | 44% | Other (908) | 1% | Other (14735) | 10% | Other (27937) | 26% | | | |
| 24890 | 2019 | cf | Sex Not Available | nope | type1 | p1 | 11 | nope | | | | |
| 24891 | 2019 | cf | Male | noope | type2 | p1 | 11 | nope | | | | |
| 24892 | 2019 | cf | Male | pre | type1 | p1 | 11 | nope | | | | |
| 24893 | 2019 | cf | Male | noope | type1 | p4 | 11 | nope | | | | |
| 24894 | 2019 | cf | Joint | pre | type1 | p1 | 11 | nope | | | | |
| 24895 | 2019 | cf | Joint | pre | type1 | p1 | 11 | nope | | | | |
| 24896 | 2019 | cf | Joint | pre | type1 | p3 | 11 | nope | | | | |
| 24897 | 2019 | cf | Female | noope | type1 | p4 | 11 | nope | | | | |
| 24898 | 2019 | cf | Joint | noope | type1 | p3 | 11 | nope | | | | |
| 24899 | 2019 | cf | Sex Not Available | noope | type3 | p3 | 11 | nope | | | | |

7

ПОБУДОВА МОДЕЛЕЙ

Реалізовані моделі:

- Logistic Regression
- Random Forest
- XGBoost
- Штучні нейронні мережі

Інструменти:

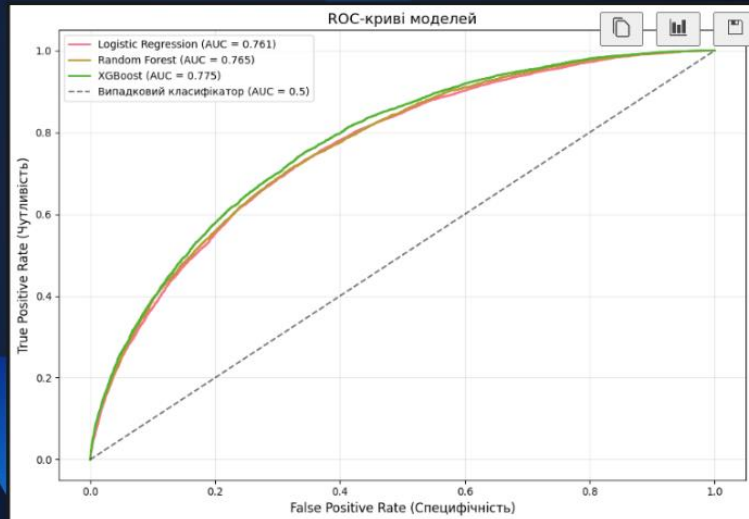
- Python
- Scikit-learn
- TensorFlow
- Pandas

8

АНАЛІЗ РЕЗУЛЬТАТІВ

Критерії оцінки:

- Accuracy
- Precision
- Recall
- F1-score
- ROC-AUC



9

АНАЛІЗ РЕЗУЛЬТАТІВ

Логістична регресія:

Accuracy: 0.7859
Precision (клас 1): 0.61
Recall (клас 1): 0.22
F1-score (клас 1): 0.37

Random Forest:

Accuracy: 0.7894
Precision: 0.63
Recall: 0.19
F1-score: 0.30

XGBoost:

Accuracy: 0.7919
Precision: 0.63
Recall: 0.22
F1-score: 0.33

10

ВИСНОВКИ

Мета роботи досягнута:

- Реалізовано прототип системи скорингу
- Проведено тестування моделей ML
- Отримано високу точність прогнозування
- Запропоновано альтернативу комерційним рішенням

Перспективи:

- Інтеграція з банківськими API
- Підключення до кредитних бюро
- Розширення типів вхідних даних

11

ДЯКУЮ ЗА УВАГУ!

Комаренко Анастасія Романівна | КА – 15 Керівник: Гуськова В.Г.

12