

Міністерство освіти і науки України
Національний технічний університет
«Дніпровська політехніка»

Факультет інформаційних технологій
(факультет)

Кафедра системного аналізу та управління
(повна назва)

ПОЯСНЮВАЛЬНА ЗАПИСКА
кваліфікаційної роботи ступеня бакалавра

Здобувача вищої освіти Хижняк Марії Романівни

академічної групи 124-21-1

спеціальності 124 Системний аналіз

за освітньо-професійною програмою Системний аналіз

на тему: «Розробка інтелектуальної системи прийняття рішення щодо встановлення кредитного ліміту»

Керівники	Прізвище, ініціали	Оцінка за шкалою		Підпис
		рейтинговою	інституційною	
кваліфікаційної роботи	<i>к.т.н., доц. Желдак Т.А.</i>			
розділів:				
Інформаційно- аналітичний	<i>к.т.н., доц. Желдак Т.А.</i>			
Спеціальний розділ	<i>к.т.н., доц. Желдак Т.А.</i>			
Експериментально- аналітичний	<i>к.т.н., доц. Желдак Т.А.</i>			
Рецензент	<i>к.т.н., доц. Приходченко С.Д.</i>			
Нормоконтролер	<i>к.ф.-м.н., доц. Хом'як Т.В.</i>			

Дніпро
2025

ЗАТВЕРДЖЕНО:

завідувач кафедри

Системного аналізу та управління

(повна назва)

к.т.н., доц. Желдак Т.А.

(підпис)

(прізвище, ініціали)

« » 20 року

ЗАВДАННЯ
на кваліфікаційну роботу
ступеня бакалавра

здобувачу вищої освіти Хижняк М. Р. академічної групи 124- 21-1

спеціальності: 124 Системний аналіз

за освітньо-професійною програмою Системний аналіз

на тему «Розробка інтелектуальної системи прийняття рішення щодо

встановлення кредитного ліміту»

затверджену наказом ректора НТУ «Дніпровська політехніка» від 05.05.2025 р.

№336-с

Розділ	Зміст	Терміни виконання
1. Інформаційно-аналітичний розділ	<i>Проаналізувати структуру об'єкта дослідження. Визначити предметну область дослідження та проблему, що розв'язується.</i>	23.12.2024 – 01.03.2025
2. Спеціальний розділ	<i>Обґрунтувати метод виконання поставленого завдання. Розв'язати поставлену задачу: сформулювати математичну модель, реалізувати алгоритм дерева рішень</i>	01.03.2025 – 28.04.2025
3. Експериментально-аналітичний	<i>Перевірити якість побудованої моделі за різними критеріями, дослідити вплив параметрів на результат та порівняти з альтернативними методами класифікації.</i>	28.04.2025 – 02.06.2025

Завдання видано

(підпис)

доц. Желдак Т.А.

(прізвище, ініціали)

Дата видачі: 23.12.2024 р.

Дата подання до екзаменаційної комісії: 18.06.2025 р.

Прийнято до виконання

(підпис студента)

Хижняк М. Р.

(прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 72 с., 19 рис., 6 табл., 4 додатки, 36 джерел.

Об'єктом дослідження в роботі є процес прийняття рішень про кредитування клієнтів фінансовими установами.

Предметом дослідження є інтелектуальна система попередньої оцінки кредитного ризику фізичних осіб.

Метою даної кваліфікаційної роботи є підвищення якості попереднього оцінювання кредитного ризику шляхом розроблення підходу, що базується на характеристиках фінансової поведінки клієнта і забезпечує точність, послідовність та обґрунтованість результатів.

Методи дослідження: аналіз даних, моделювання за допомогою дерева рішень, крос-валідація, оцінювання за метриками класифікації, порівняльний аналіз.

В *інформаційно-аналітичному розділі* проаналізовано стан кредитної діяльності банку — бази проходження практики, розкрито внутрішні процедури, інформаційне забезпечення та стратегічні напрями, що дозволило виявити обмеження чинної моделі оцінювання ризику.

У *спеціальному розділі* сформовано концепцію інтелектуальної системи для попередньої оцінки кредитного ризику, здійснено її математичне формулювання, обґрунтовано вибір алгоритму дерева рішень та реалізовано побудову моделі на основі реалістичних вхідних характеристик.

Експериментально-аналітичний розділ присвячено перевірці якості побудованої моделі за різними критеріями, оцінці її стійкості та порівнянню з альтернативними підходами для обґрунтування ефективності обраного рішення.

Практична цінність отриманих результатів полягає в тому, що запропонований підхід до первинної оцінки кредитного ризику забезпечує поєднання точності та пояснюваності рішень, що робить його придатним для інтеграції в автоматизовані системи банківського скорингу.

Ключові слова: КРЕДИТНИЙ РИЗИК, ІНТЕЛЕКТУАЛЬНА СИСТЕМА, ДЕРЕВО РІШЕНЬ, КЛАСИФІКАЦІЯ, СКОРИНГ.

ABSTRACT

Explanatory note: 72 pages, 19 figures, 6 tables, 4 appendices, 36 references.

The object of the study is the decision-making process regarding lending to clients by financial institutions.

The subject of the study is an intelligent system for preliminary assessment of credit risk of individuals.

The purpose of this qualification work is to improve the quality of preliminary credit risk assessment by developing an approach based on the characteristics of the client's financial behavior, ensuring accuracy, consistency, and justification of results.

Research methods: data analysis, decision tree modeling, cross-validation, evaluation using classification metrics, comparative analysis.

The information-analytical section analyzes the state of credit activity of the bank used as the practice base, outlines internal procedures, information support, and strategic directions, which made it possible to identify the limitations of the current risk assessment model.

The special section presents the concept of an intelligent system for preliminary credit risk assessment, provides its mathematical formulation, justifies the choice of the decision tree algorithm, and implements the model based on realistic input features.

The experimental-analytical section focuses on evaluating the quality of the constructed model using various criteria, assessing its stability, and comparing it with alternative approaches to justify the effectiveness of the chosen solution.

The practical value of the obtained results lies in the fact that the proposed approach to primary credit risk assessment combines accuracy and interpretability of decisions, making it suitable for integration into automated banking scoring systems at the preliminary analysis stage.

Keywords: CREDIT RISK, INTELLIGENT SYSTEM, DECISION TREE, CLASSIFICATION, SCORING.

ЗМІСТ

ВСТУП	7
1 ІНФОРМАЦІЙНО-АНАЛІТИЧНИЙ РОЗДІЛ	10
1.1 Історія та еволюція Абанку	10
1.2 Кредитні продукти в структурі фінансових сервісів банку	12
1.3 Організаційна структура банку в контексті управління кредитною діяльністю	14
1.4 Інформаційне забезпечення процесу оцінки кредитоспроможності клієнтів	19
1.5 Аналіз фінансових показників кредитної активності банку	24
1.6 Стратегія розвитку банку в сфері споживчого кредитування	27
1.7. Недоліки чинної кредитної моделі та напрямки її покращення.....	29
1.8 Висновок до розділу 1	32
2 СПЕЦІАЛЬНИЙ РОЗДІЛ	34
2.1 Змістовне формулювання задачі	34
2.2 Математична формалізація задачі класифікації.....	37
2.3 Вибір методу та його обґрунтування	38
2.4 Теоретичні основи алгоритму.....	41
2.5 Реалізація та структура моделі	43
2.6 Висновок до розділу 2	48
3 ЕКСПЕРИМЕНТАЛЬНО-АНАЛІТИЧНИЙ РОЗДІЛ	50
3.1 Перевірка базової продуктивності моделі	50
3.2 Аналіз впливу параметра глибини дерева	53
3.3 Перевірка стабільності моделі за допомогою крос-валідації	56

3.4 Аналіз важливості ознак.....	57
3.5 Перевірка справедливості моделі щодо соціально чутливих характеристик	59
3.6 Порівняння якості класифікації з іншими моделями	62
3.7 Висновок до розділу 3	63
ВИСНОВОК.....	65
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	67
ДОДАТОК А	71
ДОДАТОК Б.....	72
ДОДАТОК В	73
ДОДАТОК Г.....	74
Г.1 Програмний код моделі дерева рішень	74
Г.2 Опис програмної реалізації моделі.....	80

ВСТУП

Сучасна банківська система стикається з численними викликами, пов'язаними з необхідністю оперативного та обґрунтованого прийняття рішень щодо кредитування фізичних осіб. Зростання обсягів споживчого кредитування, поява нових форм взаємодії з клієнтами та посилення конкуренції між фінансовими установами вимагають від банківських структур впровадження інструментів, що поєднують точність, пояснюваність і адаптивність. Особливе значення це має в умовах нестабільної економіки України, де правильне визначення кредитного ризику є ключовим як для мінімізації втрат банку, так і для збереження фінансової доступності для громадян.

Актуальність даної теми зумовлена потребою вдосконалення підходів до оцінки кредитоспроможності, які застосовуються на етапі попереднього прийняття рішення про надання кредитного ліміту. Незважаючи на широку автоматизацію процесів, саме в початковій точці аналізу залишається високий ризик похибки, пов'язаної з обмеженим обсягом інформації та необхідністю швидкої реакції. У таких умовах значення набувають не лише точність розрахунку, а й прозорість механізму, що стоїть за результатом, — адже він має бути зрозумілим, стабільним і таким, що піддається перевірці. Відповідно, дослідження спрямоване на формування підходу, здатного забезпечити надійні результати на основі реалістичних вхідних характеристик і відповідати практичним потребам банківського сектору України.

Метою роботи є підвищення якості попереднього оцінювання кредитного ризику шляхом розроблення підходу, який базується на доступних характеристиках фінансової поведінки клієнта і забезпечує точність, послідовність та обґрунтованість результатів.

Для досягнення цієї мети передбачено виконання таких завдань:

- проаналізувати умови, у яких приймаються рішення щодо кредитування в розрізі обмеженості доступних даних та вимог до якості оцінювання;

- сформувати підхід до попереднього оцінювання кредитного ризику на основі релевантних характеристик фінансової поведінки клієнтів;
- реалізувати побудований підхід та забезпечити можливість оцінки обґрунтованості отриманих результатів;
- перевірити точність і стабільність класифікації на реальних даних;
- визначити потенціал практичного застосування підходу в контексті банківських задач.

Об'єктом дослідження є процес прийняття рішень про кредитування клієнтів фінансовими установами.

Предметом дослідження виступає інтелектуальна система попередньої оцінки кредитного ризику фізичних осіб на основі відкритих і доступних на момент звернення клієнта ознак.

Методи дослідження, застосовані в роботі, включають: аналіз даних для виявлення ключових характеристик фінансової поведінки клієнтів; моделювання за допомогою дерева рішень та подальшим уточненням структури; крос-валідацію для перевірки узагальнювальної здатності моделі; оцінювання за метриками якості класифікації; методи порівняльного аналізу з іншими алгоритмами.

Наукова новизна полягає в розробленні підходу до попередньої оцінки кредитного ризику з елементами пояснення результатів і формуванням базових рекомендацій залежно від типу клієнта. Удосконалено використання моделі дерева рішень шляхом її подальшого обрізання з метою підвищення стабільності класифікації за збереження інтерпретованості. Запропонований підхід розрахований на використання в умовах обмеженого обсягу вхідної інформації та враховує практичні вимоги до швидкості й обґрунтованості первинного рішення в банківському середовищі.

Практичне значення отриманих результатів полягає в можливості використання запропонованого підходу як складової автоматизованих систем первинного оцінювання кредитного ризику в банках та інших фінансових установах. Побудована модель забезпечує баланс між точністю, стабільністю та

зрозумілістю рішень, що дає змогу застосовувати її в умовах обмеженої вхідної інформації на ранньому етапі кредитного аналізу. Отримані результати можуть бути використані для прийняття попередніх рішень щодо доцільності надання кредиту та для виявлення категорій клієнтів, що потребують додаткової перевірки або альтернативних умов кредитування.

1 ІНФОРМАЦІЙНО-АНАЛІТИЧНИЙ РОЗДІЛ

1.1 Історія та еволюція Абанку

Абанк, офіційно зареєстрований як АКЦІОНЕРНЕ ТОВАРИСТВО «АКЦЕНТ-БАНК», був заснований 28 травня 1992 року під назвою «Київський приватний банк «Київприватбанк». З моменту створення банк неодноразово змінював як свою юридичну форму, так і найменування відповідно до законодавчих змін і стратегічних рішень керівництва. Згідно з відповідними реєстраційними документами, першим із таких кроків стало перейменування установи 19 травня 1993 року на «Український кредитний банк». 6 серпня 2001 року банк був реорганізований у закрите акціонерне товариство під назвою «Український кредитний банк». У 2007 році, внаслідок фінансових труднощів і згодом за ініціативи Національного банку України, було проведено санацію банку з боку Приватбанку. Це вплинуло на подальші структурні перетворення: назву банку змінено на ЗАТ «АКЦЕНТ-БАНК», а головний офіс перенесено з Києва до Дніпра. Ця реорганізація була задокументована 3 вересня 2007 року. 17 вересня 2009 року банк змінив свою юридичну форму на публічне акціонерне товариство, зареєстроване як ПАТ «АКЦЕНТ-БАНК». Остання значна зміна відбулася 27 червня 2018 року, коли банк перейшов від публічного акціонерного товариства до приватного акціонерного товариства. З цього часу банк функціонує під назвою АКЦІОНЕРНЕ ТОВАРИСТВО «АКЦЕНТ-БАНК» (АТ «А-БАНК»).

Наразі Абанк зареєстрований за адресою: 49074, Україна, м. Дніпро, вул. Батумська, 11. Станом на 31 грудня 2024 року, банк нараховує 213 відокремлених підрозділів. До складу банку також входять спеціалізовані каси перерахунку та центри інкасації, які забезпечують виконання критично важливих операцій.

Основні напрямки діяльності Абанку охоплюють такі сфери, як залучення депозитів, відкриття та ведення клієнтських рахунків, кредитування, гарантійні операції, а також проведення розрахунково-касових операцій і торгівлю цінними

паперами та іноземною валютою. Усі ці операції регулюються Національним банком України, під контролем якого банк знаходиться з моменту отримання банківської ліцензії №16 від 26 жовтня 2011 року.

Відповідно до чинних регуляторних вимог, Акцент-банк входить до державної системи гарантування вкладів, маючи статус спеціалізованого ощадного банку, діяльність якого зосереджена на обслуговуванні фізичних осіб та роботі з їхніми депозитними внесками. А 5 березня 2020 року заклад отримав статус системно важливого банку — установи, стабільність якої має вирішальне значення для всієї банківської системи, що зобов'язує його дотримуватися підвищених нормативів, зокрема щодо розміру основного капіталу.

Крім того, АТ «А-БАНК» активно інтегрований у міжнародні фінансові мережі, будучи учасником платіжних систем VISA та Mastercard, а також національної платіжної системи «ПРОСТІР» [32, с. 7-8].

12 грудня 2024 року незалежне рейтингове агентство «Кредит-Рейтинг» підтвердило довгостроковий кредитний рейтинг АТ «А-БАНК» на рівні uaAA- за національною шкалою, що означає дуже високу кредитоспроможність у межах країни, наближену до максимальної (uaAAA). Вперше цей рівень рейтингу було присвоєно банку 20 червня 2019 року, тоді ж прогноз оцінки був визначений як «стабільний», а у 2022–2023 роках рейтинг тимчасово супроводжувався негативним прогнозом, що вказував на можливе зниження оцінки за умов збереження ризиків, а згодом — прогнозом «у розвитку», який свідчив про підвищену ймовірність зміни рейтингу. Наразі прогноз залишається стабільним, що вказує на відсутність очікуваних змін у короткостроковій перспективі [23]. Детальну динаміку змін прогнозу кредитного рейтингу представлено на рисунку 1.1.

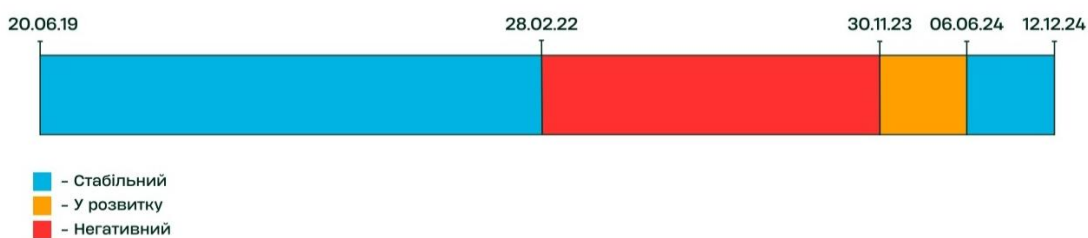


Рисунок 1.1 – Зміни прогнозу кредитного рейтингу АТ «А-БАНК» за період 2019–2024 років

22 квітня 2025 року агентство також підтвердило рейтинг надійності банківських вкладів АТ «А-БАНК» на рівні 4 за п’ятибальною шкалою. Такий рівень характеризує вклад як достатньо захищений: ймовірність виникнення труднощів із його своєчасним поверненням оцінюється як невисока, хоча залишається чутливість до зовнішніх фінансових та економічних чинників [24].

1.2 Кредитні продукти в структурі фінансових сервісів банку

Як уже зазначалося в підрозділі 1.1 Історія та еволюція Абанку, діяльність АТ «А-БАНК» ґрунтується на трьох основних напрямках: обслуговуванні платіжних карток, кредитуванні та депозитних продуктах. Саме ці напрямки формують основу продуктової політики установи й визначають її позиціонування на ринку. Разом із тим, банк розвиває окремі рішення для приватних та корпоративних клієнтів, розширюючи спектр фінансових інструментів відповідно до попиту.

Отже, по-перше, платіжні картки становлять окрему категорію продуктів, що функціонують як електронні платіжні інструменти, що забезпечують зарахування коштів, здійснення платежів та розрахунків за власні кошти або за рахунок встановленого кредитного ліміту. Оскільки дослідження зосереджене на аналізі кредитних процесів, доцільно виокремити картки з кредитною функцією, а саме: «Зелену» картку, що є дебетово-кредитним продуктом із лімітом до 200 000 грн, орієнтованим на фізичних осіб, та преміальні картки, які передбачають розширений ліміт — від 500 000 до 10 млн грн. Додатково банк пропонує дебетову «Жовту» картку, а також кобрендингові рішення («АТБ», «Вигода»

(«Епіцентр»), «Volt»), соціальні картки («Підтримка», «Відновлення», «Національний кешбек», «Дія.Картка») та спеціалізований продукт — дитячу картку [25].

Другим основним напрямом діяльності банку є депозитні програми, що передбачають розміщення клієнтських коштів на рахунках з нарахуванням відсотків відповідно до умов договору. Абанк пропонує кілька типів депозитів, які відрізняються тривалістю розміщення коштів, порядком виплати доходу, а також умовами поповнення та дострокового зняття: «Online», «Стандарт», «Стандарт строковий», «Скарбничка», «А-вклад», «Рантьє», «Допомога», «КіберЗахист» [22].

По-третє, у рамках дослідження доцільно окреслити перелік кредитних продуктів, що пропонуються банком, аби краще зрозуміти структуру надання фінансування та підходи до кредитування в цілому. Серед таких продуктів вирізняються наступні позиції:

- Кредит «Плати частинами» дозволяє клієнтам розділити вартість товару на декілька рівних частин, які сплачуються щомісяця. Такий кредит видається для купівлі товарів у партнерських магазинах.
- Послуга «Швидка готівка» — це кредит, що видається напряму на банківську картку клієнта із можливістю швидкого затвердження. Цей продукт забезпечує тимчасову фінансову підтримку для клієнтів, що потребують швидкого доступу до коштів.
- Кредит «Під депозит» виступає як опція для тих клієнтів, які мають депозити в банку і хочуть отримати кредит під заставу цих депозитів. Це забезпечує банку додаткову гарантію повернення позичених коштів, а клієнту — можливість отримати кредит на вигідних умовах, зокрема, з нижчою процентною ставкою.
- Кредит «Розстрочка на витрати» надає можливість перетворити здійснені витрати за допомогою банківської картки на кредит, що сплачується у розстрочку. Клієнти можуть використовувати цю послугу

для покриття витрат, які були зроблені не пізніше ніж 30 днів тому, з метою оптимізації своїх фінансових навантажень.

- Кредит «Розстрочка «Вигода» — програма, що специфічно орієнтована на клієнтів, які роблять покупки в мережі магазинів «Епіцентр» та «Нова лінія». Ця кредитна послуга дає змогу покупцям розбити загальну вартість товарів на рівні частини, що сплачуються щомісяця без передоплати. Продукт відрізняється від інших схожих пропозицій здатністю перетворювати попередні витрати в кредит [26].

Крім основних банківських продуктів, Абанк надає комплекс додаткових послуг, орієнтованих на потреби бізнес-клієнтів та юридичних осіб. Зокрема, банк здійснює відкриття та обслуговування рахунків, підтримує зарплатні проекти, надає доступ до торгового еквайрингу, інкасації, профільних депозитних програм, супроводу зовнішньоекономічних контрактів, а також можливість впровадження кредитних програм типу «Плати частинами». Серед спеціалізованих рішень передбачено мобільний додаток АBank24, оренду сейфових скриньок, фінансовий лізинг автомобілів з лімітом фінансування до 800 тис. грн, страхові послуги, а також пенсійні програми з можливістю зарахування виплат на картковий рахунок [21, 29].

1.3 Організаційна структура банку в контексті управління кредитною діяльністю

Для розуміння логіки управлінських процесів у банку, в тому числі в частині розгляду кредитних заявок, слід звернутися до його організаційної структури. Вона формує підґрунтя для функціонування як операційних, так і контролюючих підрозділів, розподіляючи повноваження між стратегічними, виконавчими та допоміжними рівнями. У структурі АТ «А-БАНК» виділяються як централізовані органи управління, відповідальні за загальне керівництво, так і спеціалізовані напрями, що забезпечують виконання окремих функцій, у тому

числі пов'язаних із кредитуванням. Більш докладно, ключові елементи управлінської структури включають:

- a) Наглядову раду, яка відіграє роль стратегічного керівництва, забезпечуючи загальне управління та контроль за діяльністю банку.
- b) Правління банку та Голову правління, які відповідають за оперативне управління та реалізацію стратегічних завдань.
- c) Службу внутрішнього аудиту, яка контролює внутрішню діяльність банку, забезпечує дотримання внутрішніх правил і процедур.
- d) Управління ризиками, яке відповідає за виявлення, оцінку та моніторинг різних видів ризиків, що супроводжують банківську діяльність, включаючи один із найвагоміших — кредитний. Цей напрям розподіляється на такі структурні одиниці:
 - 1) Департамент Compliance, що здійснює функції комплаєнс — забезпечує своєчасне виявлення, вимірювання, моніторинг, контроль, пом'якшення та звітування щодо комплаєнс-ризиків, тобто ризику невідповідності внутрішнім та зовнішнім регуляторним вимогам.
 - 2) Напрямок ризик-менеджменту, що відповідає за управління кредитним ризиком, ризиком процентної ставки банківської книги, ризиком ліквідності, ринковими ризиками, операційним ризиком та іншими. Важливо, що саме в межах цього напрямку реалізуються основні підходи до оцінки кредитоспроможності позичальника — включаючи аналіз фінансового стану, структури власності, заставного майна, а також використання скорингових моделей під час ухвалення кредитних рішень.
 - 3) Напрямок технологій стягнення проблемних заборгованостей, що розробляє та впроваджує стратегії роботи з непрацюючими активами та контролює процеси стягнення простроченої заборгованості.
 - 4) Департамент фінансового моніторингу, що здійснює постійний нагляд за фінансовими операціями відповідно до внутрішніх політик

банку та вимог законодавства у сфері запобігання фінансовим зловживанням.

У межах виконавчої вертикалі банку, в структурі Правління АТ «А-БАНК», функціонують колегіальні органи, які забезпечують ухвалення стратегічно важливих рішень. Особливе значення мають кредитні комітети, діяльність яких безпосередньо впливає на реалізацію кредитної політики банку, а саме:

- Кредитний комітет визначає правила та підходи до кредитування, включаючи аналіз кредитних ризиків, затвердження внутрішніх політик і погодження кредитних заявок, що виходять за межі делегованих повноважень. Його компетенція також охоплює затвердження лімітів активних операцій, надання та пролонгація кредитів, погодження банківських гарантій, а також управління групами пов'язаних осіб і оцінка прибутковості продуктів.
- Малий кредитний комітет розглядає кредитні питання меншого обсягу — переважно в межах звернень від ФОП і ТОВ — забезпечуючи оперативність у прийнятті рішень щодо затвердження, збільшення або коригування вже встановлених лімітів.

Крім кредитної ланки, у складі виконавчих органів діють комітети, відповідальні за окремі напрямки операційного управління:

- Комітет з управління активами та пасивами здійснює планування структури балансу банку, ухвалює рішення щодо умов залучення та розміщення ресурсів і оцінює пов'язані фінансові ризики.
- Комітет з управління інформаційною безпекою визначає політики захисту банківських даних і координує реагування на загрози, що виникають у цифровому середовищі.
- Комітет з управління непрацюючими активами відповідає за розгляд проблемних активів, включаючи списання боргів і аналіз тривалих прострочень.

- Малий комітет з управління непрацюючими активами функціонує для оперативного розгляду індивідуальних кейсів та трансформації заборгованості. [33, с. 4-10].

У цілому, організаційна модель управління в АТ «А-БАНК» побудована на фіксованих структурних принципах, що визначають взаємозв'язки між основними рівнями. Насамперед, її характер є горизонтальним: кілька ключових рівнів функціонують паралельно – у цьому випадку це реалізовано через головний офіс і мережу регіональних відділень, між якими розподілено управлінські повноваження. Структура також має дивізіональний характер, тобто поділяється на клієнто- та функціонально орієнтовані напрями, що забезпечують виконання ключових завдань відповідно до типу клієнтів або функціональної спеціалізації [33, с. 10-11]. Візуальне представлення загальної організаційної структури наведено на рисунку 1.2.

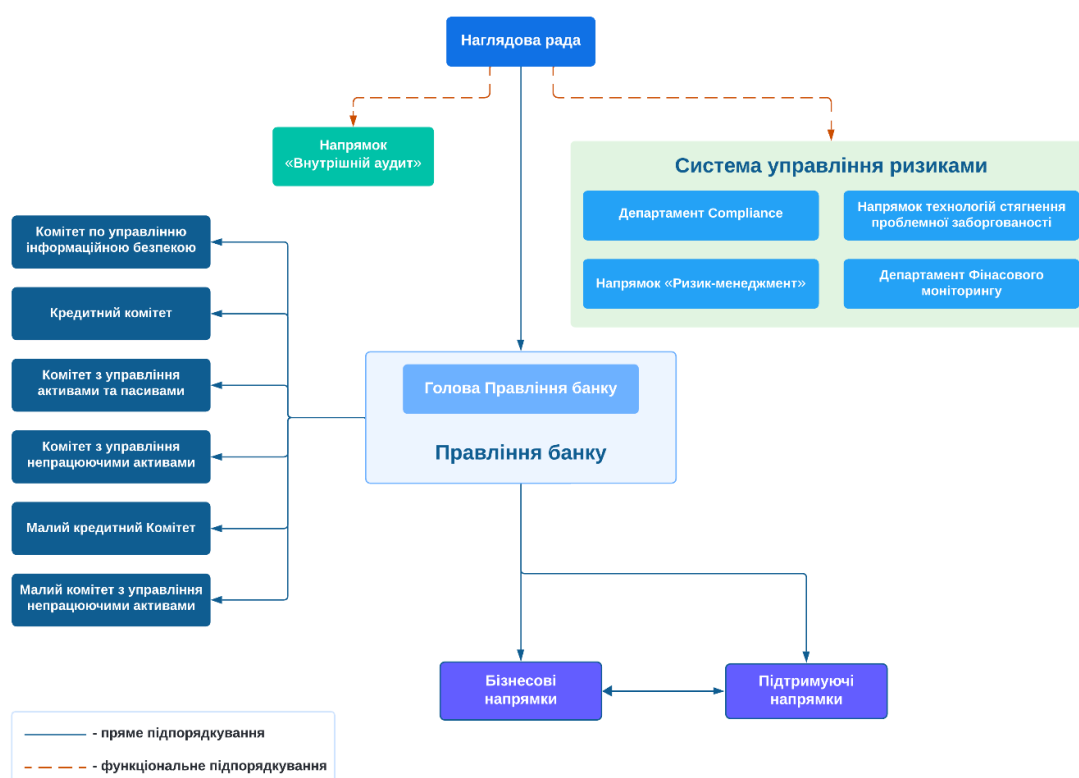


Рисунок 1.2 – Узагальнена організаційна структура управління Абанку

Відповідно, в організаційній ієрархії банку виокремлюються два основні типи напрямків: бізнесові та підтримуючі. Бізнес-напрямки мають безпосередній зв'язок з клієнтами та формують основу доходів банку. Підтримуючі напрямки, у свою чергу, створюють умови для функціонування банку в цілому, забезпечуючи роботу основних процесів, не взаємодіючи напряму з клієнтами.

Щодо кредитного напрямку, то в межах бізнес-блоку функціонують кілька структур, безпосередньо залучених до процесу надання позик:

- Бізнес А-Кредит, що спеціалізується на видачі кредитів фізичним особам у торгових точках, відділеннях банку та через онлайн-канали;
- напрямок Корпоративного бізнесу, що забезпечує обслуговування і кредитування юридичних осіб;
- напрямок малого та середнього бізнесу, відповідальний за залучення та кредитування малого та середнього бізнесу;
- напрямок споживчого кредитування, що надає нецільові позики фізичним особам.

Окремо можна виділити напрямок «Кредитні картки», який зосереджений на просуванні карткових продуктів банку. Попри те, що діяльність цього напрямку пов'язана переважно з обслуговуванням карток, значна частина таких продуктів включає кредитний ліміт, що функціонально наближує його до кредитного сегмента.

Серед інших бізнес-направків, орієнтованих на надання банківських продуктів і послуг, можна виділити напрямок «Заощадження» (робота з депозитами фізичних осіб), «Роздрібні продажі» (платіжні операції та грошові перекази), «Internet banking», а також «Казначейство», «Касове обслуговування та інкасація», «Операційний центр» (Call-center) і регіональні напрями, якими опікуються відповідні директори.

Підтримуючі підрозділи можна умовно згрупувати відповідно до характеру їхніх внутрішніх завдань. Зокрема, у сфері роботи з персоналом діють напрямки «Трудових ресурсів (HR)», напрямки корпоративного навчання та напрямки «Охорона праці». Облікову й фінансову діяльність забезпечують напрямки

бюджетування, «Бухгалтерія», департамент обліку внутрішньобанківських операцій і департамент з організації роботи архівів. До підрозділів, які супроводжують операційні процеси, належать напрямок «Розвиток робочих процесів», департамент «Вхідна кореспонденція», напрямок «Процесінг» та Back-office. Окремо функціонують юридичний департамент і напрямок «Інформаційні технології», який об'єднує декілька підрозділів, відповідальних за технічну підтримку й цифрову інфраструктуру [34, с. 5].

Отже, аналіз організаційної структури АТ «А-БАНК» дозволяє простежити, як у межах взаємодії управлінських рівнів, колегіальних органів і бізнес-напрямків реалізується процес ухвалення кредитних рішень. Загальні підходи до оцінки ризиків і визначення допустимих меж кредитування формуються на рівні управління ризиками, тоді як затвердження ключових параметрів — таких як кредитні ліміти, умови фінансування й зміни до політик — відбувається через діяльність кредитних комітетів. Безпосередню роботу з клієнтами здійснюють відповідні бізнес-напрями, що відповідають за обслуговування фізичних і юридичних осіб, а також представників малого та середнього бізнесу. У рамках цього процесу розгляд індивідуальних заявок супроводжується використанням цифрових інструментів, зокрема скорингових моделей, які забезпечують узгодженість управлінських рішень і адаптацію до конкретних характеристик позичальника.

1.4 Інформаційне забезпечення процесу оцінки кредитоспроможності клієнтів

Абанк використовує розгалужену інформаційну систему, яка охоплює усі рівні банківської діяльності — від обслуговування клієнтів до внутрішнього аналізу та супроводу операцій. Її архітектура базується на сукупності взаємопов'язаних компонентів: баз даних, програмних модулів, платформ обробки та інтегрованих моделей, що спільно забезпечують збирання, зберігання, обробку й аналіз інформації.

У структурі цієї системи функціонують окремі елементи, відповідальні за різні етапи обігу даних. Блоки зберігання охоплюють центральні бази даних, зокрема клієнтську базу та системи обліку транзакцій і кредитних операцій. Архітектура обробки реалізується через платформи, які забезпечують автоматизацію бізнес-процесів, управління операціями в реальному часі та підтримку налаштувань банківських продуктів. Крім того, до системи інтегруються модулі аналітичного оцінювання, що використовуються для моніторингу ризиків і формування управлінських рішень на основі цифрових моделей.

На рисунку 1.3 представлено загальну архітектуру інформаційної системи банку, яка складається з кількох взаємопов'язаних модулів, розміщених у хмарному середовищі. Вони виконують різні функції: зберігають та обробляють дані клієнтів, управляють транзакціями, координують банківські процеси й забезпечують роботу цифрових каналів. Ці елементи тісно інтегровані між собою та взаємодіють із внутрішньою інфраструктурою банку в режимі реального часу.

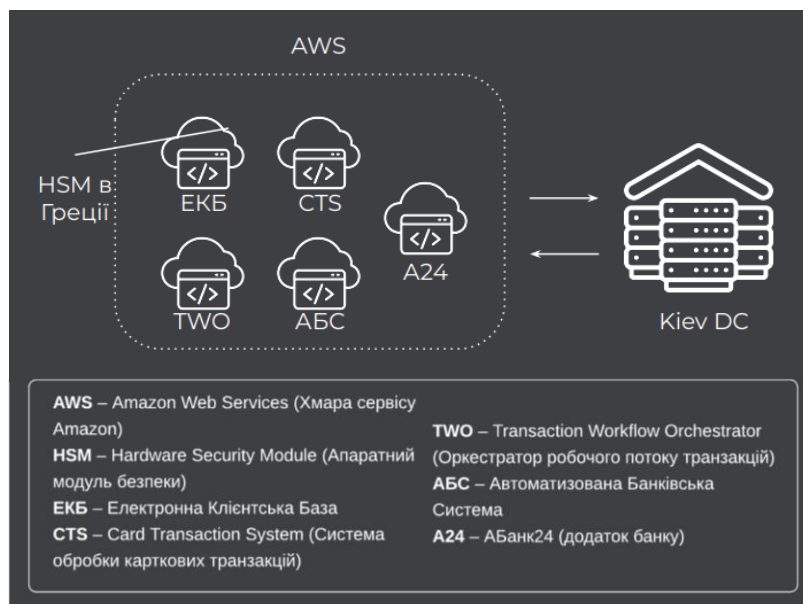


Рисунок 1.3 – Схематичне представлення інформаційної архітектури АТ «А-БАНК»

Центральним модулем архітектури є Автоматизована Банківська Система (АБС), яка є базовою ланкою обліку та управління основними операціями:

відкриттям рахунків, проведенням платежів, кредитуванням і супутніми процесами. Зберігання та структурування інформації про клієнтів забезпечує Електронна Клієнтська База (ЕКБ), що виступає ключовим джерелом даних для інших систем. Паралельно функціонує система обробки карткових транзакцій (Card Transaction System, CTS), яка відповідає за фіксацію та контроль платежів за банківськими картками. Координацію між усіма модулями забезпечує спеціалізований оркестратор (Transaction Workflow Orchestrator, TWO), який синхронізує виконання операцій і гарантує послідовність дій у межах єдиного цифрового середовища. Інформація, зафіксована в АБС, ЕКБ і CTS, у подальшому використовується для аналітичних цілей, включаючи скорингові та поведінкові моделі, що застосовуються під час оцінки кредитоспроможності клієнтів [31].

Подальша обробка, маршрутизація й узгодження внутрішніх процесів у банку здійснюється за допомогою спеціалізованих платформ, а саме Scrooge та Corezoid. Система Scrooge виконує функції централізованого управління операціями: вона забезпечує бухгалтерський облік, обробку платежів, формування аналітичної звітності та підтримку внутрішніх сервісів банку [20]. Платформа Corezoid використовується для налаштування й автоматизації бізнес-процесів у реальному часі — вона дозволяє будувати адаптивні схеми обробки подій та інтегрувати зовнішні API без прямого програмування [3]. Обидві системи виступають основою для інтеграції аналітичних моделей, у тому числі тих, що застосовуються при аналізі кредитоспроможності та розгляді заявок клієнтів.

Відтак, цифрова інфраструктура включає також аналітичний центр, що здійснює оцінку кредитного ризику, прогнозування потенційних збитків та формування резервів. У результаті моделювання визначаються ключові параметри: ймовірність порушення зобов'язань, очікувана сума заборгованості на момент дефолту та можливий рівень втрат з урахуванням заставного забезпечення. Моделі адаптуються до динаміки середовища, поєднуючи історичні дані з поточними спостереженнями та прогнозними

макроекономічними показниками. Отримані оцінки узгоджуються з внутрішньою стратегією управління ризиками, з урахуванням ризик-апетиту, нормативних вимог і внутрішніх обмежень [32, с. 82].

Кредитна політика банку в частині оцінки ризику зорієнтована на автоматизоване ухвалення рішень і зменшення впливу суб'єктивного чинника через цифрові алгоритми. У цьому контексті важливе значення мають принципи прозорості, послідовності та зрозумілості. Прозорість передбачає відкритість методів управління ризиком та можливість перевірки рішень. Послідовність гарантує єдність підходів на всіх рівнях прийняття рішень, збереження стабільності оцінок та їх відповідність встановленим критеріям. Принцип зрозумілості означає доступність логіки аналізу та формування рішень для всіх учасників процесу — від розробників моделей до фахівців, які працюють з клієнтами [32, с. 84]. Разом ці засади повинні забезпечувати об'єктивність і передбачуваність оцінювання кредитного ризику, що є основою сталого функціонування цифрової моделі прийняття рішень.

Система оцінки кредитного ризику в АТ «А-БАНК» ґрунтується на комплексі методів, що охоплюють як індивідуальний аналіз позичальника, так і агреговану оцінку портфеля в цілому. Для кредитів з однорідними характеристиками застосовуються групові підходи, які дозволяють оцінювати ризик за єдиними правилами. Водночас для великих або нестандартних позицій використовується індивідуальний підхід, що передбачає розрахунок резервів відповідно до чинних регуляторних вимог.

Оцінювання охоплює як очікувані, так і неочікувані кредитні збитки, з урахуванням характеристик боржника, виду кредитного продукту, а також макроекономічного контексту. Для додаткової перевірки надійності модельних оцінок проводиться регулярне стрес-тестування портфеля, що дає змогу прогнозувати потенційні втрати за негативних сценаріїв. Окремо аналізується концентрація ризиків за галузями, регіонами та видами забезпечення, а також здійснюється контроль за якістю заставного майна.

При цьому використовуються три основні параметри ризику:

- PD (Probability of Default) — імовірність порушення зобов'язань за кредитом;
- LGD (Loss Given Default) — частка втрат у разі дефолту, з урахуванням застави та витрат;
- EAD (Exposure at Default) — очікувана сума заборгованості на момент дефолту [32, с. 33].

У межах цифрового середовища банку реалізовано декілька типів моделей, орієнтованих на підтримку процесу ухвалення рішень:

- заявочні моделі — прогнозують імовірність прострочення на етапі подання заявки;
- поведінкові моделі — аналізують динаміку клієнта під час користування продуктом (наприклад, прогнозують прострочення на 90+ днів або суму використаного кредитного ліміту);
- скорингові моделі — оцінюють загальний рівень ризику на основі статистики клієнта, операцій та макропараметрів.

Серед прикладних моделей, що використовуються в процесі оцінювання кредитного ризику, виокремлюються такі:

- прогноз суми непідтвердженого доходу (заявочна та поведінкова моделі);
- модель кредитів готівкою (заявочна): оцінювання ймовірності прострочення окремо для зовнішніх і внутрішніх клієнтів;
- модель кредитних карток (поведінкова): прогноз виходу на прострочення;
- модель товарного кредиту (поведінкова) — «Розстрочка миттєва», «Плати частинами», «Кредит на товар»: прогноз прострочення на 12 місяців;
- модель кредитів готівкою (поведінкова): оцінка ймовірності прострочення 90+ днів;
- модель прогнозу рівня використання кредитного ліміту — EAD (поведінкова);

- скорингова модель: оцінка ймовірності дефолту (PD) з урахуванням макроекономічних змін;
- модель матриці міграції: розрахунок показників ймовірності дефолту та очікуваного рівня втрат (LGD) на 12 місяців і на весь строк дії зобов'язань [32, с. 86].

У такий спосіб завершується повний цикл перетворення даних — від збору до аналітичної підтримки кредитного процесу.

1.5 Аналіз фінансових показників кредитної активності банку

Як було вже з'ясовано в попередніх розділах, процес кредитування в банку є багаторівневою структурою, що охоплює як формування загальної політики, так і ухвалення конкретних рішень щодо кредитного ліміту для окремих клієнтів. Він поєднує цифрові інструменти з аналітичною роботою спеціалізованих підрозділів і відповідно передбачає постійне узгодження між стратегічними пріоритетами банку й оперативною роботою з реальними запитами позичальників, що формує складну систему прийняття рішень. Враховуючи масштаб і складність цієї системи, доцільним є аналіз фінансових показників, які відображають фактичні результати реалізації кредитної стратегії в динаміці.

Загальний огляд кредитної активності банку подано в таблиці 1.1, на якій відображено динаміку основних фінансових показників, пов'язаних із кредитуванням, за період з 2020 по 2024 рік.

Таблиця 1.1

Кредитні показники зі звіту про фінансовий стан на 31 грудня 2020–2024 років, тис. грн

Показник	2020	2021	2022	2023	2024
Кредити та заборгованість клієнтів	6 931 104	8 983 168	5 516 412	8 567 591	15 828 835
Процентні доходи (всього)	2 632 198	3 185 635	3 474 153	4 143 331	6 198 346
Доходи від кредитування	2 564 771	3 039 835	2 697 891	2 222 293	4 766 371

У таблиці 1.1 простежується нерівномірна, але загалом зростальна динаміка поданих фінансових величин. Обсяг кредитів та заборгованості клієнтів протягом 2020–2024 років зазнавав коливань: після зростання у 2021 році, у 2022-му спостерігалось помітне зниження майже на 39%, що може свідчити про обмеження обсягів кредитування. Подібна динаміка характерна і для сукупних процентних доходів: протягом трьох років темпи зростання залишались помірними, однак у 2023 році вони сповільнилися ще сильніше. Доходи безпосередньо від кредитування також у 2023 році знизилися більш ніж на 17% порівняно з попереднім роком [32, с. 3-4, 55]. Хоча у 2024 році всі три показники продемонстрували значне покращення, зафіксовані раніше зниження можуть відображати вплив економічної нестабільності.

З огляду на такі коливання, доцільно звернутися до внутрішньої структури кредитного портфеля за категоріями протягом того ж періоду часу. У таблиці 1.2 представлено розподіл валової балансової вартості кредитів за категоріями продуктів для фізичних осіб станом на 31 грудня 2020–2024 років.

Таблиця 1.2

Показники категорій кредитів зі звіту про кредити та заборгованість банків та клієнтів на 31 грудня 2020–2024 років, тис. грн

Категорія кредиту	2020	2021	2022	2023	2024
Кредитні картки	4 206 982	5 154 549	5 183 273	6 507 005	10 466 583
Споживчі кредити фізичним особам	2 877 784	3 330 603	2 282 935	2 622 086	5 142 317
Автокредитування фізичних осіб	316 877	212 109	148 705	249 996	416 086
Фінансовий лізинг фізичним особам	201 866	444 760	323 821	198 081	123 386
Іпотечне кредитування	84 045	71 771	100 853	104 266	116 136
Інші кредити фізичним особам	214 260	557 370	698 924	912 208	1 097 891

Таким чином, таблиця 1.2 демонструє розподіл популярності різних кредитних продуктів для фізичних осіб. Найбільші обсяги припадають на

кредитні картки та споживче кредитування, тоді як інші види — автокредити, фінансовий лізинг і іпотека — залишаються менш поширеними. Щодо динаміки змін, як і у випадку з узагальненими показниками у таблиці 5.1, у 2022–2023 роках спостерігалися помітні коливання. У ряді категорій, таких як споживче кредитування, фінансовий лізинг та автокредитування, фіксувалися спади: обсяги цих продуктів зменшилися приблизно на 30% у 2022 році порівняно з попереднім роком. Кредитні картки не зазнали зниження, проте розвиток, характерний для попередніх років, у цей період тимчасово уповільнився — темп приросту у 2022 році становив лише 0,6%, порівняно з 22% роком раніше [32, с. 62].

Ураховуючи такі зміни в 2022-2023 роках, банк мав заздалегідь визначити прийнятний рівень втрат у кредитній політиці, до якого готовий у межах своєї стратегії — так званий ризик-апетит. Відповідно, у таблиці 1.3 представлено співвідношення затверджених і фактичних значень ризик-апетиту в розрізі останніх років.

Таблиця 1.3

Затверджений і фактичний ризик-апетит станом на 1 січня 2022–2025 років по портфелю за кредитними продуктами, тис. грн

Рік	Затверджений ризик-апетит	Фактичний ризик-апетит
2022	565 000	325 227
2023	1 787 800	1 275 960
2024	785 600	249 100
2025	636 300	535 400

Аналіз таблиці 1.3 свідчить про стабільну тенденцію дотримання банком меж визначеного ризик-апетиту протягом 2022–2025 років. У 2022 році (дані на 01.01.2023) фактичні втрати становили близько 58% від затвердженого рівня, у 2023 — 71%, а в 2024 — лише 32%, що свідчить про посилення контролю над ризиками. Натомість за 2024 рік зафіксовано підвищення до 84% від встановленого ліміту, що при менш амбітному плановому значенні може свідчити про більший вплив зовнішніх чинників [32, с. 86].

Узагальнюючи результати аналізу, можна сказати, що, як і передбачалося, у 2022–2023 роках мала місце певна нестабільність у динаміці кредитної діяльності, яка згодом була цілком врегульована. Поступове покращення фінансових результатів не лише відновило позиції банку, а й забезпечило перевищення рівнів, зафіксованих до періоду спаду — зокрема, станом на 2024 рік ключові показники перевищили рівень 2021 року.

1.6 Стратегія розвитку банку в сфері споживчого кредитування

На тлі раніше проаналізованих фінансових даних, що відображають зміну кредитної активності банку протягом останніх років, варто звернути увагу на стратегічні орієнтири його подальшого розвитку. Це дає можливість не лише розуміти загальну траєкторію дій банку, а й зіставити зафіксовані кількісні результати з напрямками, які визначаються як пріоритетні в межах кредитної стратегії.

У межах таких пріоритетів банк формулює чітку генеральну мету, що полягає у досягненні 1 300 000 активних задоволених клієнтів [34, с. 13]. Цей орієнтир закладено в офіційному плані розвитку та відображено на рисунку 1.4, де наведено прогресивне зростання клієнтської бази у період до 2025 року.



Рисунок 1.4 – Графіка зміни запланованої активної клієнтської бази АТ «А-БАНК»

Відповідно до цієї цілі, банк розробив бізнес-модель розширення, яка охоплює як механізми роботи з фізичними особами у межах побудови довгострокових відносин. Потенційні клієнти залучаються за допомогою фізичних відділень, цифрової реклами, партнерських мереж та реферальної програми. Паралельно з цим утримання нової та чинної бази ґрунтується на

чотирьох ключових компонентах: депозитних продуктах, комунікації, а також кредитних програмах «Швидка готівка» та «Плати частинами». Такий акцент на споживчих кредитах свідчить про їх вагомую роль у подальшому розвитку банку, що підтверджується планами суттєвого нарощування обсягів відповідних сегментів, представленими в таблиці 1.4.

Таблиця 1.4

Заплановане розширення кредитного балансу на 2026 рік, млн грн

Категорія кредиту	Факт на 01.01.25	План на 01.01.26	Приріст, %
Кредитні картки	10 200	13 000	+27,5%
«Швидка готівка»	2 541	3 599	+42%
«Розстрочка на витрати»	1 385	1 626	+17,5%
«А-Кредит» («Розстрочка», «Плати частинами»)	983	1 275 760	+30%
Фінансовий лізинг фізичним особам	537	537	-

Як видно з таблиці 1.4, дійсно найвищі темпи зростання прогнозуються для програми «Швидка готівка» та продукту «Плати частинами» (у складі програми «А-Кредит»), які мають приріст у 42% та 30% відповідно. Також популярність кредитного ліміту на кредитних картках має зрости на 27,5% [32, с. 17]. Це дещо менше, ніж у згаданих лідерів, однак з огляду на те, що ця категорія залишається найпопулярнішим кредитним продуктом банку, таке зростання є суттєвим і за обсягом, і за впливом на загальний баланс.

Роблячи висновок із представленого плану розвитку, можна стверджувати, що саме кредитні продукти формують одну з найвагоміших складових загальної стратегії банку. Акцент на їх активне розширення означає не лише потенційне зростання прибутковості, але й суттєве збільшення обсягів коштів, залучених у зону ризику. За таких умов підвищуються вимоги до системи управління ризиками, яка має адаптуватися до зростаючого навантаження як у плані обробки запитів, так і в забезпеченні якості кредитних рішень.

1.7. Недоліки чинної кредитної моделі та напрямки її покращення

Кредитування фізичних осіб займає важливе місце у спектрі банківських послуг Абанку, забезпечуючи фундаментальну підтримку у вирішенні фінансових потреб клієнтів. Згідно з підрозділом 1.6 Стратегія розвитку банку в сфері споживчого кредитування, цей напрям залишається одним із ключових у контексті реалізації стратегії розвитку банку. Однак ефективність даного сегмента в окремі періоди виявлялася зниженою — що підтверджується як фінансовими показниками, наведеними у підрозділі 1.5 Аналіз фінансових показників кредитної активності банку, так і висновками кредитного агентства (див. підрозділ 1.1 Історія та еволюція Абанку). Така ситуація актуалізує необхідність глибшої роботи з ризиками, адже при намірах масштабного розширення саме слабкі місця в механізмах оцінки кредитоспроможності можуть становити загрозу для фінансової стабільності банку.

У таблиці 1.5, що знаходиться нижче, представлена динаміка основних кредитних параметрів, згідно з офіційною статистикою Національного банку України для АТ «А-БАНК» [27].

Таблиця 1.5

Розподіл кредитів у тисячах гривень, наданих фізичним особам у національній та іноземній валютах, та розміру кредитного ризику за класами боржника відповідно до Положення № 351 станом на 01.06.2020 – 01.06.2024

Рік	Назва показника	Клас боржника фізичної особи				
		1	2	3	4	5
-1-	-2-	-3-	-4-	-5-	-6-	-7-
2020	Сума кредитної заборгованості	4 812 832	54 248	70 524	89 044	782 073
	Кредитний ризик	229 586	13 753	26 636	56 317	682 496
2021	Сума кредитної заборгованості	7 174 498	41 448	130 748	73 876	919 794
	Кредитний ризик	311 895	9 864	51 202	46 564	804 019

Продовження табл. 1.5

-1-	-2-	-3-	-4-	-5-	-6-	-7-
2022	Сума кредитної заборгованості	8 469 070	182 159	116 423	89 335	954 780
	Кредитний ризик	379 032	41 508	45 907	56 217	829 401
2023	Сума кредитної заборгованості	5 163 215	30 100	97 207	73 102	3 840 595
	Кредитний ризик	298 189	7 413	43 790	46 582	3 312 237
2024	Сума кредитної заборгованості	10 773 272	57 562	165 775	123 989	1 102 008
	Кредитний ризик	417 592	15 093	66 628	77 820	965 358

Таким чином, аналіз фінансових звітів за останні роки показує зміну у частці кредитів, що припадає на клієнтів з критичним фінансовим станом, а саме тих, хто належить до 5-го класу ризику. Клієнти цієї категорії характеризуються такими критеріями, як сукупний обсяг боргового навантаження, що значно перевищує їх чистий дохід, тривалі прострочення платежів та низькі показники платоспроможності. Наприклад, у 2020 році ця категорія становила 13,46% від загальної суми кредитної заборгованості, частка знизилась до 9,73% у 2022 році, але різко зросла до 41,73% у 2023 році, що може вказувати на збільшення кількості клієнтів, які опинилися у фінансовій скруті. Зростання частки боржників з високим класом ризику не обов'язково свідчить про те, що банк змінив власну політику видачі кредитів і почав обирати менш надійних позичальників. Ймовірніше, така зміна є наслідком загального погіршення фінансової ситуації серед населення. Подібні зрушення не можуть залишатися лише на розгляд цифрових моделей, що працюють за послідовним алгоритмом — вони потребують регулярного аналізу з боку кредитних комітетів і відповідальних підрозділів. Саме ці органи мають брати до уваги подібні тенденції як основу для адаптації кредитної політики до змін зовнішнього середовища.

Фінансові дані, відображені на рис. 1.5, демонструють, що у 2022 році зменшилась частка непрострочених кредитів, натомість зросла частка прострочень у межах від 90 до 179 та до 360 днів [35, с. 63-64]. У 2023 році

загальна ситуація з непростроченістю покращилась, однак одночасно спостерігалось значне зростання обсягу кредитів із простроченням понад 360 днів. Це може свідчити про те, що частина проблемних кредитів, які вже у 2022 році мали ознаки ризику, в наступному році перейшла в найважчу категорію прострочення. Відповідно, рішення, прийняті у 2022 році або наприкінці 2021 року, виявилися недостатньо адаптованими до економічних зрушень. Така ситуація вказує на необхідність застосування більш гнучких моделей оцінювання ризику.

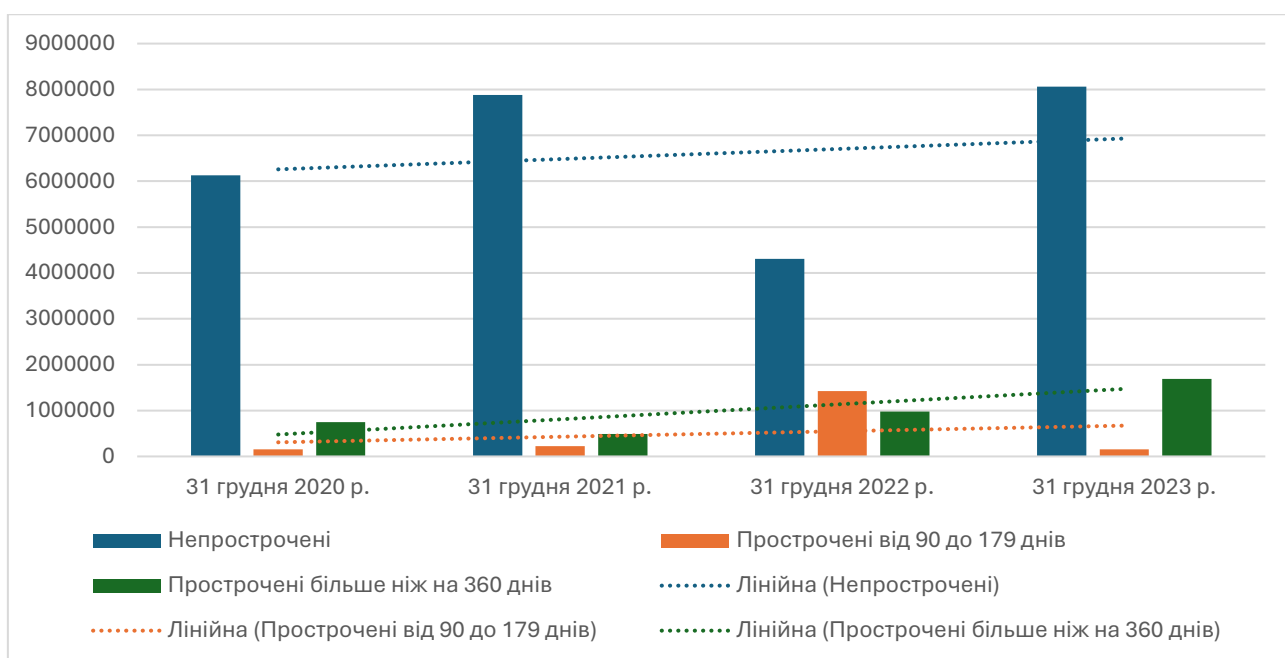


Рисунок 1.5 – Кредити, видані клієнтам фізичним особам, які оцінюються за амортизованою вартістю у тисячах гривень

З огляду на стратегію розвитку банку, сформульовану у підрозділі 1.6 Стратегія розвитку банку в сфері споживчого кредитування, яка передбачає активне розширення клієнтської бази та зростання частки задоволених користувачів, важливо враховувати також рівень клієнтського сприйняття окремих елементів банківських послуг. Детальний аналіз відгуків клієнтів на платформах Google Play та App Store за період з 1 березня 2024 до 1 березня 2025 року виявив, що трохи менше 5% висловили невдоволення саме системою визначення кредитного ліміту. Ці відгуки, хоч і становлять малу частку загальної

кількості, мають особливе значення, оскільки зосереджені виключно на проблемах з кредитуванням, тоді як скарги на інші аспекти обслуговування карток чи депозитних програм мінімальні (менше 0,5%) або ж відсутні. Це свідчить не стільки про потенційні слабкі місця самої моделі визначення кредитного ризику, скільки про обмежену її зрозумілість для користувача. Відсутність чіткого пояснення причин прийняття рішень щодо кредитування може знижувати довіру навіть до коректно працюючих моделей, що суперечить цілі банку забезпечити прозору систему взаємодії.

Зважаючи на результати проведеного дослідження, можна визначити кілька ключових напрямів удосконалення поточної моделі ухвалення кредитних рішень. По-перше, в умовах суттєвих змін фінансового стану клієнтів модель має бути достатньо зрозумілою та інтерпретованою для фахівців відповідного департаменту — це критично для забезпечення оперативного й обґрунтованого реагування на нові ризики. По-друге, вона повинна бути точною у періоди нестабільності, коли стандартні шаблони поведінки клієнтів втрачають передбачуваність. І, нарешті, модель має містити елементи рекомендаційного характеру для кінцевого користувача — клієнта банку, що дозволить не лише підвищити прозорість процесу, але й зміцнити довіру до системи кредитування загалом.

1.8 Висновок до розділу 1

У першому розділі було здійснено комплексний аналіз кредитної діяльності АТ «А-БАНК» з урахуванням організаційної структури, наявного продуктового портфеля, цифрової інфраструктури та ключових фінансових показників. Послідовне розкриття історії розвитку банку, його функціональних підрозділів і моделей оцінювання кредитоспроможності дозволило простежити логіку прийняття рішень у межах чинної кредитної політики.

На основі проведеного аналізу можна стверджувати, що сучасна система кредитування в банку є результатом взаємодії двох взаємопов'язаних рівнів: з

одного боку, стратегічних рішень профільних підрозділів і колегіальних органів, відповідальних за визначення підходів до кредитування, а з іншого — автоматизованих цифрових механізмів, які забезпечують аналіз даних і ухвалення рішень щодо конкретних клієнтів. Такий формат функціонування створює складну, але послідовну структуру, де скорингові, поведінкові та заявочні моделі виступають інструментами реалізації загальної кредитної стратегії.

Також огляд фінансових показників за останні роки виявив як позитивну динаміку зростання обсягів кредитування, так і періоди суттєвих коливань, зокрема зростання частки простроченої заборгованості у 2022–2023 роках та збільшення кількості клієнтів із критичним рівнем ризику. Подані тенденції вказують на обмежену точність окремих моделей у періоди нестабільності, коли поведінка позичальників не відповідає звичним патернам. Крім того, не менш помітною є проблема прозорості: як засвідчив аналіз відгуків клієнтів, саме складність у розумінні причин відмови або обмеженого кредитного ліміту може знижувати рівень довіри навіть до технічно коректної системи.

Отже, подальший розвиток моделі ухвалення кредитних рішень має орієнтуватися на підвищення точності в умовах нестабільності, а також інтерпретованість і зрозумілість для внутрішніх фахівців і клієнтів.

2 СПЕЦІАЛЬНИЙ РОЗДІЛ

2.1 Змістовне формулювання задачі

У процесі кредитування фізичних осіб банк постійно стикається з потребою оцінки надійності клієнта ще до прийняття рішення про фінансування. Йдеться про так званий кредитний ризик — наявний або потенційний ризик для надходжень і капіталу, який виникає через неспроможність сторони, що взяла на себе зобов'язання, виконати умови будь-якої фінансової угоди із банком. Цей ризик є домінуючим елементом у системі банківських ризиків і виступає центральним фактором при ухваленні рішень про кредитування [19, 30].

Як уже було частково розглянуто в підрозділі 1.4 Інформаційне забезпечення процесу оцінки кредитоспроможності клієнтів, його оцінка здійснюється на підставі низки вхідних параметрів, які відображають фінансову поведінку клієнта. На момент звернення до банку позичальник надає основну інформацію про себе: вік, дохід, тип зайнятості, а також дозволяє доступ до кредитної історії. Завдання банку полягає в тому, щоб на основі цих даних спрогнозувати, чи виконає клієнт свої фінансові зобов'язання — тобто, чи буде вчасно обслуговувати кредит без прострочень і порушень умов договору [7, 30].

Оцінка кредитного ризику — це прогноз негативного сценарію в майбутньому на основі доступної на сьогодні інформації. У реальних умовах негативним вважається не лише повний дефолт, а й систематичні затримки платежів, використання кредитних коштів у нестабільному або граничному режимі, високий рівень заборгованості або непрогнозована фінансова поведінка [19, 33, с. 33]. Водночас, банківські установи зацікавлені у збереженні обсягу виданих кредитів і прибутковості, тому модель оцінки повинна забезпечувати точне розмежування між надійними й ризиковими клієнтами, не знижуючи рівень кредитування без потреби.

У змістовному сенсі задача полягає в тому, щоб автоматично класифікувати нового клієнта на основі його характеристик як такого, що має високу або низьку ймовірність дефолту. Це завдання не зводиться до аналізу одного чи двох факторів, на рішення можуть впливати десятки різних чинників, зокрема:

- загальні соціально-економічні характеристики, що допомагають оцінити стабільність клієнта;
- особливості формування й розвитку кредитної історії, тривалість та активність взаємодії з банківськими продуктами;
- прояви фінансової дисципліни — наявність затримок, регулярність погашення зобов'язань, характер попередньої поведінки;
- ступінь зацікавленості у кредитуванні та інтенсивність звернень до фінансових установ;
- навантаження, яке клієнт вже має, та загальний стан його поточних фінансових зобов'язань. [33, с. 56]

Усе це формує багатовимірний профіль клієнта, який потрібно зіставити з попереднім досвідом фінансової установи — прикладами клієнтів, які вже отримували кредити і мали позитивні або негативні результати. Саме на такому підході базується машинне навчання: воно дозволяє побудувати модель, яка, навчившись на історичних прикладах, буде здатна класифікувати нові випадки [5].

Відповідно, в межах цієї роботи відтворюється скорингова система оцінки ймовірності дефолту, яка була представлена в підрозділі 1.4 Інформаційне забезпечення процесу оцінки кредитоспроможності клієнтів і використовується банком для визначення рівня кредитного ризику. Для симуляції умов функціонування такої системи використовується набір даних споживчого кредитного ризику за поновлюваним лімітом оборотними кредитами, який став загальнодоступним завдяки компанії FICO — американському розробнику скорингових систем. Цей датасет був сформований як частина FICO Analytic Cloud, платформи для створення аналітичних моделей, а згодом у 2018 році використаний для міжнародного конкурсу Explainable Machine Learning

Challenge, організованого FICO у співпраці з Google та провідними університетами. Ініціатива була спрямована на розробку точних і пояснюваних моделей кредитного скорингу. Дані стосуються продукту HELOC (Home Equity Line of Credit) — кредитної лінії, що видається під заставу нерухомості, і передбачає багаторазове використання коштів у межах встановленого ліміту [10].

Масив даних є реальним, але анонімізованим: персональні дані клієнтів вилучено, натомість збережено всі фінансові характеристики, які зазвичай використовуються у процесі оцінки кредитоспроможності. У ньому представлено 5 844 записів, кожен з яких містить набір ознак, що характеризують поточний стан, кредитну історію та поведінкові шаблони клієнта. Цільова змінна Risk_Flag набуває значення 0, якщо клієнт допустив дефолт, або 1, якщо зобов'язання виконувались вчасно [11]. Таким чином, задача має вигляд бінарної класифікації.

У моделі використовуються 27 змінних що описують соціально-економічні характеристики клієнта та стан його кредитного навантаження. Таблиця 2.1 містить перелік ознак, що використовуються як вхідні дані для моделі, разом із короткими поясненнями.

Таблиця 2.1

Опис ознак, що використовуються у моделі

№	Назва ознаки	Характеристика ознаки
-1-	-2-	-3-
1	App_Month	Місяць подання кредитної заявки
2	App_Day	День подання кредитної заявки
3	Age	Вік клієнта на момент подання заявки
4	Income	Щомісячний дохід
5	Region	Регіон проживання клієнта
6	Debt_Ratio	Співвідношення заборгованості до доходу
7	CB_Age_Oldest_TL	Кількість місяців з моменту відкриття найстарішого кредиту
8	CB_Age_Newest_TL	Кількість місяців з моменту відкриття найновішого кредиту
9	CB_Avg_Mos_File	Середній вік кредитної історії
10	CB_Nb_Sat_TL	Кількість задовільних кредитів
11	CB_Nb_60_Plus_TL	Кількість кредитів із простроченням понад 60 днів
12	CB_Nb_90_Plus_TL	Кількість кредитів із простроченням понад 90 днів
13	CB_Pct_Sat_TL	Частка задовільних кредитів серед усіх
14	CB_Mos_Since_Dlq	Кількість місяців з моменту останньої прострочки

Продовження табл. 2.1

-1-	-2-	-3-
15	CB_Max_Dlq_12_Mos	Найбільша прострочка за останні 12 місяців
16	CB_Max_Dlq_Ever	Найбільша прострочка за весь період кредитної історії
17	CB_Nb_Total_TL	Загальна кількість кредитних ліній
18	CB_Nb_TL_Open_12	Кількість кредитів, відкритих за останні 12 місяців
19	CB_Pct_IL_TL	Частка інсталяційних (з фіксованими платежами) кредитів
20	CB_Nb_Inq_6_Mos	Кількість кредитних запитів за останні 6 місяців
21	CB_Nb_Inq_6_Mos_excl_7_Days	Те саме, без урахування запитів за останні 7 днів
22	CB_Rev_Util	Рівень використання револьверних (поновлюваних) кредитів
23	CB_IL_Util	Рівень використання інсталяційних (з фіксованими платежами) кредитів
24	CB_Nb_Rev_TL_w_Bal	Кількість револьверних (поновлюваних) кредитів з активним балансом
25	CB_Nb_IL_TL_w_Bal	Кількість інсталяційних (з фіксованими платежами) кредитів з активним балансом
26	CB_Nb_Rev_TL_75_Pct_Limit	Кількість револьверних (поновлюваних) кредитів, використаних понад 75% ліміту
27	CB_Pct_TL_w_Bal	Частка кредитів, на яких є залишок боргу

2.2 Математична формалізація задачі класифікації

Задача прогнозування кредитного ризику позичальника може бути формалізована як задача двокласової класифікації, у якій кожен клієнт описується множиною числових і категоріальних ознак, а мета полягає в передбаченні ймовірності виникнення негативної події — порушення зобов'язань за кредитом.

Нехай кожен об'єкт (клієнт) представлений вектором вхідних даних:

$$X = (x_1, x_2, \dots, x_n),$$

де x_1 — значення i -ої ознаки, що описує фінансовий стан, кредитну історію та інші релевантні характеристики позичальника. У рамках даної роботи $n = 27$ ознак.

Цільова змінна позначається як

$$Y \in \{0, 1\},$$

де 0 — клієнт допустив дефолт, 1 — клієнт обслуговував зобов'язання регулярно.

Таким чином, постановка задачі полягає у побудові функції класифікації:

$$f(x) \rightarrow Y,$$

де на основі вхідного вектору ознак X дозволяє передбачити клас об'єкта — тобто, оцінити ймовірність належності нового клієнта до групи надійних або ризикованих позичальників. Функція f описує модель машинного навчання, яка відображає простір ознак у простір класів, і формується на основі навчальної вибірки, де кожен приклад містить значення цільової змінної та відповідний вектор ознак. Модель навчається шляхом узагальнення наявного історичного досвіду [15, с. 4].

У результаті вирішення задачі формується модель, яку можливо використовувати як елемент автоматизованої системи підтримки рішень у процесі кредитного скорингу.

2.3 Вибір методу та його обґрунтування

У попередньому підрозділі було визначено, що подана задача прогнозування кредитного ризику належить до задач бінарної класифікації. Такі задачі є типовими для розділу машинного навчання.

Машинне навчання — це підхід, за якого модель будується на основі прикладів із відомим результатом. Його основна ідея полягає в тому, щоб автоматично знаходити залежності між вхідними ознаками й цільовим значенням. Під час навчання алгоритм багаторазово проходить крізь дані, порівнює передбачення з правильними відповідями та поступово змінює свої параметри, щоб зменшити кількість помилок. У результаті формується функція, яка дозволяє застосовувати узагальнені закономірності до нових прикладів. Такий підхід дає змогу працювати з великою кількістю змінних, не вимагає ручного формулювання правил і дозволяє моделювати як прості, так і складні залежності.

Серед найпоширеніших методів, що застосовуються для задач кредитного скорингу, можна виділити наступні:

- Логістична регресія — лінійний метод класифікації, який оцінює вплив кожної ознаки через ваговий коефіцієнт і використовує сигмоїдну функцію для переведення результату у ймовірність належності до одного з класів. Під час навчання модель оптимізує коефіцієнти для кращого відображення залежностей між ознаками та класом.
- Дерево рішень — алгоритм, що будує ієрархію умов на основі значень ознак. На кожному етапі дані розбиваються за критерієм, який найкраще поділяє спостереження на більш однорідні підмножини. В результаті формується структура з розгалуженням, у якій кожен вузол відповідає перевірці певної ознаки, а лист — передбаченому класу.
- Випадковий ліс — ансамблевий метод, який поєднує велику кількість незалежних дерев рішень. Кожне дерево навчається на випадковій підмножині даних і ознак. Остаточне рішення приймається шляхом голосування: клас, що отримав більшість голосів від окремих дерев, вважається результатом моделі.
- Градієнтний бустинг — послідовний ансамблевий метод, у якому кожне нове дерево навчається на залишкових помилках попередніх. Модель додає слабкі алгоритми один за одним, коригуючи похибки і наближаючись до мінімуму функції втрат.
- Метод k найближчих сусідів (k -NN) — алгоритм прямого порівняння, який класифікує об'єкт, орієнтуючись на k найбільш схожих прикладів із навчальної вибірки. Визначення «схожості» базується на обраній метриці відстані, зазвичай евклідовій. Клас нового об'єкта встановлюється як найчастіший серед обраних сусідів [1].

Кожен із зазначених методів має власні принципи роботи та потенційні переваги, проте їх практична ефективність визначається специфікою даних і вимогами, що висуває задача.

Відповідно, у задачах кредитного скорингу використовуються численні ознаки різного типу, зокрема числові, категоріальні та похідні показники, які мають складні взаємозв'язки та різну прогностичну значущість. Значна частина ознак не має стабільного або лінійного впливу на цільову змінну: деякі характеристики виявляють свою інформативність лише в комбінації з іншими. У таких умовах методи, що базуються на припущенні про незалежний та адитивний ефект кожної ознаки, зокрема логістична регресія, виявляються недостатньо гнучкими для адекватного моделювання залежностей. Їх структура не дозволяє ефективно враховувати взаємодії між змінними без явного введення додаткових параметрів [14].

Інша складність пов'язана з розмірністю простору ознак. При наявності 27 вхідних параметрів зростає ризик виникнення ефекту розрідженості даних, що ускладнює застосування методів на основі метрик відстані, зокрема алгоритму k -найближчих сусідів. У високовимірному просторі об'єкти втрачають подібність за евклідовою чи іншими класичними метриками, що призводить до нестабільної поведінки моделі [6]. Крім того, метод k -NN не забезпечує механізму інтерпретації рішення: модель не надає інформації про вплив окремих змінних на класифікацію, що суперечить вимогам до пояснюваності у фінансових задачах.

Ансамблеві підходи, зокрема випадковий ліс та градієнтний бустинг, демонструють високу прогностичну здатність у задачах з нерівномірно розподіленими ознаками, слабкою кореляцією та перехрещеними класами. Вони здатні адаптуватися до складної структури даних, автоматично враховуючи взаємозв'язки між змінними. Проте ці методи характеризуються складною архітектурою, що включає велику кількість дерев і параметрів [2]. Це унеможлиблює точне відтворення логіки класифікації для окремих прикладів, що є критичним недоліком у задачах, де потрібна формалізована інтерпретація результату для користувача, аудитора або регуляторного органу.

У цьому контексті дерево рішень виступає як конструктивно виправданий підхід, що поєднує здатність моделювати нелінійні залежності з високим рівнем

інтерпретованості. Алгоритм забезпечує автоматичне формування послідовності логічних умов, за якими здійснюється класифікація. Кожен етап прийняття рішення відображено у структурі дерева, що дозволяє чітко відстежити вплив конкретних ознак на результат. Крім того, дерево рішень є універсальним щодо типу вхідних змінних і не вимагає складної попередньої обробки, що робить його придатним для застосування в умовах, де одночасно важливі точність, адаптивність та прозорість моделі.

Таким чином, для подальшої реалізації задачі обґрунтовано обрано метод дерева рішень як такий, що найкраще відповідає характеру даних та вимогам до пояснюваності рішень у кредитному скорингу.

2.4 Теоретичні основи алгоритму

Алгоритм дерева рішень належить до ієрархічних методів класифікації, у яких процес прийняття рішення моделюється як послідовність перевірок певних умов над вхідними ознаками. Модель має структуру дерева: кожен внутрішній вузол відповідає за перевірку умови на одну з ознак, кожна гілка — варіант виконання цієї умови, а листові вершини — результати класифікації. Починаючи з кореневого вузла, кожен об'єкт послідовно рухається гілками дерева, доки не досягне листа. Таким чином, класифікаційне рішення для нового прикладу визначається набором простих умов, сформованих у процесі навчання.

Формування дерева здійснюється рекурсивно. На кожному кроці вибірка розбивається на дві підмножини за значенням певної ознаки та відповідного порогового значення. Оптимальність розбиття визначається за приростом інформації — тобто зменшенням невизначеності щодо класів у результаті поділу. Для цього обчислюється різниця між ентропією вихідної множини та середньозваженою ентропією отриманих підмножин. Чим вищий приріст, тим кращим вважається розбиття.

Ентропія початкової множини S визначається як:

$$H(S) = - \sum_{i=1}^k p_i \log_2(p_i), \quad (2.1)$$

де p_i — частка об'єктів класу i у множині S , а k — кількість класів. У задачі бінарної класифікації $k = 2$.

При розбитті множини S за ознакою A і порогом v , обчислюється зважене значення ентропії двох отриманих підмножин:

$$H_{\text{split}}(S, A, v) = \frac{|S_L|}{|S|} H(S_L) + \frac{|S_R|}{|S|} H(S_R), \quad (2.2)$$

де S_L — підмножина об'єктів, для яких $A \leq v$, а S_R — тих, де $A > v$.

Інформаційний приріст визначається як різниця між ентропією початкової множини та зваженою сумою ентропій у підмножинах:

$$IG(S, A, v) = H(S) - H_{\text{split}}(S, A, v), \quad (2.3)$$

Ознака A та поріг v , для яких значення IG є максимальним, вважаються найкращими на поточному кроці побудови дерева [8].

Рекурсивне розбиття продовжується доти, доки не виконується одна з умов завершення:

- усі об'єкти в підмножині належать до одного класу;
- досягнуто заданої максимальної глибини дерева;
- подальший поділ не забезпечує зменшення ентропії.

У листках дерева фіксується рішення — клас, до якого віднесено об'єкти, що потрапили до відповідної гілки. Отже, структура дерева кодує послідовність умов, за якими здійснюється класифікація нових прикладів.

З технічної точки зору побудова дерева рішень передбачає послідовне виконання таких етапів:

1. Обчислення ентропії початкової вибірки як міри невизначеності.
2. Перебір можливих ознак і порогів розбиття та обчислення інформаційного приросту для кожного варіанту.
3. Вибір найкращого розбиття, що максимізує інформаційний приріст.

4. Розбиття вибірки на підмножини та рекурсивне повторення процесу для кожної з них.
5. Завершення побудови, коли досягнуто умови зупинки, та фіксація класу в листковому вузлі.

Окрім формування повної структури дерева, іноді застосовують додаткову процедуру спрощення — обрізання дерева після побудови (post-pruning). Його мета полягає у видаленні надлишкових розгалужень, що не покращують якість класифікації, і, відповідно, у підвищенні здатності моделі до узагальнення. У цьому процесі використовують валідаційну вибірку — окрему підмножину даних, яка не залучається до навчання і відрізняється від фінального тестового набору. Її призначення полягає в об'єктивному контролі якості спрощення дерева, що дає змогу уникнути перенавчання та зберегти релевантність прийнятих змін.

Обрізання здійснюється за наступним принципом:

- на вхід подається валідаційна вибірка;
- для кожного внутрішнього вузла, в якого обидві гілки ведуть до листів, перевіряється можливість його заміни на єдиний листовий вузол;
- якість дерева після такої заміни порівнюється з початковим варіантом за допомогою контрольної метрики (наприклад, точності на валідаційній множині);
- якщо спрощена структура демонструє не гірші результати, зміна вважається доцільною й фіксується [17].

У підсумку, після побудови та можливого спрощення структури, модель здатна приймати рішення для нових прикладів за простим принципом: перевірка значень ознак, послідовний рух по гілках дерева та визначення класу у листі.

2.5 Реалізація та структура моделі

Реалізація обраного методу класифікації — дерева рішень — здійснюється мовою програмування Python із застосуванням стандартних бібліотек для обробки даних і оцінювання результатів моделі. Відповідний код та його опис

представлено в додатку Г. Побудова структури виконується з урахуванням фінансових характеристик клієнтів, перелік яких наведено в підрозділі 2.1 Змістовне формулювання задачі. Перед використанням дані проходять очищення: відсутні значення замінюються, а рядки з пропущеними ознаками видаляються. Це забезпечує коректність розрахунків та уникає викривлень у структурі дерева під час подальшого моделювання.

Вся сукупність підготовлених даних розбивається на три частини — навчальну, валідаційну та тестову. Навчальна використовується для формування дерева рішень, на основі якої модель самостійно визначає правила класифікації. Валідаційна вибірка не бере участі в навчанні, але застосовується для проміжної оцінки якості моделі під час її спрощення. Тестова вибірка відіграє роль незалежної основи для фінальної перевірки точності побудованої моделі. Вона дає змогу об'єктивно оцінити якість передбачень на нових прикладах, які не входили до тренувального процесу.

Ключова логіка побудови дерева рішень полягає у послідовному створенні ієрархії умов, перевірок і відповідей для кожної підмножини даних. На кожному етапі модель працює з частиною вибірки, оцінює, чи варто продовжувати поділ, і, якщо умови зупинки виконано, формує остаточне рішення у вигляді листкового вузла. Щоб уникнути перенавчання, глибину дерева обмежено на рівні п'яти, що в цьому дослідженні використовується як базове значення для початкового аналізу.

Процес починається з перевірки: якщо всі об'єкти підмножини належать до одного класу або досягнуто граничну глибину, функція створює листковий вузол з фіксованим класом. Клас визначається як мода (найчастіший елемент) у поточній підмножині.

У випадку, коли поділ можливий, відбувається пошук найкращої ознаки та її порогового значення. У середині алгоритму перебираються всі ознаки з вихідного набору та всі унікальні значення в кожній з них. Для кожної комбінації обчислюється, наскільки ефективно відповідний поділ зменшує невизначеність у даних. Якщо поточний поділ дає більший приріст інформації, ніж попередні, він

зберігається як оптимальний. У результаті повертається така комбінація ознаки та порогового значення, яка забезпечує найкраще розділення. У подальшому ці значення використовуються для формування нового вузла дерева.

Коли найкращий поділ не вдається знайти (тобто жодна з ознак не дозволяє ефективно розділити вибірку), побудова дерева завершується на поточному вузлі, і створюється листковий елемент. У такому разі клас визначається як найчастіше зустрічаний у підмножині, а також зберігається ймовірнісний розподіл серед класів. Цей крок гарантує завершення побудови гілки в тих випадках, коли подальший поділ не має сенсу — зокрема, коли інформаційний приріст від усіх можливих розбиттів дорівнює нулю. Вузол повертається як остаточний — із фіксованим класом і збереженим розподілом ймовірностей.

Розрахунок інформаційного приросту відбувається за формулою (2.3), що використовує ентропійну оцінку як міру невизначеності. На цьому етапі підмножина ділиться на дві частини: до лівої входять ті об'єкти, значення ознаки яких не перевищує заданий поріг, до правої — ті, у яких воно більше. Обчислюється ентропія підмножини, яка ще не була поділена, а також — ентропія для кожної з частин після розбиття. Приріст інформації визначається як різниця між початковою ентропією і зваженим значенням ентропій (2.2) після поділу. Таким чином оцінюється, наскільки зменшилася невизначеність у результаті такого розбиття.

Обчислення ентропії виконується за формулою (2.1) і дає змогу оцінити рівень невизначеності в межах певної підмножини. У процесі цього обчислення аналізується розподіл класів, і як результат повертається числова оцінка невизначеності.

Отримані параметри поділу використовуються для розбиття даних на дві підмножини: до першої входять об'єкти, у яких значення вибраної ознаки не перевищує поріг, до другої — ті, у яких воно більше. На основі цього поділу створюється новий вузол дерева, у якому рекурсивно формуються ліве й праве піддерева. Відповідно, будівництво дерева відбувається наступним чином: на кожному рівні дерево ділиться, поки не буде досягнуто граничної глибини або всі

прикладі не належатимуть до одного класу. Кожен вузол зберігає інформацію про те, яку ознаку було обрано, який поріг поділу застосовано, які піддерева сформовано зліва і справа, а у випадку листкового вузла — який клас прогнозується і які ймовірності було зафіксовано для кожного класу [13].

Додатковим етапом оптимізації структури дерева є його обрізання дерева. Ідея такого кроку полягає в тому, щоб переглянути внутрішні вузли дерева, які мають двох нащадків — листки, і перевірити, чи не буде просте узагальнення (заміна піддерева одним класом) давати кращу або хоча б не гіршу якість на валідаційній вибірці. Для цього модель проходить дерево рекурсивно знизу вгору. Якщо поточний вузол має два листкових нащадки, обчислюється точність моделі на відповідній частині валідаційних даних до та після заміни цього вузла на єдиний листок з найчастішим класом. У разі збереження або покращення точності складна гілка замінюється узагальненим рішенням, а в новому вузлі фіксується найчастіший клас і відповідний розподіл ймовірностей. І вже виходячи з кінцевої структури, будується графічне представлення дерева, що відтворює повну послідовність поділів і прийнятих рішень у моделі [17].

Після завершення побудови дерева здійснюється передбачення класу для нових прикладів шляхом проходження дерева від кореневого вузла до листкового. На кожному кроці значення ознак нового об'єкта порівнюються з порогами, які були визначені під час побудови дерева. Якщо поточний вузол не є листковим, алгоритм переходить до лівої або правої гілки залежно від результату порівняння. Коли досягнуто листкового вузла, модель формує підсумок класифікації у вигляді прогнозованого класу разом із відповідною оцінкою ймовірності належності об'єкта до кожної з категорій. Ця ймовірність визначається на основі частки об'єктів кожного класу серед тих, що потрапили до відповідної підмножини під час навчання. У результаті система не лише повертає конкретне рішення, а й дає змогу кількісно оцінити рівень впевненості, що є критично важливим для задач, де точність прогнозу безпосередньо впливає на фінансові ризики.

Оскільки рішення щодо кредитування приймають безпосередньо працівники банку, важливо, щоб результат моделі інтерпретованим у зручній для

сприйняття формі. З цією метою додатково формується пояснення — шлях проходження конкретного клієнта по дереву рішень. Поданий шлях виводиться у текстовому форматі як послідовність умов: якщо значення ознаки не перевищує поріг — переходимо вліво, інакше — вправо, аж до досягнення кінцевого рішення.

На додачу до класифікації модель також формує індивідуальну рекомендацію щодо того, які зміни в параметрах могли б підвищити ймовірність позитивного рішення. Вона спирається на аналіз сприятливих гілок дерева та виявляє відмінності між ними й поточним шляхом клієнта — саме ті характеристики, які завадили отримати кращу оцінку. На основі цього виводиться перелік необхідних змін у форматі: поточне значення, порогова умова та напрям корекції [4].

Оціночним етапом роботи моделі є розрахунок показників якості передбачення, що дозволяють кількісно охарактеризувати її ефективність:

- Точність (Accuracy) — відношення кількості правильно передбачених класів до загальної кількості прикладів.
- Точність позитивного класу (Precision) — частка дійсно позитивних серед усіх передбачених як позитивні.
- Повнота (Recall) — частка дійсно позитивних, які модель змогла правильно ідентифікувати.
- AUC (Area Under the Curve) — це числова оцінка того, наскільки добре модель відрізняє один клас від іншого. Вона базується на побудові ROC-кривої — графіка, на якому по горизонталі відкладається частка хибнопозитивних передбачень, а по вертикалі — частка дійсно позитивних передбачень. Змінюючи поріг, за яким модель відносить об'єкт до певного класу, отримуються різні співвідношення цих двох величин, які формують криву. AUC — це площа під цією кривою, яка показує, наскільки впевнено модель відділяє один клас від іншого.

- Матриця помилок (Confusion Matrix) — табличне представлення результатів класифікації, яке показує співвідношення між передбаченими та фактичними класами [28].

У підсумку, сформовано завершену модель, здатну класифікувати нові приклади з урахуванням вхідних характеристик та надавати прозорі пояснення своїх рішень та практичні рекомендації для користувача.

2.6 Висновок до розділу 2

У цьому розділі було послідовно сформульовано прикладну задачу прогнозування кредитного ризику, яка полягає у бінарній класифікації позичальників за рівнем надійності на основі доступних соціально-економічних та фінансових ознак. Для її вирішення обґрунтовано обрано метод дерева рішень як найбільш відповідний умовам задачі — з урахуванням складної структури ознак, потреби в пояснюваності та гнучкості при моделюванні взаємозв'язків.

Модель побудована на основі спеціалізованого набору даних, який охоплює приклади фінансової поведінки клієнтів, що зверталися по кредитну лінію, забезпечену нерухомістю. У датасеті представлено детальні показники щодо доходу, заборгованості, кредитної історії та кредитного навантаження, які дозволяють змодельовати ймовірність виникнення дефолту. Формалізація задачі включала визначення цільової змінної, структуру вхідних ознак і побудову математичної постановки класифікації.

У межах реалізації моделі було впроваджено повний цикл обробки даних, побудови дерева, його спрощення та пояснення прийнятих рішень, що дозволило сформулювати інтерпретовану модель, здатну не лише передбачати ризик, а й надавати практичні рекомендації щодо підвищення шансів отримання кредиту. Передбачено можливість перевірки її точності на незалежній тестовій вибірці, з подальшим розрахунком показників якості — таких як Accuracy, Recall, AUC та матриця помилок — для оцінки ефективності моделі в умовах реального застосування.

Отже, поставлену задачу було повністю реалізовано: розроблено модель, що імітує логіку прийняття рішень у банківському кредитуванні, забезпечуючи як точність прогнозування, так і прозорість результатів.

3 ЕКСПЕРИМЕНТАЛЬНО-АНАЛІТИЧНИЙ РОЗДІЛ

3.1 Перевірка базової продуктивності моделі

Реалізовану модель дерева рішень, сформовану та описану в попередньому розділі, необхідно перевірити на практичну ефективність — тобто оцінити, наскільки її поведінка відповідає вимогам реального застосування, включаючи здатність узагальнювати результати і демонструвати надійність у прогнозуванні. Це потребує проведення серії тестувань, спрямованих на виявлення вразливостей моделі в різних умовах. Тому, насамперед, слід зафіксувати базові показники точності, які надалі слугуватимуть орієнтиром для оцінювання змін у якості.

Відповідно, початкове тестування виконується на даних, розподілених у пропорції 70% для навчальної вибірки, 15% — для валідаційної та 15% — для тестової. Максимальна глибина дерева на цьому етапі встановлюється на рівні 5 - така межа дає змогу уникнути перенавчання, зберігаючи водночас здатність моделі виявляти базові закономірності. Оцінювання здійснюється за показниками, визначеними в підрозділі 2.5 Реалізація та структура моделі. На рисунку 3.1 подано відповідні значення для кожної з метрик.

```

Точність (Accuracy): 0.7127
Точність позитивного класу (Precision): 0.7011
Повнота (Recall): 0.6450
AUC: 0.7320

Матриця помилок (Confusion Matrix):
                Передбачено 0   Передбачено 1
Справжній 0 (Bad)           367           110
Справжній 1 (Good)          142           258
  
```

Рисунок 3.1 – Показники якості побудованої моделі та матриця помилок для тестової вибірки

Результати свідчать про те, що модель демонструє задовільну якість класифікації у вихідній конфігурації: загальна точність становить 71,3%, що

вказує на достатній рівень узагальнення на тестовій вибірці. Значення точності позитивного класу є прийнятним, а повнота вказує на обмежену, але все ж наявну здатність моделі охоплювати такі приклади. Показник AUC підтверджує загальну спроможність моделі відрізнити один клас від іншого. Матриця помилок демонструє, що найбільша кількість правильних класифікацій припадає на клас 0, що узгоджується з метою мінімізації кредитних ризиків.

Окрім загальної оцінки, варто також звернути увагу на індивідуальну перевірку результату класифікації для випадково обраного клієнта з тестової вибірки. Його характеристики подано на рисунку 3.2, а послідовність проходження дерева рішень — на рисунку 3.3.

Вхідні дані клієнта:	
Age	27.0
Income	5667.0
Debt_Ratio	42.0
Region	4.0
CB_Age_Oldest_TL	83.0
CB_Age_Newest_TL	3.0
CB_Avg_Mos_File	40.0
CB_Nb_Sat_TL	26.0
CB_Nb_60_Plus_TL	4.0
CB_Nb_90_Plus_TL	2.0
CB_Pct_Sat_TL	80.0
CB_Mos_Since_Dlq	10.0
CB_Max_Dlq_12_Mos	4.0
CB_Max_Dlq_Ever	4.0
CB_Nb_Total_TL	30.0
CB_Nb_TL_Open_12	3.0
CB_Pct_IL_TL	33.0
CB_Nb_Inq_6_Mos	0.0
CB_Nb_Inq_6_Mos_excl_7_Days	0.0
CB_Rev_Util	3.0
CB_IL_Util	89.0
CB_Nb_Rev_TL_w_Bal	2.0
CB_Nb_IL_TL_w_Bal	4.0
CB_Nb_Rev_TL_75_Pct_Limit	0.0
CB_Pct_TL_w_Bal	58.0
App_Month	1.0
App_Day	1.0

Рисунок 3.2 – Вхідні дані випадково обраного клієнта

```

Шлях прийняття рішення:
Якщо CB_Rev_Util <= 31.00 (значення: 3.00)
  Якщо CB_Mos_Since_Dlq <= 24.00 (значення: 10.00)
    Якщо CB_Pct_Sat_TL <= 89.00 (значення: 80.00)
      -> Клас: 0

```

Рисунок 3.3 – Шлях прийняття рішення для випадково обраного клієнта

На рисунку 3.3 можна побачити, як модель, послідовно перевіряючи значення окремих ознак, приймає рішення щодо конкретного клієнта. У цьому прикладі сформовано шлях, що завершується класифікацією клієнта як ненадійного (клас 0). Відповідна оцінка впевненості моделі у вигляді ймовірнісного розподілу між класами наведена на рисунку 3.4: імовірність належності до класу 0 становить 0,6167, що перевищує порогове значення 0,5. Фактичне значення цільової змінної для цього клієнта також становить 0 і узгоджується з результатом моделі.

```

Ймовірність належності до класу 0 (дефолт): 0.6167
Ймовірність належності до класу 1 (надійний): 0.3833
Фактичний клас: 0

```

Рисунок 3.4 – Ймовірність належності до класу та фактична відповідь моделі для випадково обраного клієнта

З огляду на отриманий результат, для цього клієнта модель також сформуvala рекомендацію щодо потенційного переходу до більш сприятливого рішення — її наведено на рисунку 3.5. З метою уточнення логіки такої рекомендації на рисунку 3.6 представлено повне дерево рішень, на якому одночасно позначено фактичний шлях, пройдений клієнтом, і альтернативний маршрут до цільового класу.

```

Рекомендація для клієнта:
Змінити 'CB_Pct_Sat_TL': зараз = 80.00, потрібно > 89.00
Ймовірність у цільовому листку: клас 0 = 0.4845, клас 1 = 0.5155

```

Рисунок 3.5 – Рекомендовані зміни ознак випадково обраного клієнта для підвищення ймовірності класифікації до позитивного класу

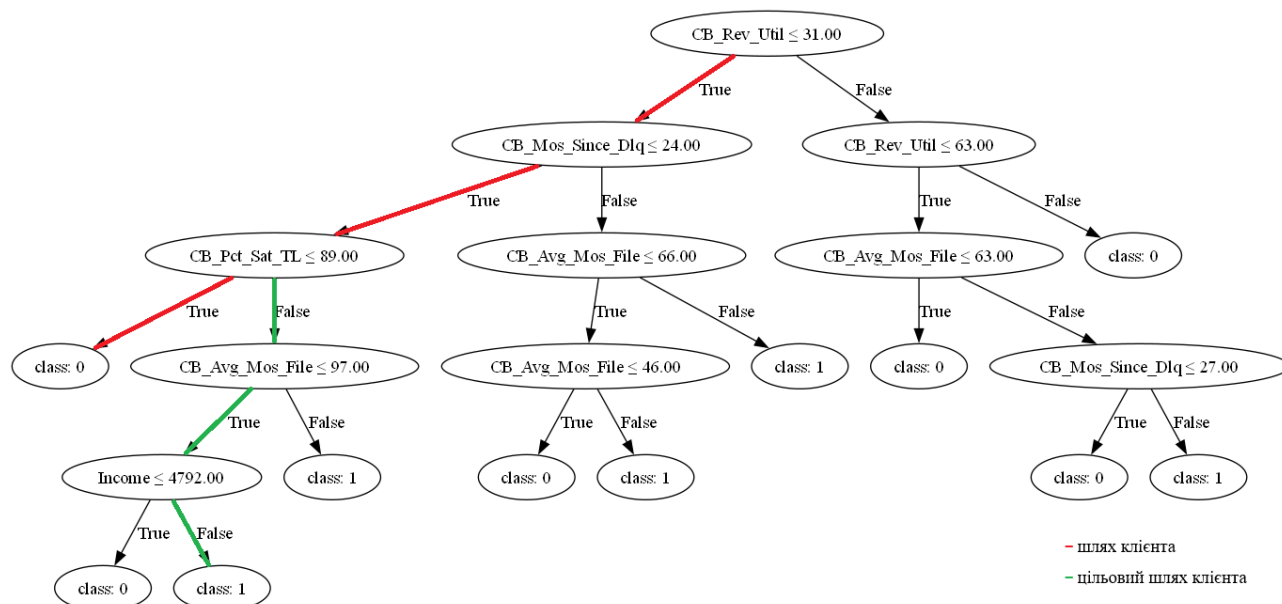


Рисунок 3.6 – Побудоване дерево рішень з виділеними фактичним та цільовим шляхами проходження для випадково обраного клієнта

У межах аналізу цього прикладу модель визначила мінімальну зміну однієї ознаки — частки задовільних кредитів, — якої достатньо для переходу до листка з більшою ймовірністю належності до класу 1. У цільовому вузлі ймовірність становить 0,5155, що перевищує порогове значення 0,5 і, відповідно, дозволяє віднести клієнта до позитивного класу.

Отже, перевірка реалізованої моделі показала її здатність до адекватного розпізнавання об'єктів різних класів. Окремі приклади підтвердили, що вона також послідовно формує обґрунтовані індивідуальні рекомендації, базуючись на дотриманні критерію мінімальності змін, тобто визначаючи найменше можливе коригування вхідних характеристик, необхідне для переходу до цільового листка або підвищення ймовірності бажаного класу. В цілому, отримані результати продемонстрували збалансовану роботу дерева рішень за основними критеріями якості, водночас все ще зберігається потенціал для подальшого вдосконалення шляхом оптимізації параметрів або аналізу впливу окремих ознак.

3.2 Аналіз впливу параметра глибини дерева

У межах загального дослідження точності моделі та можливостей покращення важливо перевірити її поведінку за різних конфігурацій параметрів. У випадку дерева рішень одним із критичних аспектів є контроль складності побудованої структури, оскільки відомо, що цей метод має схильність до перенавчання. Серед параметрів, що суттєво впливають на складність моделі, визначальною є глибина дерева — кількість рівнів розгалуження, яка впливає на здатність моделі виявляти закономірності без втрати узагальнюючих властивостей [16].

Вважається, що збільшення глибини дерева до певного рівня сприяє підвищенню точності класифікації, проте надмірне ускладнення структури зменшує її здатність до узагальнення. Аналіз цього впливу дає змогу визначити оптимальну глибину для конкретної задачі.

На підставі цього було проведено експеримент, під час якого запуск моделі виконувався з різним параметром максимальної глибини дерева в межах від 1 до 10. Оцінювання ефективності проводилося за основними показниками якості класифікації, з акцентом на аналіз динаміки зміни точності залежно від рівня ускладнення моделі.

Отримані результати наведено на рис. 3.7 – 3.8.

Максимальна глибина	Точність	Точність позитивного класу	Повнота	AUC
1	0.6853	0.6512	0.6675	0.6839
2	0.6921	0.7289	0.5175	0.7075
3	0.7035	0.7023	0.6075	0.6956
4	0.7001	0.7033	0.5925	0.7377
5	0.7127	0.7011	0.6450	0.7320
6	0.7092	0.6975	0.6400	0.7309
7	0.6967	0.6948	0.5975	0.7348
8	0.6807	0.6648	0.6050	0.7109
9	0.6613	0.6338	0.6100	0.7057
10	0.6556	0.6263	0.6075	0.6934

Рисунок 3.7 – Показники якості класифікації для різних значень максимальної глибини дерева

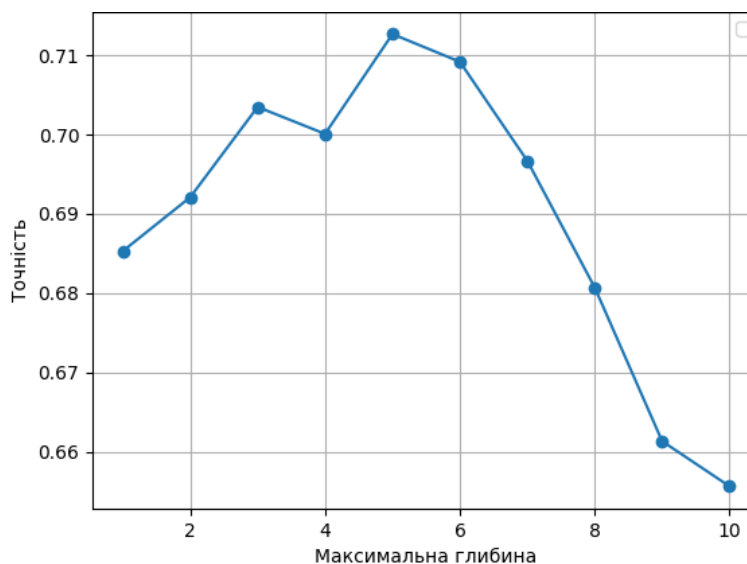


Рисунок 3.8 – Залежність показника точності класифікації від параметра максимальної глибини дерева

Найвищу точність класифікації було досягнуто при глибині 5 — саме на цьому рівні структура дерева виявилася найбільш збалансованою: вона зберігає достатню гнучкість для виявлення закономірностей, не втрачаючи здатності генералізувати. Значення точності позитивного класу і повноти у цій конфігурації залишаються близькими, що підтверджує стабільність моделі в умовах реальних прикладів. Після цього рівня подальше ускладнення структури призводить до помітного падіння загальної точності: від 71,3% на глибині 5 — до менш як 66% на глибині 10. Це свідчить про перенавчання, модель починає надмірно підлаштовуватися під навчальні приклади, втрачаючи здатність адаптуватися до нових ситуацій. І все ж на глибинах 6–7 точність ще залишається на прийнятному рівні, однак уже спостерігається поступове зниження повноти, що знижує практичну ефективність рішення. А при занадто малій глибині (1–2) спостерігається інший недолік: хоча загальна точність не є критично низькою, повнота падає до рівня близько 52%, що означає втрату здатності моделі виявляти значну частину позитивних прикладів. У контексті кредитного скорингу це може призводити до втрати потенційно прибуткових клієнтів, чия платоспроможність залишилась недооціненою.

Загалом, конфігурації з глибиною 5 і 6 продемонстрували найвищу злагодженість між точністю та повнотою. При цьому глибини 4 і 7 також

показали доволі прийнятні результати, хоча й зі слабшим співвідношенням цих показників. Таким чином, саме ці значення глибини можна вважати найперспективнішими для подальшого використання.

3.3 Перевірка стабільності моделі за допомогою крос-валідації

Попередній аналіз дав змогу визначити орієнтовно найкращу глибину дерева, однак ця оцінка ґрунтувалася лише на фіксованому поділі даних, що може викривлювати результати. Щоб переконатися в стабільності моделі й перевірити, наскільки обрані конфігурації демонструють стійку якість на різних підмножинах даних, варто застосувати метод крос-валідації [12]. Він надасть більш об'єктивну перевірку, оскільки передбачає багаторазове навчання й тестування моделі на різних частинах набору даних, фіксуючи характерні коливання метрик. У такий спосіб можна остаточно переконатися, які рівні глибини справді є найнадійнішими, і чи зберігається досягнута точність поза межами однієї конкретної вибірки.

Отже, перевірка стійкості моделі здійснюється шляхом п'ятикратної крос-валідації, у межах якої дані щоразу діляться на навчальну, валідаційну та підмножину для обрізання дерева. Як було встановлено в підрозділі 3.2, найкращі результати демонструють дерева з максимальною глибиною від 4 до 7, тому саме ці значення використовуються для порівняння при запусках моделі. А основною метрикою слугує точність класифікації, яка оцінювалася як середнє значення серед усіх п'яти фолдів, характерні коливання цієї метрики враховуються через розрахунок стандартного відхилення.

Зведений результат представлено на рис. 3.9.

```

Результати 5-кратної крос-валідації:
Максимальна глибина = 4: Точність = 0.6853 ± 0.0156
Максимальна глибина = 5: Точність = 0.6886 ± 0.0155
Максимальна глибина = 6: Точність = 0.6865 ± 0.0163
Максимальна глибина = 7: Точність = 0.6843 ± 0.0215

```

Рисунок 3.9 – Результати 5-кратної крос-валідації за метрикою точності при різній глибині дерева

За результатами крос-валідації найвищий середній показник точності було досягнуто при глибині 5 — це значення є максимальним серед перевірених конфігурацій і супроводжується найменшим стандартним відхиленням, тобто найстабільнішим результатом. Близькі результати спостерігаються також для глибини 6, однак при збільшенні глибини до 7 якість класифікації поступово знижується, а розкид результатів зростає, що підтверджує можливість перенавчання моделі. Глибина 4, у свою чергу, демонструє нижчу якість — точність залишається найменш виразною серед перевірених варіантів. У підсумку можна сказати, що перевірка показала - для цієї вибірки оптимальні значення глибини дерева перебувають у межах 5–6 рівнів, забезпечуючи збалансовану структуру моделі та стабільну класифікацію на різних підмножинах.

3.4 Аналіз важливості ознак

Розуміння впливу окремих характеристик на результат класифікації сприяє кращому розумінню побудованої моделі та логіки її рішень, що, у свою чергу, може бути корисним під час перевірки обґрунтованості кредитного рішення або швидкого аналізу його доцільності. Оцінка ролі кожної змінної дає змогу встановити, наскільки виправданим є її включення до моделі, а характеристики з мінімальним впливом можуть бути виключені на подальших етапах з метою спрощення структури без суттєвого зниження якості класифікації [9].

Відповідно, було проведено дослідження з визначення важливості кожної вхідної ознаки для дерева рішень з максимальною глибиною 5. Ваги ознак розраховуються шляхом підсумовування інформаційного приросту, що досягається при поділі вузлів за відповідною характеристикою. Чим частіше певна ознака забезпечує ефективне розділення прикладів, тим більшою є її роль у структурі моделі. На рисунку 3.10 подано порівняння отриманих значень важливості.

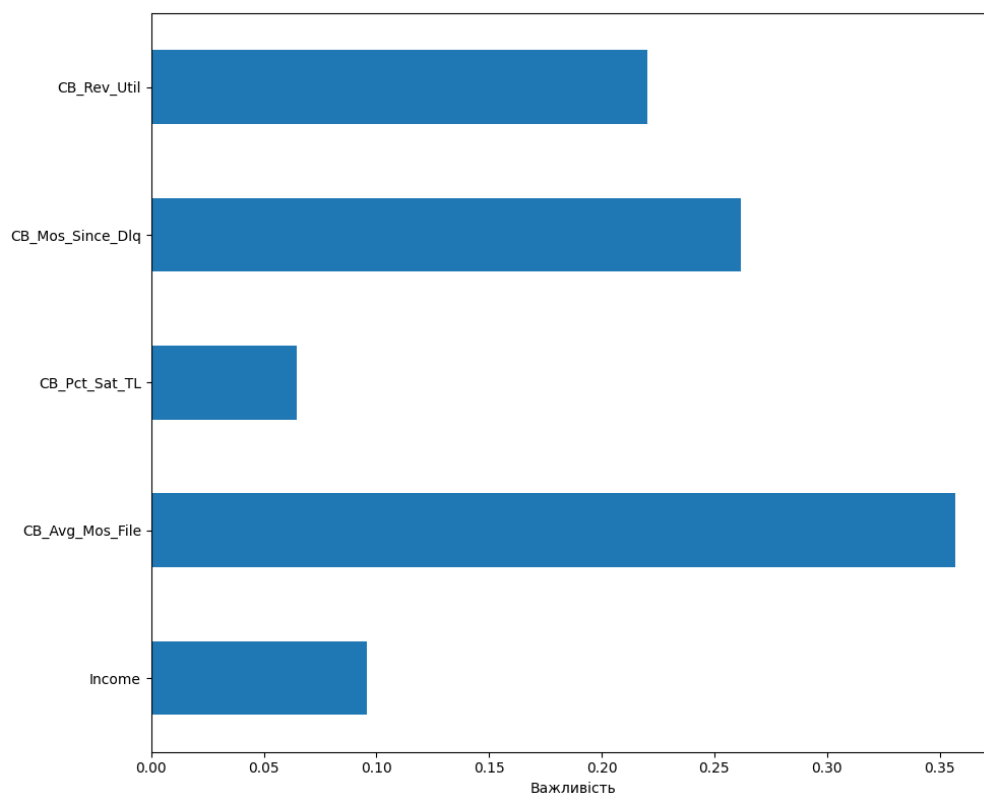


Рисунок 3.10 – Порівняння важливості ознак для дерева рішення з глибиною 5

Як видно з рисунка 3.10, модель надає найбільшу вагу змінній `CB_Avg_Mos_File`, що відображає середній вік кредитної історії клієнта. Високі значення також отримали показники `CB_Mos_Since_Dlq` (кількість місяців від останньої прострочки) та `CB_Rev_Util` (ступінь використання кредитного ліміту), що підтверджує їхню ключову роль у визначенні класу. Навпаки, ознаки `CB_Pct_Sat_TL` (частка задовільних кредитних ліній) та `Income` (дохід клієнта) виявились менш впливовими, хоч і не повністю неінформативними. Ця оцінка узгоджується з побудованим деревом рішень, наведеним в підрозділі 3.1. Перевірка базової продуктивності моделі, де структура моделі з усіма п'ятьма ознаками вже виглядає узгодженою та логічною. Тож усі змінні демонструють певний рівень інформативності, а їх загальна кількість є обмеженою, тому проведення додаткового обрізання ознак у цьому випадку видається недоцільним — спрощення не призводить до зменшення складності дерева, але може погіршити розподільну здатність моделі.

3.5 Перевірка справедливості моделі щодо соціально чутливих характеристик

У темі кредитування дедалі частіше піднімається питання етичності автоматизованих рішень. Коли йдеться про фінансову відповідальність перед клієнтом, будь-яке необґрунтоване обмеження доступу до кредиту сприймається як несправедливе — особливо якщо воно базується не на поведінкових або фінансових показниках, а на характеристиках, які не залежать від самого позичальника [18]. Саме тому багато установ дотримуються вимоги перевіряти моделі на відсутність дискримінації — прихованої або явної — щодо груп, що мають соціально чутливий статус. Йдеться про ті ознаки, які не повинні впливати на оцінку платоспроможності, оскільки самі по собі не є її маркером, але при цьому можуть спровокувати упереджене ставлення. Рейтингові агентства й кредитні платформи все частіше вводять додаткові обмеження, аби переконатися, що модель не створює систематичної переваги для одних і не занижує шанси інших лише через належність до певної групи. У межах цієї моделі такими характеристиками є регіон та вік — ознаки, які, з одного боку, не містять прямого економічного змісту, але, з іншого, можуть істотно впливати на рішення, якщо модель використовує їх необережно. Щоб виявити можливі відхилення у ставленні до різних підгруп, модель перевіряється окремо для кожної з них. Зокрема, якість класифікації оцінюється за метриками точності та повноти, що дозволяє визначити, чи зберігається послідовність результатів при зміні соціального профілю клієнта.

Результати цього аналізу подані у вигляді графіків на рис. 3.11 – 3.13: окремо для підгруп, сформованих за віком (18–24, 25–39, 40–59, 60+), та за типом регіону (1–5).

Кількість прикладів у кожній віковій групі:
 18-24: загалом = 12, клас 1 (Good) = 3, клас 0 (Bad) = 9
 25-39: загалом = 322, клас 1 (Good) = 132, клас 0 (Bad) = 190
 40-59: загалом = 446, клас 1 (Good) = 214, клас 0 (Bad) = 232
 60+: загалом = 97, клас 1 (Good) = 51, клас 0 (Bad) = 46

Кількість прикладів у кожному регіоні:
 Регіон 1.0: загалом = 114, клас 1 (Good) = 56, клас 0 (Bad) = 58
 Регіон 2.0: загалом = 62, клас 1 (Good) = 24, клас 0 (Bad) = 38
 Регіон 3.0: загалом = 101, клас 1 (Good) = 60, клас 0 (Bad) = 41
 Регіон 4.0: загалом = 380, клас 1 (Good) = 170, клас 0 (Bad) = 210
 Регіон 5.0: загалом = 220, клас 1 (Good) = 90, клас 0 (Bad) = 130

Рисунок 3.11 – Кількість прикладів у тестовій вибірці за віком та регіоном

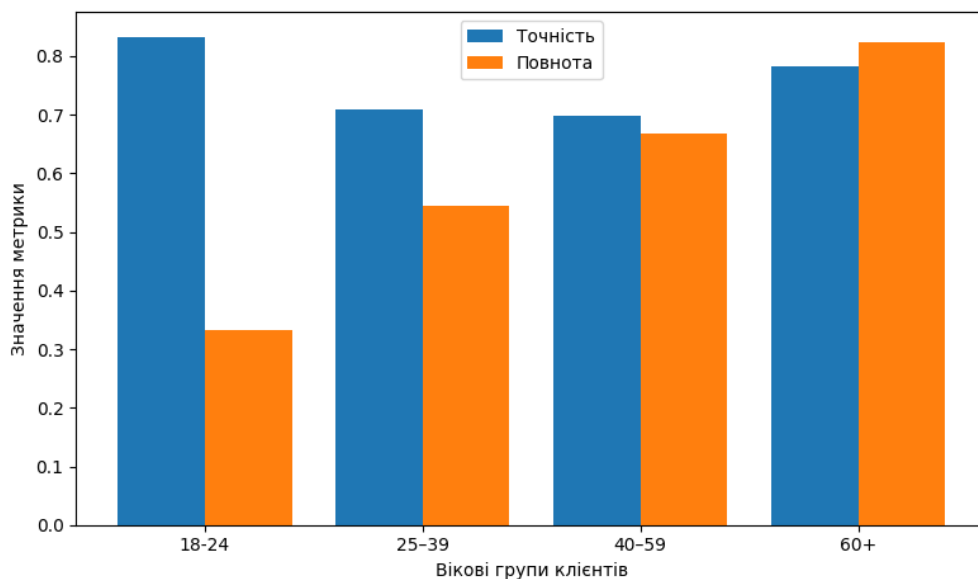


Рисунок 3.12 – Точність і повнота моделі дерева рішень з максимальною глибиною 5 за віковими групами клієнтів

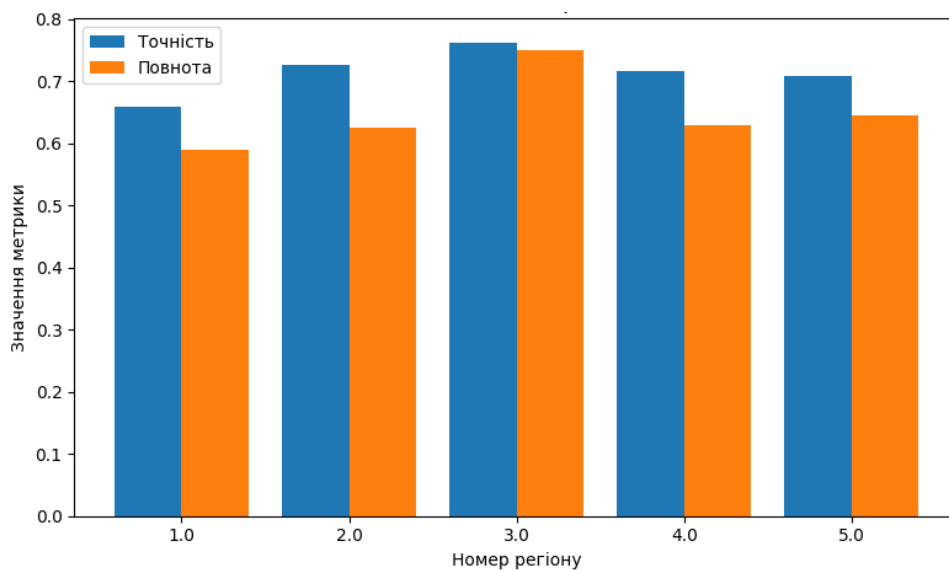


Рисунок 3.13 – Точність і повнота моделі дерева рішень з максимальною глибиною 5 за віковими групами клієнтів

Аналіз результатів за віковими підгрупами показує нерівномірність роботи моделі залежно від того, до якої категорії належить позичальник. Найбільш помітне відхилення спостерігається серед клієнтів віком 18–24 роки: попри високу загальну точність класифікації, показник повноти в цій категорії є найнижчим серед усіх груп. Модель рідко правильно виявляє приклади позитивного класу саме в цьому сегменті. Водночас ця група є й найменш представлена у вибірці — вона включає лише 12 прикладів, з яких позитивних усього три. За такого дисбалансу результати класифікації можуть бути випадковими або нестабільними, і це радше свідчить про обмеженість даних, а не про наявність реального упередження моделі. У старших групах — 25–39 і 40–59 років — значення метрик стають більш вирівняними, хоча повнота все ще трохи відстає від точності. Найвищі показники спостерігаються у категорії 60+, де модель демонструє впевненість і в загальній точності, і у виявленні позитивного класу. У цілому, всі суттєві відхилення виявляються саме в підгрупах із малою кількістю прикладів, тому говорити про системну упередженість або потребу в коригуванні ознак на цьому етапі передчасно.

У розрізі регіонів модель демонструє ще стабільнішу поведінку. Значення точності й повноти в кожному з територіальних сегментів перебувають на близьких рівнях, без різких коливань або систематичних відхилень. Незначна варіативність між регіонами не виходить за межі очікуваної статистичної розбіжності, а кількість прикладів у кожному з них є достатньою для базового аналізу. У такому контексті не виявлено ознак, які б вказували на упередженість моделі щодо певного регіону. Загалом, результати вказують на відносну рівномірність якості класифікації за географічною ознакою.

Тобто, результати перевірки справедливості моделі не дають підстав стверджувати про наявність суттєвого зміщення або дискримінаційного впливу з боку вікових чи регіональних ознак. Виявлені відхилення в якості класифікації збігаються з тими підгрупами, які виявилися найменш представленими у вибірці, що скоріше вказує на обмеженість емпіричної бази, ніж на системну упередженість самої моделі. У більш репрезентативних категоріях значення

метрик залишаються стабільними, що дозволяє вважати модель загалом рівномірною щодо проаналізованих соціально чутливих характеристик.

3.6 Порівняння якості класифікації з іншими моделями

Після аналізу продуктивності базової реалізації дерева рішень та перевірки доцільності зміни параметрів і структури ознак можна зробити висновок, що модель демонструє достатню якість класифікації й не потребує суттєвих коригувань. Проте для повноти оцінки її ефективності доцільно зіставити отримані результати з альтернативними алгоритмами класифікації. Порівняння із загальноприйнятими методами допоможе встановити, наскільки модель на основі дерева рішень поступається або відповідає їм за основними кількісними метриками. До таких методів належать логістична регресія, метод k найближчих сусідів (k -NN), випадковий ліс та градієнтний бустинг.

Таким чином, зіставлення буде проведено з використанням власної реалізації дерева рішень, яка працює з оптимально підбраною глибиною - 5 та базовим розподілом вибірки: 70% для навчання, 15% для валідації та 15% для тестування. Для об'єктивності оцінки її продуктивність порівнюється з бібліотечною реалізацією дерева рішень із пакету `sklearn`, яка працює з тією самою глибиною (5) та аналогічним розподілом даних. Інші згадані методи також реалізовано за допомогою стандартних засобів бібліотеки `sklearn`. Якість кожної з моделей оцінюється за тими самими метриками, що застосовувались раніше. На рисунку 3.14 подано результати порівняння.

Порівняння метрик класифікації:							
	Дерево рішень (власне)	Дерево рішень	Логістична регресія	k-NN	Випадковий ліс	Градієнтний бустинг	
Точність	0.7127	0.6693	0.7138	0.6271	0.7104	0.7332	
Точність позитивного класу	0.7011	0.6463	0.6830	0.6213	0.7086	0.7196	
Повнота	0.6450	0.6075	0.6950	0.4675	0.6200	0.6800	
AUC	0.7320	0.6643	0.7123	0.6143	0.7031	0.7289	

Рисунок 3.14 – Порівняння показників точності для різних моделей

Результати порівняння демонструють, що власна реалізація дерева рішень перевершує бібліотечну за всіма основними метриками. Логістична регресія

показує трохи вищі результати, хоча поступається у точності позитивного класу. Метод k-NN демонструє значно гірші показники порівняно з усіма іншими моделями, особливо в частині виявлення позитивного класу. Ансамблеві підходи, зокрема градієнтний бустинг, забезпечують найвищі значення більшості метрик і випереджають власну модель за якістю класифікації. А випадковий ліс демонструє результати, подібні до власної реалізації, проте з дещо вищою точністю. У цілому, хоча ансамблеві моделі й показують перевагу за деякими характеристиками, власне дерево рішень залишається конкурентним підходом, поєднуючи хорошу якість із високою прозорістю та інтерпретованістю.

3.7 Висновок до розділу 3

У межах третього розділу було всебічно досліджено якість побудованої моделі дерева рішень та її інтерпретованість у контексті задачі оцінки кредитного ризику. Перевірка базової конфігурації підтвердила, що структура моделі дозволяє досягати задовільних результатів на тестових даних: точність класифікації перевищує 70%, а показники повноти й AUC свідчать про стабільне розпізнавання об'єктів обох класів. Індивідуальний приклад клієнта показав, що модель здатна не лише приймати послідовні рішення, а й формувати конкретні рекомендації щодо підвищення ймовірності позитивної класифікації шляхом мінімальних змін у вхідних характеристиках.

Аналіз впливу параметра глибини дерева дозволив виявити оптимальні межі складності моделі: при глибині 5 структура виявилася найбільш збалансованою, забезпечуючи гармонійне поєднання точності та узагальнюваності. Подальше ускладнення призводить до перенавчання, тоді як занадто мала глибина не дозволяє виявити достатньо позитивних прикладів. Додаткова перевірка через крос-валідацію підтвердила, що конфігурації з глибиною 5–6 є не лише точними, а й найстабільнішими на різних підмножинах даних.

У процесі дослідження важливості ознак було встановлено, що модель найбільше спирається на показники, пов'язані з кредитною історією, зокрема середній вік кредитної історії клієнта, кількість місяців від останньої прострочки та ступінь використання ліміту. Менш важливими виявилися дохід та частка задовільних кредитних ліній, однак усі ознаки зробили певний внесок у класифікацію, тому подальше скорочення їх кількості не є доцільним.

Оцінка справедливості моделі засвідчила відносну рівномірність класифікації за регіонами та більшість вікових категорій. Незначні відхилення у повноті спостерігалися серед найменш репрезентованих груп, що швидше вказує на нестачу даних, аніж на наявність дискримінації. У більш повних підгрупах результати залишаються стабільними, що підтверджує етичність моделі щодо аналізованих соціально чутливих ознак.

Порівняння з іншими моделями продемонструвало, що власна реалізація дерева рішень є конкурентною – вона перевершує як базову бібліотечну версію, так і прості моделі на кшталт логістичної регресії та методу найближчих сусідів за більшістю ключових метрик, і незважаючи на деяке відставання від градієнтного бустингу чи випадкового лісу, модель поєднує задовільну точність із прозорістю та можливістю формування індивідуальних рекомендацій.

Загалом, результати дослідження підтвердили практичну придатність побудованої моделі для задачі кредитного скорингу. Вона демонструє стабільну якість, зберігає інтерпретованість і не виявляє ознак упередженості, що робить її ефективним інструментом для прийняття кредитних рішень у фінансовому секторі.

ВИСНОВОК

У межах виконання кваліфікаційної роботи було розглянуто задачу підтримки прийняття рішення щодо встановлення кредитного ліміту для клієнтів банку на основі аналізу їхніх характеристик. З урахуванням потреб реального банківського середовища, особлива увага приділялася не лише точності прогнозування, а й прозорості, доступності результату для подальшої інтерпретації та можливості формування обґрунтованих рекомендацій. Робота охоплює три послідовні етапи — аналітичний, методичний і прикладний — що в сукупності забезпечують цілісне вирішення поставленої задачі.

Перший розділ був присвячений аналізу поточної кредитної політики банку, структури портфеля, організації процесу ухвалення рішень і цифрової інфраструктури, що забезпечує обробку та оцінювання кредитних заявок. На основі цього аналізу сформульовано головну проблему — потребу в точному, але інтерпретованому підході до оцінки надійності клієнтів, який би дозволяв не лише прогнозувати ризик, а й будувати практичні рекомендації на індивідуальному рівні.

У другому розділі було сформовано підхід до вирішення цієї задачі. В основі моделі лежить алгоритм дерева рішень, який реалізовано з урахуванням перевірки якості моделі, візуалізації шляху прийняття рішення для конкретного клієнта і можливості виведення рекомендацій. Структура системи дозволяє адаптувати її до потреб кредитного фахівця, забезпечуючи зрозумілі й узгоджені оцінки. Параметри побудови були підібрані експериментально, з урахуванням балансу між складністю моделі та стабільністю результатів.

У третьому розділі проведено всебічне тестування розробленої моделі. Було перевірено її базову продуктивність, протестовано вплив глибини дерева на точність, оцінено стабільність за допомогою крос-валідації, а також проаналізовано важливість вхідних ознак. Модель продемонструвала впевнені результати як за точністю, так і за здатністю до пояснення своїх рішень. Порівняння з іншими підходами підтвердило її конкурентоспроможність: попри

нижчу складність порівняно з ансамблевими методами, вона забезпечує достатній рівень якості та повністю задовольняє вимоги до прозорості.

Таким чином, побудована система є логічно завершеним рішенням, яке поєднує класифікацію ризику з формуванням обґрунтованих індивідуальних висновків для клієнтів. Вона може бути використана як окремий аналітичний інструмент або як частина більшої системи кредитного скорингу. Перевагою реалізованого підходу є здатність надати не лише оцінку ризику, а й рекомендацію, що дозволяє клієнту зрозуміти, які аспекти його фінансової поведінки потребують покращення для підвищення шансу на отримання бажаного ліміту.

Застосування такої моделі може стати практично цінним доповненням до процесу ухвалення рішень, сприяючи скороченню кількості невиправдано ризикованих рішень, покращенню якості оцінювання клієнтів і формуванню більш стійкого кредитного портфеля. У перспективі можлива адаптація моделі до змін економічного середовища, а також інтеграція з іншими системами — зокрема, з експертними правилами або методами пояснюваного машинного навчання — для розширення її функціональності та підвищення гнучкості.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. 5 Classification Algorithms for Machine Learning. URL: <https://builtin.com/data-science/supervised-machine-learning-classification> (дата звернення: 02.06.2025).
2. A novel framework for enhancing transparency in credit scoring: Leveraging Shapley values for interpretable credit scorecards. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11318906/#:~:text=Studies%20comparing> (дата звернення: 02.06.2025).
3. Corezoid Actor Engine Anatomy. URL: <https://corezoid.com/product/> (дата звернення: 02.06.2025).
4. Counterfactual Explanations: The What-Ifs of AI Decision Making. URL: <https://kpmg.com/ch/en/insights/artificial-intelligence/counterfactual-explanation.html> (дата звернення: 02.06.2025).
5. Credit Risk Analysis with Machine Learning. URL: <https://medium.com/@evelyn.eve.9512/credit-risk-analysis-with-machine-learning-5fc4fbe06a99> (дата звернення: 02.06.2025).
6. Credit Risk Assessment in R with KNN. URL: <https://medium.com/@charlesdirenzo/credit-risk-assessment-in-r-with-knn-53c0ad06ecf7> (дата звернення: 02.06.2025).
7. Credit scoring models in banking BPM. URL: <https://www.infosysbpm.com/blogs/financial-services/credit-scoring-models.html> (дата звернення: 02.06.2025).
8. Decision Tree Algorithms. URL: <https://www.geeksforgeeks.org/decision-tree-algorithms/> (дата звернення: 02.06.2025).
9. Feature Importance in Decision Tree. URL: <https://medium.com/@krisalay/feature-importance-in-decision-tree-8e60f2174717> (дата звернення: 02.06.2025).
10. FICO is letting anyone download anonymized credit applications for AI research. URL: <https://qz.com/1150883/fico-is-running-a-competition-with-google-to-fix-how-credit-is-granted> (дата звернення: 02.06.2025).

11. Home Equity Line of Credit (HELOC). URL: <https://www.kaggle.com/datasets/averkiyoliabev/home-equity-line-of-creditheloc> (дата звернення: 02.06.2025).
12. How to choose ideal Decision Tree depth without overfitting?. URL: <https://www.geeksforgeeks.org/how-to-choose-ideal-decision-tree-depth-without-overfitting/> (дата звернення: 02.06.2025).
13. How to program a decision tree in Python from 0. URL: <https://anderfernandez.com/en/blog/code-decision-tree-python-from-scratch/> (дата звернення: 02.06.2025).
14. Interpretable Selective Learning in Credit Risk. URL: <https://arxiv.org/abs/2209.10127#> (дата звернення: 02.06.2025).
15. Machine Learning Models in Risk Management. GARP White Paper. 2022. URL: https://www.garp.org/hubfs/Whitepapers/a2r5d000003s85tAAA_RiskIntell.WP.MLModels.Feb24.22.pdf (дата звернення: 02.06.2025).
16. Overfitting in Decision Tree Models. URL: <https://www.geeksforgeeks.org/overfitting-in-decision-tree-models/> (дата звернення: 02.06.2025).
17. Post-Pruning and Pre-Pruning in Decision Tree. URL: <https://medium.com/analytics-vidhya/post-pruning-and-pre-pruning-in-decision-tree-561f3df73e65> (дата звернення: 02.06.2025).
18. Promises and Challenges of Causality for Ethical Machine Learning. URL: <https://montrealethics.ai/promises-and-challenges-of-causality-for-ethical-machine-learning/> (дата звернення: 02.06.2025).
19. What Is Credit Scoring? Purpose, Factors, and Role In Lending. URL: https://www.investopedia.com/terms/c/credit_scoring.asp (дата звернення: 02.06.2025).
20. ABC Scrooge. URL: <https://lime-systems.com/products-post/abs-scrooge/> (дата звернення: 02.06.2025).

21. Бізнесу. URL: https://a-bank.com.ua/business?utm_source=google&utm_medium=cpc&utm_campaign=224 (дата звернення: 02.06.2025).
22. Депозити. URL: https://a-bank.com.ua/deposits?utm_source=google&utm_medium=cpc&utm_campaign=2244 (дата звернення: 02.06.2025).
23. Історія рейтингу АКЦІОНЕРНЕ ТОВАРИСТВО "АКЦЕНТ-БАНК". URL: https://www.credit-rating.ua/ua/rate_history/21/713/ (дата звернення: 02.06.2025).
24. Історія рейтингу. URL: https://www.credit-rating.ua/ua/rate_history/24/713/ (дата звернення: 02.06.2025).
25. Картки. URL: https://a-bank.com.ua/cards?utm_source=google&utm_medium=cpc&utm_campaign=224450 (дата звернення: 02.06.2025).
26. Кредити. URL: https://a-bank.com.ua/credit?utm_source=google&utm_medium=cpc&utm_campaign= (дата звернення: 02.06.2025).
27. Наглядова статистика. URL: <https://bank.gov.ua/ua/statistic/supervision-statist> (дата звернення: 02.06.2025).
28. Поняття машинного навчання. URL: <https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall?hl=uk%5D> (дата звернення: 02.06.2025).
29. Послуги. URL: https://a-bank.com.ua/services?utm_source=google&utm_medium=cpc&utm_campaign=22445066764&utm_content=176378656485&utm_term=%D0%B0%20%D0%B1%D0%B0%D0%BD%D0%BA&gad_source=1&gad_campaignid=22445066764&gbraid=0AAAAACrrxPj9Kvh1gnuUVdFHIXeK2y04o&gclid=Cj0KCQjwgIXCBhDBARIsAELC9ZjrEMxFzReWUnUXLzEFKUsFd5VHyfehR45GPr-NIUudaZFX-vZRMW0aAov9EALw_wcB (дата звернення: 02.06.2025).

- 30.Скоринг в банках і фінансових організаціях: види, складові, вплив на кредитну історію. URL: <https://banker.ua/uk/skoring-v-ukra%D1%97nskix-bankax/> (дата звернення: 02.06.2025).
- 31.Стратегія. URL: <https://a-bank.com.ua/about/strategy> (дата звернення: 02.06.2025).
- 32.Фінансова звітність за рік, що закінчився 31 грудня 2024 р. : внутр. документ. Дніпро : АТ «А-БАНК», 2024. 108 с.
- 33.Річна інформація емітента цінних паперів за 2023 рік : внутр. документ. Дніпро : АТ «А-БАНК», 2023. 110 с.
- 34.Звіт керівництва (звіт про управління) за рік, що закінчився 31 грудня 2023 року : внутр. документ. Дніпро : АТ «А-БАНК», 2023. 31 с.
- 35.Фінансова звітність за рік, що закінчився 31 грудня 2023 р. : внутр. документ. Дніпро : АТ «А-БАНК», 2023. 113 с.
36. Кваліфікаційна робота бакалавра [Електронний ресурс] : методичні рекомендації для здобувачів ступеня бакалавра освітньо-професійної програми «Системний аналіз» зі спеціальності 124 Системний аналіз / уклад.: Т.А. Желдак, Т.В. Хом'як, А.В. Малієнко ; М-во освіти і науки України, Нац. техн. ун-т «Дніпровська політехніка». – Дніпро : НТУ «ДП», 2025. – 32 с.

ДОДАТОК А

Відомість матеріалів кваліфікаційної роботи

№ з/п	Позначення				Найменування	Кількість аркушів	Примітки		
1									
2					Документація				
3									
4	САУ.КР.25.53.ПЗ				Пояснювальна записка	82	Формат А4		
5									
6	САУ.КР.25.53.ДМ				Демонстраційний матеріал	16	Презентація на CD-R		
7									
8	САУ.КР.25.53.КР				Копія роботи	1	Диск CD-R		
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
					САУ.КР.25.53.ДА.ПЗ.				
Змін.	Аркуш	№ докум.	Підпис	Дата	Матеріали кваліфікаційної роботи	Літ.	Аркуш	Аркушів	
Розроб.		Хижняк М. Р.							
К. розд.		Желдак Т. А.							
Керівн.		Желдак Т. А.							
Н.контр.		Хом'як Т. В.							
Зав. каф.		Желдак Т. А.							
						НТУ «ДП», 12; 124-21-1			

ДОДАТОК Б

Відгук
на кваліфікаційну роботу бакалавра
здобувача вищої освіти групи 124 – 21 – 1
спеціальності 124 Системний аналіз

Тема кваліфікаційної роботи: «Розробка інтелектуальної системи прийняття рішення щодо встановлення кредитного ліміту».

Обсяг кваліфікаційної роботи 82 стор.

Мета кваліфікаційної роботи: підвищення якості попереднього оцінювання кредитного ризику шляхом розроблення підходу, що базується на характеристиках фінансової поведінки клієнта і забезпечує точність, послідовність та обґрунтованість результатів.

Актуальність теми обумовлена поточним станом кредитної діяльності банку, що характеризується обмеженістю чинної моделі оцінювання ризику.

Тема кваліфікаційної роботи безпосередньо пов'язана з об'єктом діяльності бакалавра спеціальності 124 Системний аналіз, оскільки в ній використані методи аналізу даних, моделювання за допомогою дерева рішень, крос-валідація, оцінювання за метриками класифікації та порівняльний аналіз.

Виконані в кваліфікаційній роботі завдання відповідають вимогам ступеня бакалавра. Оригінальність наукових рішень полягає в розробленні підходу до попередньої оцінки кредитного ризику з елементами пояснення результатів і формуванням базових рекомендацій залежно від типу клієнта. Удосконалено використання моделі дерева рішень шляхом її подальшого обрізання з метою підвищення стабільності класифікації за збереження інтерпретованості.

Практичне значення результатів кваліфікаційної роботи полягає в тому, що запропонований підхід до первинної оцінки кредитного ризику забезпечує поєднання точності та пояснюваності рішень, що робить його придатним для інтеграції в автоматизовані системи банківського скорингу.

Висновки підтверджують можливість використання результатів роботи в роботі фінансових установ різної форми власності.

Оформлення пояснювальної записки та демонстраційного матеріалу до неї виконано згідно з вимогами. Роботу виконано самостійно, відповідно до завдання та у повному обсязі.

У роботі не відзначено суттєвих недоліків, які можуть вплинути на її оцінювання.

Кваліфікаційна робота в цілому заслуговує оцінки: «відмінно» (100 балів).

З урахуванням викладеного авторка заслуговує присвоєння кваліфікації «бакалавр з системного аналізу».

Керівник кваліфікаційної роботи,
К.т.н., доц., зав.каф. САУ

_____ / Желдак Т.А.

ДОДАТОК В

Рецензія
на кваліфікаційну роботу бакалавра від АТ «А-БАНК»
здобувача вищої освіти групи 124 – 21 – 1
спеціальності 124 Системний аналіз

Робота справляє позитивне враження як з точки зору обраної теми, так і за змістом. Проблематика оцінки кредитного ризику залишається актуальною для банківського сектору, і запропонований підхід загалом відповідає логіці тих процесів, які реально використовуються у фінансових установах.

Матеріал подано послідовно, з розумінням предметної області. В роботі простежується прагнення до поєднання теоретичної частини з практичною реалізацією. Сама модель виглядає як реалістична основа для прототипу, який у разі доопрацювання міг би бути адаптований під внутрішні задачі банку або використаний як елемент допоміжного аналізу.

Загалом можна відзначити, що робота має прикладне спрямування та демонструє розуміння основних принципів, на яких будуються автоматизовані системи підтримки кредитних рішень

Рецензент,
Голова Правління
АТ «А-Банк»

_____ / Кандауров Ю. В.

ДОДАТОК Г

Г.1 Програмний код моделі дерева рішень

```

import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, precision_score, recall_score, roc_auc_score, confusion_matrix
from graphviz import Digraph

df = pd.read_csv("RiskData.csv")

drop_columns = [
    'FICO_Score', 'Prior_Custom_Score', 'Current_Custom_Score',
    'Loan_Amt', 'Loan_Amt_Req', 'LTV', 'Interest_Revenue', 'Sampling_Weight',
    'Nb_Borrowers', 'Bank_Relationship'
]
df.drop(columns=drop_columns, inplace=True, errors='ignore')
df['Region'] = df['Region'].str.extract(r'(\d+).').astype(float)

df['Application_Date'] = pd.to_datetime(df['Application_Date'], format='%Y%m%d', errors='coerce')
df['App_Month'] = df['Application_Date'].dt.month
df['App_Day'] = df['Application_Date'].dt.day
df.drop(columns=['Application_Date'], inplace=True)
df.replace(['.', 'NA', ''], np.nan, inplace=True)
df.dropna(inplace=True)
print(len(df))

df['Risk_Flag'] = df['Risk_Flag'].astype(int)
X = df.drop(columns=['Risk_Flag']).astype(float)
y = df['Risk_Flag']

X_temp, X_test, y_temp, y_test = train_test_split(X, y, test_size=0.15, random_state=42)
val_ratio = 0.15 / 0.85
X_train, X_val, y_train, y_val = train_test_split(X_temp, y_temp, test_size=val_ratio, random_state=42)

train_data = X_train.copy()
train_data['target'] = y_train
validation_data = X_val.copy()
validation_data['target'] = y_val
test_data = X_test.copy()
test_data['target'] = y_test

def entropy(y):

```

```

_, counts = np.unique(y, return_counts=True)
probs = counts / counts.sum()
return -np.sum(probs * np.log2(probs + 1e-9))

```

```

def information_gain(data, split_feature, split_value):
    left = data[data[split_feature] <= split_value]
    right = data[data[split_feature] > split_value]
    if len(left) == 0 or len(right) == 0:
        return 0
    h_before = entropy(data['target'])
    h_left = entropy(left['target'])
    h_right = entropy(right['target'])
    h_after = len(left)/len(data) * h_left + len(right)/len(data) * h_right
    return h_before - h_after

```

```

def best_split(data, features):
    best_gain = -1
    best_feature = None
    best_value = None
    for feature in features:
        values = data[feature].unique()
        for val in values:
            gain = information_gain(data, feature, val)
            if gain > best_gain:
                best_gain = gain
                best_feature = feature
                best_value = val
    return best_feature, best_value

```

```

class Node:
    def __init__(self, feature=None, value=None, left=None, right=None, prediction=None, class_probs=None):
        self.feature = feature
        self.value = value
        self.left = left
        self.right = right
        self.prediction = prediction
        self.class_probs = class_probs

```

```

def build_tree(data, features, depth=0, max_depth=5):
    if len(np.unique(data['target'])) == 1 or depth >= max_depth:
        prediction = int(data['target'].mode()[0])
        counts = data['target'].value_counts(normalize=True).to_dict()
        class_probs = {0: counts.get(0, 0.0), 1: counts.get(1, 0.0)}
        return Node(prediction=prediction, class_probs=class_probs)
    feature, value = best_split(data, features)
    if feature is None:

```

```

prediction = int(data['target'].mode()[0])
counts = data['target'].value_counts(normalize=True).to_dict()
class_probs = {0: counts.get(0, 0.0), 1: counts.get(1, 0.0)}
return Node(prediction=prediction, class_probs=class_probs)
left = data[data[feature] <= value]
right = data[data[feature] > value]
return Node(
    feature=feature,
    value=value,
    left=build_tree(left, features, depth+1, max_depth),
    right=build_tree(right, features, depth+1, max_depth)
)

def predict_tree(node, x):
    if node.prediction is not None:
        return node.prediction
    if x[node.feature] <= node.value:
        return predict_tree(node.left, x)
    else:
        return predict_tree(node.right, x)

def predict_dataset(tree, data):
    return np.array([predict_tree(tree, row.drop('target')) for _, row in data.iterrows()])

def prune_tree(node, validation_data):
    if node.left is not None and node.right is not None:
        left_data = validation_data[validation_data[node.feature] <= node.value]
        right_data = validation_data[validation_data[node.feature] > node.value]

        node.left = prune_tree(node.left, left_data)
        node.right = prune_tree(node.right, right_data)

    if node.left.prediction is not None and node.right.prediction is not None:
        pred_before = predict_dataset(node, validation_data)
        acc_before = accuracy_score(validation_data['target'], pred_before)

        majority = int(validation_data['target'].mode()[0])
        pred_after = np.full(len(validation_data), majority)
        acc_after = accuracy_score(validation_data['target'], pred_after)

        if acc_after >= acc_before:
            counts = validation_data['target'].value_counts(normalize=True).to_dict()
            class_probs = {0: counts.get(0, 0.0), 1: counts.get(1, 0.0)}
            return Node(prediction=majority, class_probs=class_probs)

    return node

```

```

def predict_proba_manual(node, x):
    if node.prediction is not None:
        probs = node.class_probs
        if not isinstance(probs, dict) or 0 not in probs or 1 not in probs:
            return {0: 0.0, 1: 0.0}
        return probs
    if x[node.feature] <= node.value:
        return predict_proba_manual(node.left, x)
    else:
        return predict_proba_manual(node.right, x)

def explain_decision_path(node, x, depth=0):
    if node.prediction is not None:
        print(" " * depth + f"-> Клас: {node.prediction}")
        return
    val = x[node.feature]
    print(" " * depth + f"Якщо {node.feature} <= {node.value:.2f} (значення: {val:.2f})")
    if val <= node.value:
        explain_decision_path(node.left, x, depth + 1)
    else:
        print(" " * depth + f"Інакше (переходимо вправо)")
        explain_decision_path(node.right, x, depth + 1)

def get_all_leaf_paths(node, path=None):
    if path is None:
        path = []
    if node.prediction is not None:
        return [(path.copy(), node)]
    results = []
    if node.left is not None:
        results += get_all_leaf_paths(node.left, path + [(node.feature, node.value, "<=")])
    if node.right is not None:
        results += get_all_leaf_paths(node.right, path + [(node.feature, node.value, ">")])
    return results

def visualize_tree_graphviz(root):
    dot = Digraph()
    node_counter = [0]

    def add_nodes(node):
        if node is None:
            return None
        node_id = f"node{node_counter[0]}"
        node_counter[0] += 1

```

```

if node.prediction is not None:
    label = f"class: {node.prediction}"
else:
    label = f"{node.feature} ≤ {node.value:.2f}"

dot.node(node_id, label)

if node.left:
    left_id = add_nodes(node.left)
    dot.edge(node_id, left_id, label="True")

if node.right:
    right_id = add_nodes(node.right)
    dot.edge(node_id, right_id, label="False")

return node_id

add_nodes(root)
dot.attr(rankdir='TB')
return dot

features = list(X.columns)
tree_root = build_tree(train_data, features, max_depth=5)
pruned_tree_root = prune_tree(tree_root, validation_data)

y_true = test_data['target'].values
y_pred = predict_dataset(pruned_tree_root, test_data)

accuracy = accuracy_score(y_true, y_pred)
precision = precision_score(y_true, y_pred)
recall = recall_score(y_true, y_pred)
y_proba = np.array([
    predict_proba_manual(pruned_tree_root, x)[1]
    for _, x in test_data.drop(columns='target').iterrows()
])
auc = roc_auc_score(y_true, y_proba)
cm = confusion_matrix(y_true, y_pred)
cm_df = pd.DataFrame(
    cm,
    index=["Справжній 0 (Bad)", "Справжній 1 (Good)"],
    columns=["Передбачено 0", "Передбачено 1"]
)

print(f"Точність (Accuracy): {accuracy:.4f}")
print(f"Точність позитивного класу (Precision): {precision:.4f}")
print(f"Повнота (Recall): {recall:.4f}")

```

```

print(f'AUC: {auc:.4f}')
print("\nМатриця помилок (Confusion Matrix):")
print(cm_df)

print()
print("="*40)
random_index = np.random.randint(0, len(test_data))
sample = test_data.drop(columns='target').iloc[random_index]
true_class = test_data['target'].iloc[random_index]
print("\nВхідні дані клієнта:")
print(sample)
print()
print("="*40)
print("\nШлях прийняття рішення:")
explain_decision_path(tree_root, sample)

print()
print("="*40)
probs = predict_proba_manual(tree_root, sample)
print(f"\nЙмовірність належності до класу 0 (дефолт): {probs[0]:.4f}")
print(f"\nЙмовірність належності до класу 1 (надійний): {probs[1]:.4f}")
print(f"\nФактичний клас: {true_class}")
print()
print("="*40)

sample_prediction = predict_tree(tree_root, sample)
sample_probs = predict_proba_manual(tree_root, sample)[1]

leaf_paths = get_all_leaf_paths(tree_root)

if sample_prediction == 0:
    candidate_paths = [(conds, leaf) for conds, leaf in leaf_paths if leaf.prediction == 1]
else:
    candidate_paths = [(conds, leaf) for conds, leaf in leaf_paths if leaf.prediction == 1 and leaf.class_probs[1] > sample_probs]

recommendations = []
for conds, _ in candidate_paths:
    needed_changes = []
    for feature, value, direction in conds:
        current_value = sample[feature]
        if direction == "<=" and current_value > value:
            needed_changes.append((feature, f"<= {value:.2f}", current_value))
        elif direction == ">" and current_value <= value:
            needed_changes.append((feature, f"> {value:.2f}", current_value))
    if needed_changes:
        recommendations.append((needed_changes, _))

```

```

shortest_change = min(recommendations, key=lambda x: len(x[0])) if recommendations else ([], None)
formatted_recommendation = [
    f'Змінити '{feature}': зараз = {current:.2f}, потрібно {cond}'
    for feature, cond, current in shortest_change[0]
]

print("\nРекомендація для клієнта:")
if formatted_recommendation:
    for line in formatted_recommendation:
        print(line)
    new_probs = shortest_change[1].class_probs
    print(f'\nЙмовірність у цільовому листку: клас 0 = {new_probs[0]:.4f}, клас 1 = {new_probs[1]:.4f}')
else:
    print("Рекомендаційних змін не знайдено.")

dot = visualize_tree_graphviz(tree_root)
dot.render("my_tree", format="png", cleanup=True)
dot.view()

```

Г.2 Опис програмної реалізації моделі

У ході реалізації моделі дерева рішень для класифікації кредитного ризику було побудовано повноцінну систему, що охоплює етапи обробки даних, побудови дерева, обрізки (pruning), прогнозування та генерації інтерпретованих рекомендацій для клієнтів. Реалізація виконана з нуля, без використання готових алгоритмів з бібліотеки `sklearn`.

На етапі попередньої обробки було очищено вхідні дані: вилучено змінні, які можуть містити інформацію про вже ухвалене кредитне рішення або бути недоступними на момент звернення клієнта (`FICO_Score`, `Loan_Amt`, тощо), оброблено пропущені значення та витягнуто окремі компоненти з дати подачі заявки. Також набір було розділено на тренувальну, валідаційну та тестову вибірки в пропорції 70%, 15% та 15% відповідно.

Функція `entropy(y)` обчислює ентропію цільової змінної, яка використовується як міра невизначеності. На її основі реалізовано критерій

інформаційного приросту (`information_gain`), який дозволяє обрати оптимальне розбиття вибірки за ознакою та значенням.

Функція `best_split` виконує перебір усіх можливих пар «ознака – поріг» і повертає той варіант, що забезпечує найбільший приріст інформації. Цей підхід є жадібним аналогом ID3/CART.

Основна побудова дерева реалізована у функції `build_tree`. Вона працює рекурсивно, створюючи об'єкти класу `Node` для кожного внутрішнього вузла. Якщо досягнуто максимальної глибини або всі приклади належать до одного класу, створюється листовий вузол із класом-прогнозом і розподілом ймовірностей.

Для уникнення перенавчання реалізовано процедуру обрізки (`prune_tree`) на валідаційній вибірці. Якщо об'єднання двох дочірніх листів у єдиний вузол не погіршує точність на перевірці — дерево спрощується. Це дозволяє зменшити складність моделі та покращити її узагальнення.

Функції `predict_tree` і `predict_dataset` реалізують процес класифікації для одного прикладу та для всього набору відповідно. Для оцінки моделі використано точність (`accuracy`), точність позитивного класу (`precision`), повноту (`recall`) та площу під ROC-кривою (`AUC`), що дозволяє оцінити якість класифікації за різними аспектами.

Для пояснення класифікаційного рішення реалізовано функцію `explain_decision_path`, яка показує шлях, пройдений об'єктом у дереві, включаючи всі умови. Додатково реалізовано функцію `predict_proba_manual`, яка повертає ймовірності належності до кожного класу, що зберігаються у листових вузлах.

Інтерпретованість моделі доповнено модулем рекомендацій. Якщо об'єкт класифіковано до негативного класу (дефолт), алгоритм здійснює пошук найближчих листків дерева з позитивним прогнозом. На основі цього пропонується мінімальний набір змін у значеннях ознак, які могли б покращити результат для конкретного клієнта. Якщо ж приклад уже належить до класу 1, порівнюються листки з вищими ймовірностями цього класу, і рекомендації формуються лише у тому разі, якщо можлива більш стабільна класифікація.

Таким чином, система надає зворотний зв'язок щодо потенційно покращуваних параметрів, дозволяючи клієнту або аналітику оцінити, наскільки легко можна змінити прогноз на краще.

Нарешті, для побудови візуального подання дерева використовується зовнішня бібліотека `graphviz`, що дозволяє створювати орієнтовані графи у вигляді зображень. Функція `visualize_tree_graphviz` рекурсивно обходить усі вузли дерева та генерує графічну структуру з підписами ознак і рішень, яка експортується у форматі PNG.

У результаті реалізовано повноцінну систему прийняття рішень на основі дерева рішень, що охоплює побудову моделі, її оцінку, валідацію, пояснення прогнозів і генерацію дій на основі них.