Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)
Кафедра Штучного інтелекту
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти перший (бакалаврський)

Інтелектуальна система моніторингу та виявлення аномальних банківських транзакцій на основі методів машинного навчання
(тема)

Виконав:
здобувач четвертого року навчання,
групи ІТШІ-22-3
Арина КУРОЧКІНА
(власне ім'я, прізвище)

Спеціальність 122 Комп'ютерні науки
(код і повна назва спеціальності)
Тип програми освітньо-професійна
Освітня програма Штучний інтелект
(повна назва освітньої програми)
Керівник доц. Олена ВОЛОЩУК
(посада, власне ім'я, прізвище)

Допускається до захисту

Завідувач кафедри ШІ _____ Лариса ЧАЛА
(підпис) (власне ім'я, прізвище)

2026 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук

Кафедра Штучного інтелекту

Рівень вищої освіти перший (бакалаврський)

Спеціальність 122 Комп'ютерні науки
(код і повна назва)

Тип програми освітньо-професійна

Освітня програма Штучний інтелект
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
«_____» _____ 20 ___ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві Курочкіній Арині Русланівні
(прізвище, ім'я, по батькові)

1. Тема роботи Інтелектуальна система моніторингу та виявлення аномальних банківських транзакцій на основі методів машинного навчання

затверджена наказом університету від 5 червня 20 26 р. № 488Ст

2. Термін подання здобувачем роботи до екзаменаційної комісії 8 липня 20 26 р.

3. Вихідні дані до роботи Науково-технічні публікації та дослідження у сфері виявлення фінансового шахрайства, дані Інтернет-джерел і сучасних антифрод-систем, документація Python та використаних програмних бібліотек, набір даних Credit Card Transactions Fraud Detection Dataset для навчання та тестування моделі машинного навчання, засоби розробки Python, Streamlit, SQLite, Pandas, Scikit-learn та Plotly

4. Перелік питань, що потрібно опрацювати в роботі
1) Теоретичні основи виявлення фінансового шахрайства
2) Проєктування системи моніторингу шахрайських операцій
3) Розробка та реалізація програмної системи

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Строк / термін виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	08.06.2026	виконано
2	Аналіз предметної галузі	09.06.2026	виконано
3	Аналіз існуючих систем моніторингу фінансових	11.06.2026	виконано
	операцій та вибір підходу до розв'язання задачі		
4	Підготовка набору даних та навчання моделі	14.06.2026	виконано
	машинного навчання		
5	Проєктування архітектури системи та бази даних	15.06.2026	виконано
6	Розробка програмної системи моніторингу	16.06.2026	виконано
	шахрайських транзакцій		
7	Написання пояснювальної записки	18.06.2026	виконано
8	Перевірка на академічний плагіат	20.06.2026	виконано
9	Нормоконтроль	22.06.2026	виконано
10	Підготовка презентації та доповіді	25.06.2026	виконано
11	Попередній захист	26.06.2026	виконано
12	Рецензування	29.06.2026	виконано
13	Захист перед ЕК	08.07.2026	

Дата видачі завдання 8 червня 2026 р.

Здобувач _____
(підпис)

Керівник роботи _____ доц. Олена ВОЛОЩУК
(підпис) (посада, власне ім’я, прізвище)

РЕФЕРАТ

Пояснювальна записка: 86 с., 21 рис., 12 табл., 3 дод., 22 джерела.

АНАЛІЗ ДАНИХ, ВИЯВЛЕННЯ ШАХРАЙСТВА, МАШИННЕ НАВЧАННЯ, МОНІТОРИНГ ТРАНЗАКЦІЙ, ФІНАНСОВЕ ШАХРАЙСТВО, CREDIT CARD FRAUD, FRAUD DETECTION, MACHINE LEARNING, RANDOM FOREST, TRANSACTION MONITORING.

Об’єкт дослідження – процес виявлення шахрайських фінансових транзакцій.

Предмет дослідження – методи та засоби машинного навчання для автоматизованого виявлення шахрайських операцій у фінансових системах.

Мета роботи – розробка програмної системи моніторингу фінансових транзакцій із використанням методів машинного навчання для автоматичного виявлення потенційно шахрайських операцій.

Методи дослідження – теоретичний аналіз наукових джерел, методи інтелектуального аналізу даних, машинного навчання, статистичного аналізу та програмної розробки з використанням мови Python, бібліотек Scikit-learn, Pandas, Streamlit і системи керування базами даних SQLite.

У роботі проведено аналіз сучасних підходів до виявлення фінансового шахрайства, досліджено методи машинного навчання для задач класифікації транзакцій та виконано проєктування системи моніторингу шахрайських операцій. Розроблено програмну систему, яка забезпечує автоматичний аналіз транзакцій, оцінювання рівня ризику, формування попереджень і ведення історії фінансових операцій.

ABSTRACT

Bachelor’s thesis contains: 86 pp., 21 fig., 12 tabl., 3 ann., 22 references.

CREDIT CARD FRAUD, DATA ANALYSIS, FINANCIAL FRAUD, FRAUD DETECTION, MACHINE LEARNING, RANDOM FOREST, TRANSACTION MONITORING.

Object of research – the process of detecting fraudulent financial transactions.

Subject of research – machine learning methods and tools for automated detection of fraudulent operations in financial systems.

Purpose of the thesis – development of a financial transaction monitoring system based on machine learning methods for automatic detection of potentially fraudulent transactions.

Research methods – theoretical analysis of scientific publications, data mining methods, machine learning, statistical analysis and software development using Python, Scikit-learn, Pandas, Streamlit and SQLite database management system.

The thesis analyzes modern approaches to financial fraud detection, investigates machine learning methods for transaction classification tasks and designs a transaction monitoring system. A software system has been developed that provides automatic transaction analysis, risk assessment, alert generation and transaction history management.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів 7
Вступ 8
1 Теоретичні основи виявлення шахрайських транзакцій 10
1.1 Поняття фінансового шахрайства та його види 10
1.2 Методи виявлення шахрайства 16
1.3 Використання машинного навчання для задач виявлення шахрайства 22
1.4 Аналіз існуючих рішень у сфері виявлення фінансового шахрайства 26
2 Проєктування системи моніторингу шахрайських транзакцій 33
2.1 Постановка задачі 33
2.2 Аналіз та підготовка набору даних 36
2.3 Вибір алгоритму машинного навчання 40
2.4 Проєктування архітектури системи 44
2.5 Проєктування бази даних 47
3 Розробка та реалізація програмної системи 51
3.1 Реалізація моделі машинного навчання 51
3.2 Реалізація бази даних 55
3.3 Реалізація вебзастосунку 59
3.4 Реалізація фонового моніторингу 66
3.5 Тестування системи 69
Висновки 74
Перелік джерел посилання 76
Додаток А Програмний код навчання моделі 79
Додаток Б Програмний код створення бази даних 84
Додаток В Відомість кваліфікаційної роботи 86

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

БД – база даних;

ШІ – штучний інтелект;

AI – Artificial Intelligence – штучний інтелект;

API – Application Programming Interface – програмний інтерфейс застосунку;

CNP Fraud – Card-Not-Present Fraud – шахрайство без фізичного використання платіжної картки;

CNN – Convolutional Neural Network – згорткова нейронна мережа;

GUI – Graphical User Interface – графічний інтерфейс користувача;

ML – Machine Learning – машинне навчання;

RF – Random Forest – випадковий ліс;

SQL – Structured Query Language – мова структурованих запитів;

SQLite – Lightweight Structured Query Language Database – вбудована реляційна система керування базами даних.

ВСТУП

Сучасний розвиток цифрових технологій та фінансових сервісів сприяв значному збільшенню кількості електронних транзакцій. Разом із цим зростає кількість випадків фінансового шахрайства, що створює суттєві ризики для банківських установ, платіжних систем та їх клієнтів. Традиційні методи контролю фінансових операцій не завжди забезпечують своєчасне виявлення нових шахрайських схем, тому актуальним напрямом є використання методів машинного навчання для автоматизації процесів моніторингу транзакцій.

Актуальність роботи обумовлена необхідністю підвищення ефективності виявлення шахрайських фінансових операцій шляхом застосування сучасних алгоритмів аналізу даних та машинного навчання. Використання таких підходів дозволяє автоматично аналізувати великі обсяги інформації, оцінювати ризики та своєчасно виявляти підозрілу активність.

Метою роботи є розробка програмної системи моніторингу фінансових транзакцій із використанням методів машинного навчання для автоматичного виявлення потенційно шахрайських операцій.

Для досягнення поставленої мети було виконано аналіз предметної галузі, досліджено сучасні методи виявлення фінансового шахрайства, обрано алгоритм машинного навчання, спроєктовано архітектуру системи та базу даних, реалізовано програмну систему й проведено її тестування.

Об'єктом дослідження є процес виявлення шахрайських фінансових транзакцій.

Предметом дослідження є методи та засоби машинного навчання для автоматизованого виявлення шахрайських операцій у фінансових системах.

Практичне значення роботи полягає у створенні програмної системи, яка забезпечує автоматичний аналіз транзакцій, оцінювання рівня ризику та формування попереджень про потенційно шахрайські операції. Результати

роботи можуть бути використані під час розробки та вдосконалення антифрод-систем у банківській сфері та фінансових сервісах.

Робота складається зі вступу, трьох розділів, загальних висновків, списку використаних джерел та додатків. У роботі розглянуто теоретичні основи виявлення фінансового шахрайства, виконано проєктування системи моніторингу транзакцій та реалізовано програмне забезпечення з використанням алгоритмів машинного навчання.

1 ТЕОРЕТИЧНІ ОСНОВИ ВИЯВЛЕННЯ ШАХРАЙСЬКИХ ТРАНЗАКЦІЙ

1.1 Поняття фінансового шахрайства та його види

Сучасний розвиток інформаційних технологій суттєво змінив принципи функціонування фінансової сфери. Активне впровадження електронних платіжних систем, мобільного банкінгу, онлайн-переказів та цифрових фінансових сервісів забезпечило значне підвищення швидкості та зручності проведення фінансових операцій. Разом із позитивними наслідками цифровізації фінансового сектору з'явилися нові загрози, пов'язані з несанкціонованим доступом до фінансових ресурсів, викраденням персональних даних та здійсненням шахрайських операцій. Саме тому проблема фінансового шахрайства сьогодні належить до найбільш актуальних викликів для банківських установ, платіжних систем та фінансових організацій у всьому світі [1].

Фінансове шахрайство являє собою сукупність умисних протиправних дій, спрямованих на отримання матеріальної вигоди шляхом обману, приховування інформації, маніпулювання даними або використання довіри користувачів фінансових послуг. Основною метою шахрайських дій є незаконне заволодіння грошовими коштами, персональними даними клієнтів або отримання доступу до фінансових ресурсів. Особливістю сучасного фінансового шахрайства є його тісний зв'язок із цифровими технологіями та використанням мережевої інфраструктури для реалізації злочинних схем [2].

За даними сучасних досліджень, масштаби фінансового шахрайства щороку збільшуються разом із розвитком електронної комерції та дистанційних банківських послуг. Якщо раніше більшість фінансових злочинів вимагала фізичної присутності злочинця або безпосереднього контакту з жертвою, то сьогодні значна частина шахрайських операцій

здійснюється дистанційно. Це значно ускладнює процес виявлення та попередження незаконних дій, а також створює необхідність використання автоматизованих систем моніторингу транзакцій [1].

Зростання масштабів фінансового шахрайства обумовлюється сукупністю економічних, соціальних та технологічних факторів. Однією з головних причин є стрімка цифровізація фінансових послуг, яка значно спростила доступ користувачів до банківських сервісів, але одночасно створила нові можливості для зловмисників. Використання мобільного банкінгу, електронних платіжних систем та дистанційного обслуговування клієнтів призвело до збільшення кількості потенційних точок доступу до фінансової інформації, що може бути використана для реалізації шахрайських схем [1].

Важливим фактором поширення фінансового шахрайства є також людський фактор. Недостатній рівень цифрової грамотності користувачів, нехтування правилами інформаційної безпеки, використання слабких паролів та довіра до неперевірених джерел інформації значно підвищують ризик успішного здійснення шахрайських атак. Саме тому значна частина сучасних фінансових злочинів базується не лише на технічних засобах, але й на використанні методів соціальної інженерії [3].

Крім того, суттєвий вплив на розвиток фінансового шахрайства має постійне вдосконалення інформаційних технологій. Зловмисники активно використовують автоматизовані інструменти, спеціалізоване програмне забезпечення та можливості мережі Інтернет для пошуку потенційних жертв і реалізації складних багаторівневих схем незаконного отримання фінансової вигоди. У багатьох випадках шахрайські операції виконуються одночасно щодо великої кількості користувачів, що значно збільшує масштаби завданих збитків [4].

Фінансове шахрайство має серйозні негативні наслідки як для окремих користувачів, так і для фінансових установ та економіки загалом. Для клієнтів банків основними ризиками є прямі фінансові втрати,

викрадення персональних даних, погіршення кредитної історії та втрата довіри до фінансових сервісів. Для банківських установ наслідками можуть бути значні фінансові збитки, витрати на компенсацію коштів постраждалим клієнтам, репутаційні ризики та необхідність постійного вдосконалення систем безпеки.

Негативний вплив фінансового шахрайства поширюється і на економіку держави в цілому. Зростання кількості шахрайських операцій призводить до збільшення витрат на забезпечення кібербезпеки, ускладнює функціонування фінансових ринків та може знижувати рівень довіри населення до сучасних цифрових фінансових інструментів. У зв'язку з цим проблема боротьби з фінансовим шахрайством набуває не лише економічного, але й соціального значення.

Фінансове шахрайство є складним багатофакторним явищем, яке може проявлятися у різних формах залежно від способів отримання доступу до фінансових ресурсів та характеру здійснюваних операцій. Основні напрями фінансового шахрайства наведені на рисунку 1.1.

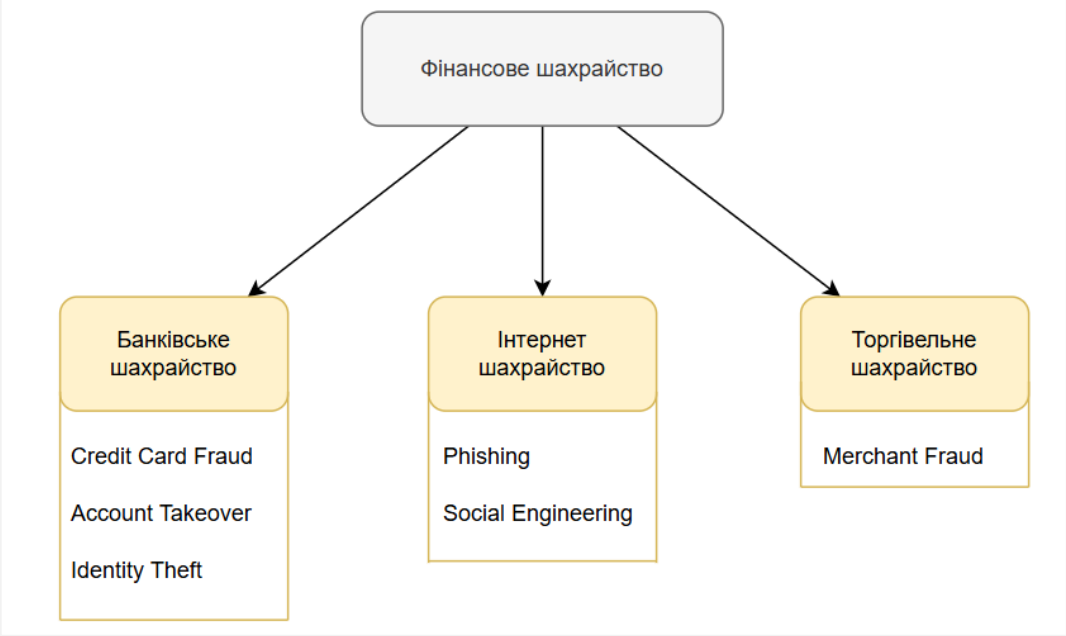


Рисунок 1.1 – Класифікація фінансового шахрайства

Серед різноманітних форм фінансового шахрайства особливу увагу привертає шахрайство з використанням платіжних карток (Credit Card Fraud). Даний вид злочинної діяльності передбачає незаконне використання реквізитів банківських карток для здійснення фінансових операцій без дозволу власника рахунку. Шахрайство з платіжними картками стало одним із найпоширеніших видів фінансових злочинів через стрімке поширення безготівкових розрахунків та електронної комерції [5].

Однією з найбільш поширених форм карткового шахрайства є Card-Not-Present Fraud. Такі операції здійснюються без фізичного використання картки та переважно характерні для онлайн-покупок. Для проведення платежу зловмиснику достатньо володіти основними реквізитами картки, зокрема її номером, строком дії та кодом безпеки. Через відсутність фізичної перевірки платіжного інструменту даний вид шахрайства характеризується високим рівнем ризику для банківських установ та платіжних систем [6].

Ще одним поширеним видом фінансового шахрайства є захоплення облікового запису користувача (Account Takeover Fraud). У даному випадку злочинці отримують доступ до банківських акаунтів клієнтів шляхом викрадення паролів, використання шкідливого програмного забезпечення або застосування методів соціальної інженерії. Після отримання контролю над обліковим записом зловмисники можуть виконувати грошові перекази, змінювати персональні дані користувача та здійснювати інші фінансові операції від його імені.

Окрему категорію становить шахрайство, пов'язане з викраденням особистих даних (Identity Theft). Даний вид шахрайства полягає у незаконному використанні персональної інформації фізичної особи для отримання кредитів, відкриття рахунків або проведення інших фінансових операцій. Особливу небезпеку становить той факт, що жертва часто дізнається про злочин лише після виникнення фінансових зобов'язань або інших негативних наслідків.

Важливу роль у сучасних схемах фінансового шахрайства відіграють фішингові атаки. Фішинг являє собою метод отримання конфіденційної інформації шляхом введення користувача в оману за допомогою підроблених вебсайтів, електронних листів, повідомлень або телефонних дзвінків. Метою фішингових атак зазвичай є отримання логінів, паролів, реквізитів банківських карток або іншої інформації, яка може бути використана для подальшого здійснення шахрайських операцій.

Характеристика найбільш поширених видів фінансового шахрайства наведена в таблиці 1.1.

Таблиця 1.1 – Основні види фінансового шахрайства

Вид шахрайства	Основна характеристика
Credit Card Fraud	Незаконне використання реквізитів банківської картки
Card-Not-Present Fraud	Проведення онлайн-платежів без фізичної картки
Account Takeover Fraud	Отримання контролю над банківським акаунтом користувача
Identity Theft	Використання викрадених персональних даних
Phishing	Отримання конфіденційної інформації шляхом обману
Merchant Fraud	Шахрайські дії з боку продавців або торговельних майданчиків

Аналіз даних, наведених у таблиці 1.1, свідчить про значну різноманітність сучасних форм фінансового шахрайства. Незважаючи на відмінності у механізмах реалізації шахрайських схем, усі вони мають спільну мету – отримання неправомірної фінансової вигоди шляхом використання недоліків інформаційних систем, платіжної інфраструктури або людського фактора. При цьому більшість сучасних шахрайських операцій пов'язана саме з використанням цифрових каналів взаємодії між клієнтами та фінансовими установами.

Особливу небезпеку становлять комбіновані схеми шахрайства, у яких одночасно використовуються декілька методів впливу на потенційну жертву. Наприклад, фішинг може застосовуватися для отримання персональних даних користувача, після чого викрадена інформація використовується для захоплення облікового запису або проведення несанкціонованих карткових операцій. Подібні багаторівневі схеми значно ускладнюють процес виявлення шахрайських дій та потребують використання комплексних засобів захисту [2].

Важливою особливістю сучасного фінансового шахрайства є його постійна еволюція. З розвитком цифрових технологій та впровадженням нових фінансових сервісів зловмисники адаптують існуючі методи та розробляють нові способи обходу механізмів безпеки. Це створює додаткові виклики для банківських установ, які змушені постійно вдосконалювати системи моніторингу та виявлення підозрілих операцій [3].

Останніми роками характер фінансового шахрайства суттєво змінився під впливом цифрової трансформації фінансового сектору. Якщо раніше більшість шахрайських операцій була пов'язана з фізичним доступом до платіжних інструментів або документів, то сьогодні переважають дистанційні способи здійснення незаконних дій. Розвиток електронної комерції, мобільних додатків та онлайн-банкінгу створив нові можливості для фінансових операцій, але одночасно збільшив кількість потенційних кіберзагроз [4].

Особливістю сучасного цифрового середовища є висока швидкість проведення фінансових транзакцій. У багатьох випадках між моментом здійснення шахрайської операції та її виявленням проходить лише декілька секунд або хвилин. За таких умов традиційні методи контролю часто виявляються недостатньо ефективними, що обумовлює необхідність використання автоматизованих систем моніторингу та аналізу транзакцій у режимі реального часу [7]. Основні фактори, що сприяють поширенню фінансового шахрайства в умовах цифровізації, наведені на рисунку 1.2.



Рисунок 1.2 – Фактори поширення фінансового шахрайства в цифровому середовищі

Таким чином, фінансове шахрайство є складною багатокомпонентною проблемою, що охоплює широкий спектр незаконних дій у сфері фінансових операцій. Зростання кількості цифрових транзакцій та ускладнення шахрайських схем обумовлюють необхідність використання сучасних технологій аналізу даних для своєчасного виявлення підозрілої активності. Одним із найбільш перспективних напрямів вирішення даної проблеми є застосування методів машинного навчання, які дозволяють автоматизувати процес виявлення шахрайських операцій та підвищити ефективність систем фінансової безпеки.

1.2 Методи виявлення шахрайства

Ефективне виявлення шахрайських фінансових операцій є одним із ключових завдань сучасних банківських установ, платіжних систем та

фінансових організацій. Постійне збільшення кількості електронних транзакцій, розвиток цифрового банкінгу та ускладнення шахрайських схем призводять до необхідності використання спеціалізованих методів контролю фінансових операцій. Основною метою таких методів є своєчасне виявлення підозрілих дій, запобігання фінансовим втратам та забезпечення безпеки клієнтів фінансових установ [4].

Історично системи виявлення шахрайства розвивалися паралельно з розвитком фінансового сектору. Спочатку контроль здійснювався виключно вручну співробітниками банківських установ, однак із зростанням кількості клієнтів та фінансових операцій такий підхід виявився малоефективним. У подальшому почали застосовуватися автоматизовані системи контролю на основі заздалегідь визначених правил та статистичних методів аналізу даних. Сьогодні поряд із традиційними підходами активно використовуються інтелектуальні системи, засновані на методах машинного навчання та аналізу великих даних [8].

У сучасній практиці методи виявлення шахрайства можна умовно поділити на три основні групи: ручний контроль, системи на основі правил та інтелектуальні методи аналізу даних. Загальну класифікацію методів виявлення шахрайства доцільно подати на рисунку 1.3.



Рисунок 1.3 – Класифікація методів виявлення фінансового шахрайства

Представлена на рисунку 1.3 класифікація демонструє еволюцію підходів до виявлення фінансового шахрайства. З розвитком банківського сектору та збільшенням кількості електронних транзакцій вимоги до систем моніторингу постійно зростали. Якщо на початкових етапах основною метою було виявлення вже здійснених шахрайських операцій, то сучасні системи орієнтовані на попередження та блокування підозрілих дій ще до завершення транзакції [1].

Традиційні методи виявлення шахрайства переважно належать до реактивного підходу. У межах такого підходу аналіз здійснюється після проведення операції або після появи ознак підозрілої активності. Основним недоліком реактивних методів є те, що фінансові втрати можуть виникнути ще до моменту виявлення шахрайства. Саме тому сучасні фінансові установи поступово переходять до проактивних систем моніторингу, які дозволяють оцінювати ризики безпосередньо під час виконання транзакції [8].

Важливою характеристикою будь-якої системи виявлення шахрайства є її здатність забезпечувати баланс між безпекою та зручністю використання фінансових сервісів. Надмірно жорсткі механізми контролю можуть призводити до блокування легітимних операцій, тоді як недостатній рівень контролю підвищує ризик успішного здійснення шахрайських дій. У зв'язку з цим сучасні системи моніторингу повинні не лише виявляти потенційно небезпечні операції, але й мінімізувати кількість помилкових спрацювань [9].

Одним із найстаріших підходів до боротьби з фінансовим шахрайством є ручний контроль транзакцій. У межах даного підходу співробітники банку або спеціалісти служби безпеки здійснюють аналіз фінансових операцій, які були попередньо визначені як потенційно підозрілі. Рішення щодо наявності або відсутності ознак шахрайства приймається людиною на основі аналізу інформації про клієнта, історії його операцій та інших доступних даних.

Основною перевагою ручного контролю є можливість врахування контекстної інформації та використання професійного досвіду спеціаліста. Однак ефективність такого підходу суттєво знижується при збільшенні кількості транзакцій. Сучасні банки можуть обробляти мільйони операцій щодня, тому ручна перевірка всіх фінансових операцій стає практично неможливою.

Наступним етапом розвитку систем виявлення шахрайства стало впровадження систем на основі правил (Rule-Based Systems). У таких системах використовуються заздалегідь визначені правила та умови, які дозволяють автоматично визначати підозрілі транзакції. Наприклад, система може позначати операцію як ризикову у випадку перевищення певної суми платежу, здійснення великої кількості операцій за короткий проміжок часу або проведення транзакції з незвичної географічної локації.

Rule-Based системи тривалий час залишалися основним інструментом боротьби з фінансовим шахрайством. Їх популярність пояснюється відносною простотою реалізації, прозорістю прийняття рішень та можливістю швидкого налаштування під конкретні вимоги організації. Крім того, використання чітко визначених правил дозволяє легко пояснити причину блокування або позначення певної транзакції як підозрілої.

Незважаючи на численні переваги, системи на основі правил мають ряд суттєвих недоліків. Насамперед вони не здатні ефективно виявляти нові типи шахрайства, які не були враховані під час формування набору правил. Крім того, зі збільшенням кількості правил зростає складність їх підтримки та оновлення. У результаті ефективність таких систем поступово знижується в умовах постійної еволюції шахрайських схем.

Отже, традиційні підходи забезпечили автоматизацію процесу виявлення шахрайства, однак їх обмежена гнучкість і складність адаптації до нових загроз стали передумовою для впровадження методів машинного навчання.Переваги та недоліки традиційних методів виявлення шахрайства наведено в таблиці 1.2.

Таблиця 1.2 – Переваги та недоліки традиційних методів виявлення шахрайства

Метод	Переваги	Недоліки
Ручний контроль	Висока точність експертної оцінки	Низька швидкість обробки
Rule-Based системи	Простота реалізації, прозорість	Неможливість адаптації до нових схем
Статистичні методи	Автоматизація аналізу	Обмежена гнучкість

Дані таблиці 1.2 демонструють, що традиційні методи виявлення шахрайства мають як переваги, так і суттєві обмеження. Незважаючи на можливість ефективного контролю окремих типів ризиків, такі підходи не забезпечують необхідного рівня адаптивності в умовах постійної зміни поведінки шахраїв. Саме тому фінансові установи поступово переходять до використання більш складних методів аналізу даних та інтелектуальних систем підтримки прийняття рішень [7].

Важливе місце серед традиційних методів займають статистичні підходи до аналізу транзакцій. Їх основною метою є виявлення аномалій та відхилень від типової поведінки користувачів. До таких методів належать аналіз середніх значень, оцінювання розподілів даних, використання порогових значень та виявлення нетипових операцій. Подібні методи дозволяють автоматизувати частину процесів моніторингу, однак часто демонструють недостатню ефективність при роботі зі складними багатовимірними наборами даних.

Однією з ключових проблем систем виявлення шахрайства є наявність помилкових класифікацій. У практиці фінансового моніторингу виділяють два основні типи помилок: False Positive та False Negative. Помилка типу False Positive виникає тоді, коли легітимна транзакція помилково визначається як шахрайська. У результаті клієнт може зіткнутися з

необґрунтованим блокуванням платежу або необхідністю проходження додаткових процедур підтвердження особи. Помилка типу False Negative є більш небезпечною, оскільки в цьому випадку шахрайська операція залишається невиявленою та успішно виконується [10].

Співвідношення між різними типами помилок є одним із головних критеріїв оцінювання ефективності антишахрайських систем. Зменшення кількості помилкових блокувань дозволяє покращити клієнтський досвід, однак надмірне послаблення контролю може призвести до зростання фінансових втрат. Саме тому сучасні системи виявлення шахрайства використовують складні механізми оцінювання ризиків, які враховують одночасно велику кількість характеристик транзакцій [4].

Для підвищення ефективності контролю фінансові установи дедалі частіше застосовують багаторівневі системи моніторингу. У таких системах кожна транзакція проходить декілька етапів перевірки. На першому рівні використовуються базові правила та порогові значення. Наступні рівні можуть включати статистичний аналіз, оцінювання поведінкових характеристик клієнта та використання інтелектуальних моделей аналізу даних. Подібний підхід дозволяє підвищити точність виявлення шахрайства та зменшити кількість помилкових спрацювань [1], [3].

Основні етапи багаторівневої перевірки фінансових транзакцій доцільно подати на рисунку 1.4.



Рисунок 1.4 – Етапи багаторівневої перевірки транзакцій

Крім того, сучасні фінансові системи дедалі частіше використовують моніторинг у режимі реального часу. Такий підхід дозволяє аналізувати транзакції безпосередньо під час їх виконання та миттєво реагувати на виявлені ризики. В умовах стрімкого розвитку електронної комерції саме швидкість прийняття рішень стає одним із найважливіших факторів ефективності систем боротьби з фінансовим шахрайством [7].

Розвиток технологій аналізу даних та збільшення обсягів фінансової інформації стали передумовою для появи сучасних підходів до виявлення шахрайства. На відміну від традиційних систем, нові методи здатні автоматично аналізувати великі масиви даних, знаходити приховані закономірності та адаптуватися до змін поведінки користувачів. Саме тому сьогодні особливого значення набувають методи машинного навчання, які забезпечують вищу точність виявлення шахрайських операцій та дозволяють ефективно працювати в режимі реального часу [8].

Таким чином, методи виявлення фінансового шахрайства пройшли тривалий шлях розвитку від ручного аналізу операцій до складних автоматизованих систем моніторингу. Кожен із підходів має власні переваги та обмеження, однак сучасні умови функціонування фінансових установ вимагають використання більш гнучких та інтелектуальних механізмів аналізу даних. Це створює передумови для широкого впровадження методів машинного навчання у сфері боротьби з фінансовим шахрайством.

1.3 Використання машинного навчання для задач виявлення шахрайства

Стрімке зростання обсягів фінансових операцій та постійне ускладнення шахрайських схем зумовили необхідність пошуку нових підходів до виявлення підозрілих транзакцій. Традиційні методи контролю, засновані на ручному аналізі або використанні фіксованих правил, поступово втрачають ефективність через обмежену здатність адаптуватися

до нових моделей поведінки шахраїв. У зв'язку з цим особливої актуальності набули методи машинного навчання, які дозволяють автоматично виявляти приховані закономірності в даних та приймати рішення на основі попереднього досвіду [10].

Машинне навчання (Machine Learning) є одним із напрямів штучного інтелекту, що передбачає створення алгоритмів, здатних навчатися на основі наявних даних без явного програмування всіх можливих сценаріїв. Основна ідея машинного навчання полягає у побудові математичних моделей, які здатні виявляти залежності між ознаками об'єктів та

У задачах виявлення фінансового шахрайства методи машинного навчання використовуються для аналізу великої кількості характеристик транзакцій, зокрема суми платежу, часу проведення операції, місця її виконання, типу торгової категорії, демографічних характеристик клієнта та інших факторів. На основі аналізу таких ознак модель визначає ймовірність того, що конкретна операція є шахрайською [6]. Основні напрями використання машинного навчання у задачах виявлення шахрайства наведені на рисунку 1.5.



Рисунок 1.5 – Основні напрямки застосування машинного навчання у системах виявлення шахрайства

Одним із найбільш поширених напрямів застосування машинного навчання у сфері фінансової безпеки є класифікація транзакцій. У даному випадку модель аналізує набір характеристик фінансової операції та визначає її належність до одного з двох класів: легітимна або шахрайська транзакція. Подібний підхід дозволяє автоматизувати процес прийняття рішень та значно скоротити навантаження на співробітників служб безпеки [12].

Не менш важливим напрямом є виявлення аномалій (Anomaly Detection). Основна ідея цього підходу полягає у пошуку транзакцій, характеристики яких суттєво відрізняються від типової поведінки користувача або загальної структури набору даних. Аномальні операції не завжди є шахрайськими, однак вони потребують додаткової перевірки та можуть свідчити про наявність ризиків для фінансової установи.

Окреме місце займає поведінковий аналіз користувачів (Behavioral Analytics). У межах даного підходу система формує профіль звичайної поведінки клієнта на основі історії його фінансових операцій. При цьому враховуються такі характеристики, як середня сума платежів, типові часові проміжки здійснення транзакцій, географічне розташування користувача та категорії товарів або послуг, які найчастіше купуються. Суттєве відхилення від сформованого профілю може розглядатися як ознака потенційного шахрайства [1].

Ще одним важливим напрямом використання машинного навчання є оцінювання ризиків фінансових операцій. У цьому випадку модель не приймає остаточного рішення щодо шахрайського характеру транзакції, а визначає рівень ризику її виконання. Отримана оцінка може використовуватися для подальшої автоматичної перевірки або передаватися співробітникам служби безпеки для детального аналізу [2].

Для вирішення задач виявлення шахрайства використовуються різні типи машинного навчання. Найбільш поширеним є навчання з учителем (Supervised Learning), яке передбачає використання попередньо

розмічених даних. У таких наборах даних кожна транзакція вже має відповідну мітку, що вказує на її належність до легітимних або шахрайських операцій. Під час навчання модель аналізує зв'язки між ознаками транзакцій та їхніми класами, після чого може використовувати отримані закономірності для прогнозування нових спостережень [12].

Навчання без учителя (Unsupervised Learning) використовується у випадках, коли інформація про належність транзакцій до певного класу відсутня. У таких умовах алгоритми самостійно виявляють структуру даних, групують подібні об'єкти або знаходять аномальні спостереження. Подібний підхід особливо корисний для виявлення нових видів шахрайства, які ще не були представлені у навчальних вибірках [4].

У практичних системах боротьби з фінансовим шахрайством дедалі частіше застосовується також напівконтрольоване навчання (Semi-Supervised Learning), яке поєднує переваги двох попередніх підходів. У даному випадку для навчання використовуються як розмічені, так і нерозмічені дані. Це дозволяє підвищити якість побудованих моделей та ефективніше використовувати великі обсяги доступної інформації [12]. Порівняльна характеристика основних типів машинного навчання наведена в таблиці 1.3.

Таблиця 1.3 – Основні типи машинного навчання у задачах виявлення шахрайства

Тип навчання	Особливості	Основна сфера застосування
Supervised Learning	Використання розмічених даних	Класифікація транзакцій
Unsupervised Learning	Відсутність розмітки	Виявлення аномалій
Semi-Supervised Learning	Частково розмічені дані	Аналіз великих наборів даних

Дані таблиці 1.3 демонструють, що вибір конкретного підходу залежить від особливостей доступних даних та поставлених завдань. У сфері фінансового моніторингу найчастіше використовуються алгоритми навчання з учителем, оскільки вони забезпечують високу точність класифікації за наявності історичних даних про шахрайські операції [10].

1.4 Аналіз існуючих рішень у сфері виявлення фінансового шахрайства

Проблема виявлення фінансового шахрайства протягом останніх років привертає значну увагу як наукової спільноти, так і представників фінансового сектору. Постійне зростання кількості електронних транзакцій, розвиток цифрового банкінгу та поява нових схем шахрайства стимулювали активний розвиток систем моніторингу фінансових операцій. У результаті було запропоновано значну кількість підходів, які відрізняються використовуваними алгоритмами, набором ознак, способами підготовки даних та методами оцінювання ефективності моделей [4].

Перші автоматизовані системи виявлення шахрайства базувалися переважно на використанні експертних правил та статистичних методів аналізу. Однак подальші дослідження показали, що подібні рішення мають обмежену здатність адаптуватися до нових моделей поведінки шахраїв. Саме тому значна частина сучасних наукових робіт присвячена застосуванню методів машинного навчання для автоматичного виявлення підозрілих транзакцій [1].

Важливим етапом розвитку систем Fraud Detection стало впровадження підходів, орієнтованих на аналіз великих обсягів історичних транзакцій. У роботі «Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy» автори запропонували підхід, який враховує особливості реальних фінансових даних та часову залежність між транзакціями. Дослідники зазначають, що багато моделей демонструють

високі результати на штучно підготовлених вибірках, однак їх ефективність може суттєво знижуватися під час використання реальних потоків фінансових операцій [13].

Особливу увагу в сучасних дослідженнях приділяють проблемі побудови відтворюваних моделей машинного навчання. У роботі «Reproducible Machine Learning for Credit Card Fraud Detection – Practical Handbook» запропоновано комплексний підхід до побудови антифрод-систем, який охоплює всі етапи життєвого циклу моделі: від збору та підготовки даних до оцінювання результатів класифікації та впровадження готового рішення у виробниче середовище [14]. Автори підкреслюють важливість забезпечення відтворюваності результатів досліджень, оскільки це дозволяє об'єктивно порівнювати різні алгоритми та підходи.

Значний інтерес становлять роботи, присвячені використанню глибокого навчання для задач виявлення шахрайства. Зокрема, у дослідженні «Anomaly Detection in Financial Transactions Using Convolutional Neural Networks» розглянуто можливість застосування згорткових нейронних мереж для аналізу фінансових операцій [15]. Отримані результати демонструють, що глибокі нейронні мережі здатні ефективно виявляти складні нелінійні залежності між ознаками транзакцій та забезпечувати високу точність класифікації. Водночас подібні рішення потребують значних обчислювальних ресурсів та великих обсягів навчальних даних.

Важливим напрямом розвитку сучасних систем виявлення фінансового шахрайства є використання поведінкового аналізу користувачів. На відміну від традиційних підходів, які оцінюють окремі характеристики транзакції, поведінкові моделі враховують довгострокову історію дій клієнта та формують його індивідуальний профіль. До таких характеристик можуть належати середня сума операцій, часові інтервали між платежами, типові географічні локації, категорії товарів та послуг, а також особливості використання банківських сервісів [6].

Перевагою поведінкового аналізу є здатність виявляти шахрайські операції навіть у тих випадках, коли вони формально не порушують встановлені правила або не містять явних аномалій. Наприклад, транзакція може бути здійснена на відносно невелику суму та не перевищувати жодних лімітів, проте суттєво відрізнятися від звичної поведінки конкретного клієнта. У такому випадку система може присвоїти операції підвищений рівень ризику та ініціювати додаткову перевірку [5].

У сучасних антифрод-рішеннях дедалі частіше використовуються комбіновані підходи, які поєднують декілька методів аналізу даних. Подібні системи можуть одночасно застосовувати експертні правила, статистичні методи, алгоритми машинного навчання та поведінковий аналіз користувачів. Поєднання різних підходів дозволяє компенсувати недоліки окремих методів та підвищити загальну ефективність виявлення шахрайських операцій.

Ще однією тенденцією розвитку сучасних антифрод-рішень є використання систем моніторингу в режимі реального часу. У традиційних підходах аналіз операцій часто здійснювався після завершення транзакції, що призводило до виникнення фінансових втрат ще до моменту виявлення шахрайства. Сучасні системи дозволяють оцінювати ризик практично миттєво, приймаючи рішення протягом часток секунди [3].

Особливо важливим моніторинг у режимі реального часу є для електронної комерції та мобільного банкінгу, де швидкість обробки транзакцій безпосередньо впливає на якість обслуговування клієнтів. Саме тому значна частина сучасних досліджень спрямована на розробку алгоритмів, здатних забезпечувати високу точність класифікації за мінімального часу прийняття рішення [4].

Окрему увагу дослідники приділяють питанням інтерпретованості моделей машинного навчання. Для фінансових установ важливо не лише отримати результат класифікації, але й розуміти причини, через які певна транзакція була визначена як підозріла. Особливо актуальною ця проблема

є для складних моделей глибокого навчання, які часто демонструють високу точність, але характеризуються низьким рівнем прозорості процесу прийняття рішень [15].

У зв'язку з цим у багатьох сучасних дослідженнях розглядаються підходи Explainable Artificial Intelligence (XAI), що дозволяють пояснювати результати роботи моделей та підвищувати рівень довіри до автоматизованих систем виявлення шахрайства. Використання інтерпретованих моделей є важливою умовою їх практичного впровадження у банківській сфері та фінансовому секторі загалом.

Окремий напрям досліджень пов'язаний із використанням ансамблевих методів машинного навчання. Особливу популярність у сфері виявлення фінансового шахрайства отримав алгоритм Random Forest, який поєднує результати роботи великої кількості дерев рішень. Завдяки високій стійкості до перенавчання та здатності працювати з великою кількістю ознак даний алгоритм активно використовується як у наукових дослідженнях, так і в практичних антифрод-рішеннях [16].

У сучасних роботах також приділяється значна увага проблемі незбалансованості даних. Для фінансових транзакцій характерна ситуація, коли кількість шахрайських операцій становить лише незначну частину від загального обсягу даних. Це створює додаткові труднощі під час навчання моделей та може негативно впливати на якість класифікації. Для вирішення даної проблеми використовуються методи Oversampling, Undersampling та SMOTE, які дозволяють підвищити якість прогнозування рідкісних подій [17].

Таким чином, сучасні дослідження у сфері виявлення фінансового шахрайства спрямовані на підвищення точності моделей, забезпечення інтерпретованості результатів та ефективну роботу з великими й незбалансованими наборами даних. Комплексне поєднання цих підходів дозволяє створювати більш надійні та адаптивні системи моніторингу

фінансових транзакцій. Порівняльну характеристику основних напрямів сучасних досліджень доцільно подати у таблиці 1.4.

Таблиця 1.4 – Порівняння сучасних підходів до виявлення фінансового шахрайства

Підхід	Основна ідея	Переваги	Недоліки
Rule-Based Systems	Набір експертних правил	Простота реалізації	Низька адаптивність
Статистичні методи	Аналіз відхилень від норми	Висока швидкість роботи	Обмежена точність
Random Forest	Ансамбль дерев рішень	Висока точність	Складність інтерпретації
Нейронні мережі	Аналіз складних залежностей	Висока гнучкість	Значні обчислювальні витрати

Аналіз даних, наведених у таблиці 1.4, показує, що універсального підходу до виявлення фінансового шахрайства наразі не існує. Кожен із методів має власні переваги та обмеження, а ефективність його використання значною мірою залежить від особливостей набору даних та специфіки фінансових операцій.

Аналіз сучасних наукових досліджень показує, що ефективність систем виявлення шахрайства оцінюється не лише за загальною точністю класифікації. У зв'язку з високою незбалансованістю фінансових даних дослідники все частіше використовують додаткові метрики оцінювання, зокрема Precision, Recall, F1-score та ROC-AUC [14].

Особливе значення має показник Recall, який характеризує здатність моделі виявляти шахрайські транзакції. Для фінансових установ пропуск навіть незначної кількості шахрайських операцій може призводити до суттєвих фінансових втрат. Водночас надмірне збільшення чутливості

моделі може спричиняти зростання кількості помилкових спрацювань, що негативно впливає на користувацький досвід клієнтів.

Саме тому сучасні антифрод-системи намагаються забезпечити баланс між ефективністю виявлення шахрайства та мінімізацією кількості помилкових блокувань легітимних транзакцій. Досягнення такого балансу є одним із ключових завдань під час розробки та впровадження систем фінансового моніторингу [14].

Крім безпосереднього виявлення шахрайських операцій, сучасні дослідження значну увагу приділяють питанням масштабованості та продуктивності антифрод-систем. Із постійним зростанням кількості електронних платежів фінансові установи стикаються з необхідністю обробки мільйонів транзакцій щоденно. У зв'язку з цим важливого значення набувають алгоритми, здатні забезпечувати високу швидкість аналізу даних без суттєвого зниження точності класифікації. Особливо актуальною дана проблема є для великих банківських установ та платіжних систем, де навіть незначні затримки під час обробки транзакцій можуть негативно впливати на якість обслуговування клієнтів.

Ще одним перспективним напрямом розвитку є інтеграція систем виявлення шахрайства з технологіями штучного інтелекту та аналітичними платформами підтримки прийняття рішень. Подібні рішення дозволяють не лише автоматично визначати підозрілі транзакції, але й формувати рекомендації для співробітників служб безпеки, оцінювати рівень ризику клієнтів та прогнозувати потенційні загрози. Використання таких підходів сприяє підвищенню ефективності фінансового моніторингу та забезпечує більш комплексний захист банківських систем від шахрайських дій.

Окрему увагу сучасні дослідження приділяють питанням автоматизації процесів прийняття рішень у сфері фінансового моніторингу. Використання інтелектуальних систем дозволяє суттєво зменшити навантаження на працівників служб безпеки, автоматизувати первинний аналіз транзакцій та скоротити час реагування на потенційні загрози.

Завдяки цьому фінансові установи можуть більш ефективно використовувати наявні ресурси та підвищувати якість контролю фінансових операцій.

Важливим фактором розвитку сучасних антифрод-систем стало поширення відкритих наборів даних. Одним із найбільш популярних є Credit Card Transactions Fraud Detection Dataset, який активно використовується для навчання та тестування моделей машинного навчання [18]. Наявність відкритих даних значно спростила проведення досліджень та сприяла розвитку нових методів виявлення шахрайських операцій (рисунок 1.6).



Рисунок 1.6 – Еволюція підходів до виявлення фінансового шахрайства

Таким чином, аналіз сучасних досліджень свідчить про поступовий перехід від традиційних методів контролю до інтелектуальних систем, заснованих на алгоритмах машинного навчання. Найбільш перспективними напрямами розвитку залишаються ансамблеві методи, нейронні мережі та системи аналізу даних у режимі реального часу. Саме ці підходи сьогодні формують основу сучасних рішень для боротьби з фінансовим шахрайством та визначають подальший розвиток антифрод-технологій.

2 ПРОЄКТУВАННЯ СИСТЕМИ МОНІТОРИНГУ ШАХРАЙСЬКИХ ТРАНЗАКЦІЙ

2.1 Постановка задачі

Сучасний фінансовий сектор характеризується активним впровадженням цифрових технологій та постійним зростанням кількості електронних фінансових операцій. Банківські установи, платіжні системи та сервіси електронної комерції щоденно обробляють значні обсяги транзакцій, що забезпечує високу швидкість проведення платежів та зручність використання фінансових послуг. Водночас збільшення кількості цифрових операцій супроводжується зростанням ризиків фінансового шахрайства, що створює додаткові виклики для фінансових організацій.

Традиційні методи контролю транзакцій, які базуються на ручному аналізі або використанні фіксованих наборів правил, не завжди дозволяють своєчасно виявляти нові шахрайські схеми. Зловмисники постійно вдосконалюють способи обходу існуючих механізмів безпеки, що знижує ефективність статичних систем контролю. У таких умовах виникає потреба у використанні інтелектуальних засобів аналізу даних, здатних автоматично виявляти підозрілі закономірності та оцінювати ризик фінансових операцій.

Одним із перспективних напрямів вирішення даної проблеми є використання методів машинного навчання для автоматизованого аналізу транзакцій. На відміну від традиційних підходів, моделі машинного навчання можуть виявляти складні залежності між характеристиками фінансових операцій та адаптуватися до змін у структурі даних. Це дозволяє підвищити ефективність виявлення шахрайських дій та зменшити кількість пропущених підозрілих операцій.

Метою проєктування є створення системи моніторингу шахрайських транзакцій, яка забезпечує автоматичний аналіз фінансових операцій із

використанням моделі машинного навчання та надає інструменти для контролю результатів моніторингу.

Для досягнення поставленої мети необхідно вирішити такі завдання:

– визначити вимоги до системи моніторингу шахрайських транзакцій;

– дослідити набір даних для навчання моделі машинного навчання;

– обрати алгоритм класифікації фінансових операцій;

– спроєктувати архітектуру програмної системи;

– спроєктувати структуру бази даних для збереження результатів аналізу;

– забезпечити можливість моніторингу транзакцій та формування повідомлень про підозрілі операції.

Об'єктом проєктування є система моніторингу фінансових транзакцій.

Предметом проєктування є методи та програмні засоби автоматизованого виявлення шахрайських фінансових операцій на основі технологій машинного навчання.

У процесі проєктування передбачається створення системи, яка дозволить виконувати автоматичну оцінку ризику фінансових транзакцій, накопичувати результати аналізу, формувати попередження про потенційно небезпечні операції та надавати користувачу засоби для перегляду статистичної та аналітичної інформації (рисунок 2.1).

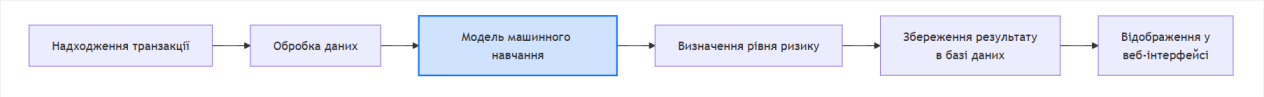


Рисунок 2.1 – Загальна схема функціонування системи моніторингу шахрайських транзакцій

Основним вхідним потоком даних для системи є фінансові транзакції, що містять набір характеристик, необхідних для оцінювання рівня ризику. Після надходження даних виконується їх обробка та підготовка до аналізу.

Далі сформований набір ознак передається до моделі машинного навчання, яка визначає ймовірність шахрайства та формує результат класифікації.

Результати аналізу використовуються для подальшого моніторингу та зберігаються у базі даних. Це забезпечує можливість накопичення історії транзакцій, проведення статистичного аналізу та формування повідомлень про підозрілі операції. Наявність централізованого сховища даних також створює основу для подальшого розширення функціональності системи.

Під час проєктування було сформовано перелік функціональних вимог до майбутньої системи. До основних функціональних вимог належать:

– автоматичне визначення рівня ризику фінансової транзакції;

– збереження інформації про виконані перевірки;

– формування повідомлень про високоризикові операції;

– ведення історії транзакцій;

– ведення переліку заблокованих рахунків;

– підтримка авторизації користувачів;

– відображення статистичної та аналітичної інформації.

Крім функціональних вимог, система повинна відповідати низці нефункціональних вимог. До них належать забезпечення прийнятної швидкості обробки транзакцій, надійність зберігання даних, зручність використання інтерфейсу, можливість подальшого розширення функціональності та підтримка роботи з великими обсягами даних.

Таким чином, у межах даного проєкту поставлено задачу створення системи моніторингу шахрайських транзакцій, яка поєднує можливості сучасних методів машинного навчання та інструментів аналізу даних для автоматизованого виявлення потенційно небезпечних фінансових операцій. Результати виконання даного етапу проєктування є основою для подальшого вибору набору даних, алгоритму машинного навчання, архітектури програмної системи та структури бази даних.

2.2 Аналіз та підготовка набору даних

Якість роботи моделей машинного навчання значною мірою залежить від характеристик набору даних, що використовується під час навчання. У задачах виявлення фінансового шахрайства особливе значення мають обсяг даних, різноманітність ознак, наявність прикладів шахрайських операцій та відповідність набору даних реальним умовам функціонування фінансових систем. Саме тому одним із ключових етапів проєктування системи моніторингу шахрайських транзакцій є вибір та підготовка навчального набору даних.

Для досліджень у сфері Fraud Detection існує декілька популярних відкритих наборів даних. Найбільш відомими є Credit Card Fraud Detection Dataset, IEEE-CIS Fraud Detection Dataset, PaySim Mobile Money Dataset та Credit Card Transactions Fraud Detection Dataset. Кожен із них має власні особливості, переваги та обмеження.

Більшість доступних наборів даних містить інформацію про фінансові операції, однак деякі з них мають суттєво обмежений набір характеристик або використовують попередньо анонімізовані ознаки, що ускладнює інтерпретацію результатів аналізу. Для проєктування системи моніторингу шахрайських транзакцій було важливо використовувати набір даних, який максимально наближений до реальних умов функціонування фінансових установ та містить як фінансові, так і поведінкові характеристики користувачів.

З урахуванням зазначених вимог для виконання роботи було обрано набір даних Credit Card Transactions Fraud Detection Dataset, розміщений на платформі Kaggle [18]. Даний набір даних містить інформацію про реальні фінансові транзакції та включає широкий набір характеристик, які можуть використовуватися для побудови моделей машинного навчання.

Основними причинами вибору саме цього набору даних стали:

– наявність великої кількості записів;

– наявність позначок шахрайських та легітимних операцій;

– широкий набір числових та категоріальних ознак;

– можливість формування додаткових характеристик транзакцій;

– популярність набору даних серед дослідників у сфері виявлення фінансового шахрайства;

– доступність для відтворення результатів дослідження.

Загальну характеристику використаного набору даних наведено в таблиці 2.1.

Таблиця 2.1 – Основні характеристики використаного набору даних

Характеристика	Значення
Джерело даних	Kaggle
Назва набору	Credit Card Transactions Fraud Detection Dataset
Тип задачі	Бінарна класифікація
Цільова ознака	is_fraud
Тип даних	Транзакції платіжних карток
Наявність шахрайських операцій	Так

Набір даних містить інформацію про фінансові операції, клієнтів банку, торговельні категорії та географічні характеристики транзакцій. Серед доступних полів присутні сума платежу, дата та час проведення операції, стать користувача, чисельність населення населеного пункту, географічні координати клієнта та продавця, а також інші параметри, що можуть бути використані для побудови моделі виявлення шахрайства.

Перед початком навчання моделі було виконано відбір ознак, які потенційно можуть впливати на ймовірність виникнення шахрайської операції. При цьому використовувалися як початкові характеристики набору даних, так і додаткові ознаки, сформовані під час попередньої

обробки даних. Для навчання моделі були використані наступні ознаки, зазначені в таблиці 2.2.

Таблиця 2.2 – Ознаки обрані для навчання моделі

Ознака	Опис
amt	сума транзакції
city_pop	чисельність населення міста
lat, long	координати клієнта
merch_lat, merch_long	координати торгової точки
age	вік клієнта
hour	година виконання операції
day_of_week	день тижня
month	місяць виконання транзакції
gender	стать користувача
category	категорія торгової операції

На етапі підготовки даних було виконано перетворення часових характеристик транзакцій. Поле trans_date_trans_time містило дату та час проведення операції. На його основі були сформовані додаткові ознаки hour, day_of_week та month, які дозволяють враховувати часові закономірності поведінки користувачів. Програмний код реалізації часових ознак поданий у лістингу 2.1.

Лістинг 2.1 – Формування часових ознак на основі дати транзакції

```
df['hour'] = df['trans_date_trans_time'].dt.hour

df['day_of_week'] = (
    df['trans_date_trans_time']
    .dt.dayofweek
)

df['month'] = (
    df['trans_date_trans_time']
    .dt.month
)
```

Крім того, на основі дати народження користувача було розраховано вік клієнта. Для цього використовувалися поля trans_date_trans_time та dob. Отримана ознака age дозволяє враховувати можливий вплив вікових особливостей клієнтів на характер здійснюваних фінансових операцій. Програмний код формування цієї ознаки поданий у лістингу 2.2.

Лістинг 2.2 – Розрахунок віку клієнта

```
df['age'] = (
    df['trans_date_trans_time'].dt.year
    - df['dob'].dt.year
)
```

Оскільки алгоритми машинного навчання не можуть безпосередньо працювати з текстовими категоріальними значеннями, було виконано кодування ознак gender та category. Для цього використовувався алгоритм LabelEncoder із бібліотеки Scikit-learn. У результаті кожному категоріальному значенню було поставлено у відповідність числовий код, що дозволило використовувати ці характеристики під час навчання моделі. Програмний код кодування категоріальних ознак поданий у лістингу 2.3.

Лістинг 2.3 – Кодування категоріальних ознак

```
gender_encoder = LabelEncoder()
category_encoder = LabelEncoder()

df['gender'] = gender_encoder.fit_transform(
    df['gender']
)

df['category'] = category_encoder.fit_transform(
    df['category']
)
```

Після завершення підготовки даних було сформовано матрицю ознак X та цільовий вектор y. Цільовою змінною виступала ознака is_fraud, яка містить інформацію про належність транзакції до шахрайських або легітимних операцій.

Для оцінювання якості майбутньої моделі набір даних було розділено на навчальну та тестову вибірки. Розподіл здійснювався у співвідношенні 80% до 20%. Під час поділу використовувався механізм стратифікації, що дозволило зберегти початкове співвідношення класів у кожній вибірці та забезпечити коректне оцінювання результатів навчання.

Таким чином, у результаті аналізу та підготовки набору даних було сформовано набір ознак, придатний для використання в задачі класифікації фінансових транзакцій. Отримані дані стали основою для подальшого навчання моделі машинного навчання та побудови системи моніторингу шахрайських операцій.

2.3 Вибір алгоритму машинного навчання

Одним із ключових етапів проєктування системи моніторингу шахрайських транзакцій є вибір алгоритму машинного навчання, який буде використовуватися для аналізу фінансових операцій та визначення ймовірності шахрайства. Від правильності вибору алгоритму значною мірою залежить якість класифікації транзакцій, швидкість обробки даних та можливість практичного використання системи в умовах реального фінансового середовища.

Задача виявлення фінансового шахрайства належить до задач бінарної класифікації, оскільки кожна транзакція повинна бути віднесена до одного з двох класів: легітимна або шахрайська операція. Для вирішення подібних задач у сучасній практиці використовуються різні алгоритми машинного навчання, серед яких найбільшого поширення набули логістична регресія, дерева рішень, метод опорних векторів, ансамблеві методи та нейронні мережі [19].

Одним із найпростіших алгоритмів класифікації є логістична регресія (Logistic Regression). Даний алгоритм використовується для оцінювання ймовірності належності об'єкта до певного класу на основі

набору вхідних ознак. Основною перевагою логістичної регресії є простота реалізації та висока інтерпретованість результатів. Водночас ефективність даного підходу може знижуватися у випадках, коли між ознаками та цільовою змінною існують складні нелінійні залежності [20].

Іншим популярним підходом є використання дерев рішень (Decision Trees). Алгоритм послідовно розділяє набір даних на підмножини на основі значень окремих ознак, формуючи структуру дерева, яка використовується для класифікації нових об'єктів. Основними перевагами дерев рішень є наочність, простота інтерпретації та можливість роботи як із числовими, так і з категоріальними даними [21].

Незважаючи на переваги дерев рішень, вони мають схильність до перенавчання, особливо у випадках використання великих наборів даних із значною кількістю ознак. Для зменшення впливу даної проблеми були розроблені ансамблеві методи машинного навчання, які поєднують результати роботи великої кількості окремих моделей [22].

До числа таких алгоритмів належить Random Forest, який є одним із найбільш популярних методів класифікації у задачах виявлення фінансового шахрайства. Основна ідея алгоритму полягає у побудові великої кількості дерев рішень та об'єднанні їх прогнозів для отримання кінцевого результату. Кожне дерево навчається на випадковій підмножині даних та використовує випадково обрані ознаки, що дозволяє підвищити стійкість моделі та зменшити ризик перенавчання [16].

Ще одним поширеним підходом є використання методу опорних векторів (Support Vector Machine, SVM). Алгоритм виконує побудову роздільної поверхні між класами та забезпечує високу якість класифікації для багатьох типів задач. Однак при роботі з великими наборами даних та значною кількістю ознак обчислювальна складність SVM може суттєво зростати, що ускладнює його використання у системах моніторингу реального часу [19].

У сучасних дослідженнях також широко застосовуються нейронні мережі та методи глибокого навчання. Подібні алгоритми здатні виявляти складні нелінійні закономірності та демонструють високі показники точності класифікації. Проте їх використання часто потребує значних обчислювальних ресурсів, великого обсягу навчальних даних та складнішої процедури налаштування моделі [19].

Порівняльну характеристику розглянутих алгоритмів наведено в таблиці 2.3.

Таблиця 2.3 – Порівняння алгоритмів машинного навчання для задач виявлення шахрайства

Алгоритм	Переваги	Недоліки
Logistic Regression	Простота реалізації, інтерпретованість	Обмежена здатність виявляти складні залежності
Decision Tree	Наочність, простота аналізу	Схильність до перенавчання
SVM	Висока точність класифікації	Значні обчислювальні витрати
Neural Networks	Виявлення складних закономірностей	Високі вимоги до ресурсів
Random Forest	Висока точність, стійкість до перенавчання	Складніша інтерпретація результатів

Аналіз даних, наведених у таблиці 2.3, показує, що жоден із алгоритмів не є універсальним рішенням для всіх задач класифікації. Вибір конкретного методу залежить від структури даних, обсягу навчальної вибірки, вимог до швидкості роботи системи та необхідного рівня точності.

Для проєктованої системи моніторингу шахрайських транзакцій було обрано алгоритм Random Forest. Дане рішення обумовлене декількома факторами. По-перше, алгоритм ефективно працює з табличними даними, які становлять основу використаного набору транзакцій. По-друге, Random Forest здатний обробляти велику кількість ознак без необхідності складної

попередньої підготовки даних. По-третє, ансамблева структура моделі дозволяє зменшити ризик перенавчання та забезпечує стабільні результати класифікації навіть за наявності шуму у даних [16].

Крім того, важливою перевагою Random Forest є можливість роботи з незбалансованими наборами даних, що є характерною особливістю задач виявлення фінансового шахрайства. Оскільки кількість шахрайських транзакцій зазвичай значно менша за кількість легітимних операцій, використання алгоритму, стійкого до дисбалансу класів, має особливе значення для забезпечення високої якості прогнозування [16].

Загальну схему роботи алгоритму Random Forest доцільно подати на рисунку 2.2.

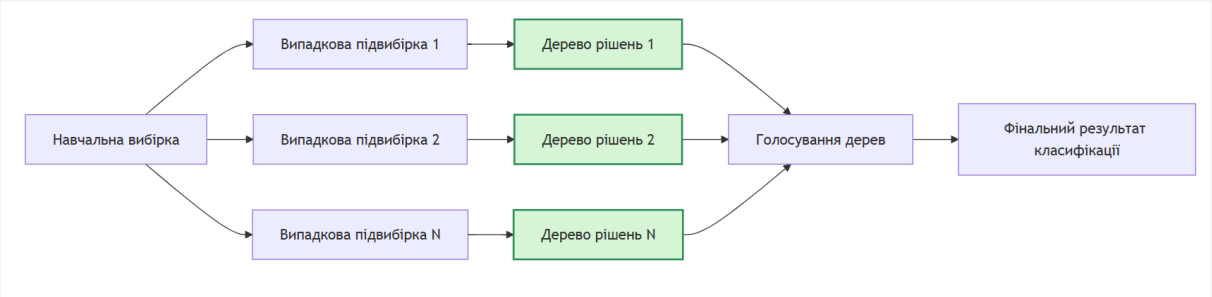


Рисунок 2.2 – Принцип роботи алгоритму Random Forest

Таким чином, у результаті аналізу сучасних алгоритмів машинного навчання для подальшого використання у системі моніторингу шахрайських транзакцій було обрано алгоритм Random Forest. Поєднання високої точності класифікації, стійкості до перенавчання, здатності працювати з великими обсягами даних та відносної простоти практичного використання робить його доцільним рішенням для задач виявлення фінансового шахрайства.

2.4 Проєктування архітектури системи

Одним із ключових етапів створення програмного забезпечення є проєктування архітектури системи. Від якості архітектурних рішень залежить зручність подальшої розробки, можливість масштабування програмного забезпечення, підтримка окремих модулів та ефективність взаємодії між компонентами системи.

Під час проєктування системи моніторингу шахрайських транзакцій було поставлено завдання створити програмний комплекс, який забезпечуватиме автоматичний аналіз фінансових операцій, збереження результатів перевірки та надання користувачу засобів моніторингу через веб-інтерфейс. Для досягнення цієї мети було обрано модульну архітектуру, у межах якої окремі компоненти відповідають за виконання визначених функцій та взаємодіють між собою через чітко визначені інтерфейси.

Використання модульного підходу дозволяє зменшити залежність між окремими частинами системи та спрощує її подальше розширення. У разі необхідності окремі модулі можуть бути модифіковані або замінені без суттєвого впливу на роботу інших компонентів програмного забезпечення. Загальну архітектуру системи наведено на рисунку 2.3.

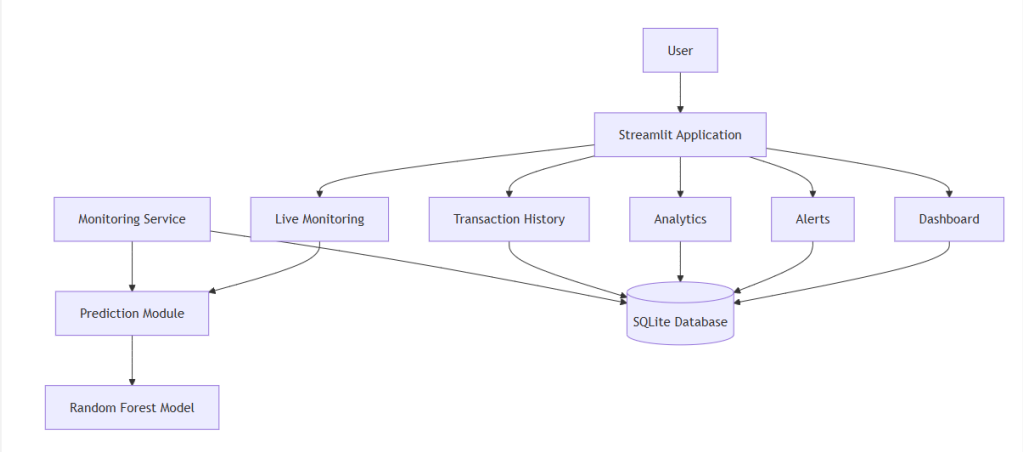


Рисунок 2.3 – Загальна архітектура системи моніторингу шахрайських транзакцій

Центральним елементом архітектури є модуль прогнозування, який виконує аналіз характеристик фінансової транзакції та визначає ймовірність її належності до шахрайських операцій. Для виконання прогнозування використовується попередньо навчена модель машинного навчання. Модуль прогнозування отримує набір ознак транзакції, виконує їх підготовку та передає сформовані дані до моделі для отримання результату класифікації.

Окремим компонентом системи є модуль моніторингу транзакцій. Його основним призначенням є забезпечення безперервного надходження та аналізу фінансових операцій. У процесі роботи модуль отримує інформацію про транзакції, передає її до модуля прогнозування та отримує результати оцінювання ризику. Після завершення аналізу інформація про транзакцію зберігається у базі даних для подальшого використання.

Важливу роль у роботі системи відіграє база даних, яка використовується для централізованого зберігання інформації. База даних забезпечує накопичення історії транзакцій, зберігання повідомлень про підозрілі операції, ведення списку заблокованих рахунків та зберігання інформації про користувачів системи. Використання єдиного сховища даних дозволяє забезпечити цілісність інформації та спрощує взаємодію між окремими компонентами програмного забезпечення.

Для взаємодії користувача із системою передбачено веб-інтерфейс. Через нього користувач отримує доступ до основних функцій моніторингу, перегляду історії транзакцій, аналізу статистичної інформації та керування повідомленнями про потенційно небезпечні операції. Веб-інтерфейс виконує роль посередника між користувачем та внутрішніми компонентами системи.

Головною сторінкою системи є інформаційна панель моніторингу, яка відображає узагальнену статистику роботи системи. На даній сторінці користувач може переглядати кількість оброблених транзакцій, кількість

активних попереджень, інформацію про заблоковані рахунки та розподіл операцій за рівнями ризику.

Для забезпечення оперативного контролю фінансових операцій у системі передбачено сторінку моніторингу транзакцій. Даний модуль призначений для відображення поточних результатів аналізу та контролю процесу надходження нових транзакцій.

Окремі компоненти інтерфейсу відповідають за роботу з історією транзакцій, повідомленнями про ризикові операції та аналітичними звітами. Поділ функціональності на окремі сторінки дозволяє спростити навігацію та забезпечує зручність використання програмного забезпечення.

Для забезпечення безпеки доступу до системи передбачено механізм автентифікації користувачів. Перед початком роботи користувач повинен пройти процедуру авторизації шляхом введення облікових даних. Після успішної перевірки система надає доступ до відповідних функцій залежно від ролі користувача. Послідовність взаємодії основних компонентів системи наведено на рисунку 2.4.

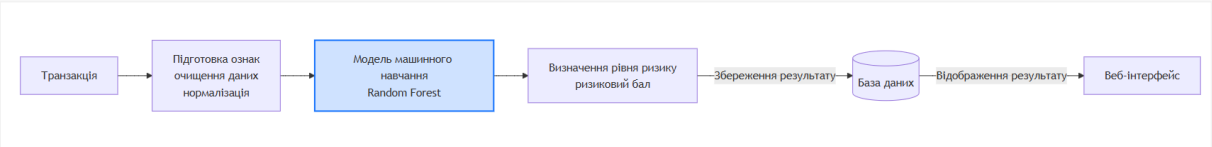


Рисунок 2.4 – Послідовність обробки транзакції в системі

Під час проєктування особлива увага приділялася забезпеченню можливості подальшого розширення функціональності системи. Запропонована архітектура дозволяє інтегрувати нові алгоритми машинного навчання, розширювати структуру бази даних та додавати нові модулі без суттєвого перепроєктування вже існуючих компонентів.

Таким чином, у результаті проєктування було сформовано модульну архітектуру системи моніторингу шахрайських транзакцій, яка забезпечує

взаємодію між модулем прогнозування, базою даних, системою моніторингу та веб-інтерфейсом користувача. Запропонована структура створює основу для подальшої реалізації програмного забезпечення та забезпечує можливість його розвитку в майбутньому.

2.5 Проєктування бази даних

Ефективне функціонування системи моніторингу шахрайських транзакцій неможливе без надійного механізму зберігання та обробки інформації. У процесі роботи система повинна накопичувати результати аналізу транзакцій, зберігати інформацію про користувачів, повідомлення про підозрілі операції та дані щодо заблокованих рахунків. Для вирішення цих завдань у межах проєкту було спроєктовано реляційну базу даних.

Основною метою проєктування бази даних є забезпечення централізованого зберігання інформації, необхідної для функціонування системи моніторингу. Використання єдиного сховища даних дозволяє забезпечити цілісність інформації, спростити взаємодію між окремими компонентами системи та створити основу для подальшого розширення функціональності програмного забезпечення.

Під час проєктування бази даних було визначено перелік сутностей, необхідних для реалізації функціональних можливостей системи. Аналіз вимог дозволив виділити такі основні категорії даних:

– інформація про користувачів системи;

– інформація про фінансові транзакції;

– інформація про повідомлення щодо ризикових операцій;

– інформація про заблоковані рахунки;

– інформація про пропущені попередження.

Виділені категорії даних охоплюють усі основні інформаційні потоки, які виникають під час роботи системи моніторингу шахрайських транзакцій. Такий підхід дозволяє забезпечити структуроване зберігання інформації,

спростити її подальшу обробку та створити умови для ефективної взаємодії між окремими програмними модулями. Крім того, чіткий розподіл даних за сутностями сприяє підвищенню надійності системи та полегшує її супровід у процесі експлуатації.

На основі виділених сутностей було сформовано структуру бази даних, яка складається з п'яти основних таблиць (рисунок 2.5).

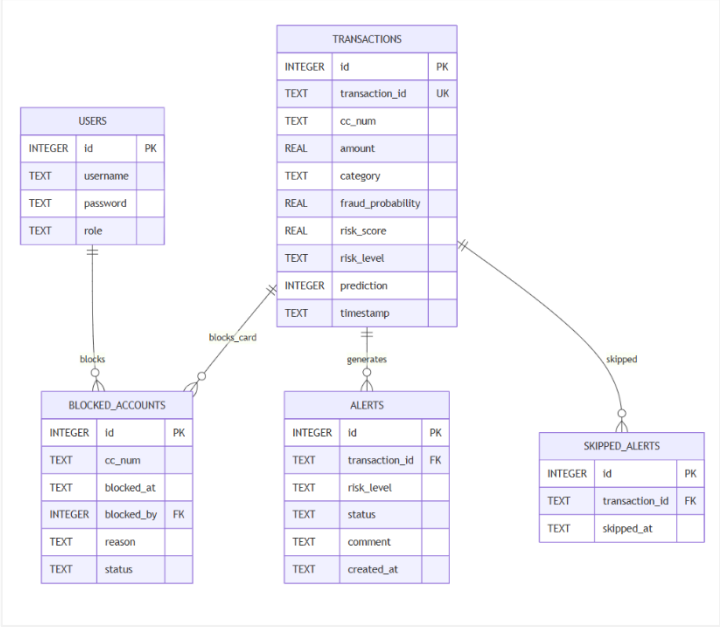


Рисунок 2.5 – ER-діаграма бази даних системи моніторингу шахрайських транзакцій

Центральним елементом бази даних є таблиця TRANSACTIONS, яка призначена для зберігання інформації про всі оброблені фінансові операції. Для кожної транзакції зберігається її унікальний ідентифікатор, номер платіжної картки, сума операції, категорія платежу, результат прогнозування моделі машинного навчання та рівень ризику. Таблиця TRANSACTIONS є основним джерелом інформації для формування статистичних звітів, побудови аналітичних показників та відображення історії фінансових операцій у користувацькому інтерфейсі.

Для забезпечення контролю доступу до системи використовується таблиця USERS. У ній зберігаються облікові записи користувачів, необхідні для проходження процедури авторизації та визначення рівня доступу до функціональних можливостей програмного забезпечення. Використання ролей дозволяє розмежувати права доступу до окремих функцій системи та забезпечити додатковий рівень безпеки.

Для роботи з підозрілими фінансовими операціями передбачено таблицю ALERTS. Вона використовується для зберігання інформації про транзакції, які були визначені як потенційно небезпечні та потребують додаткового аналізу. Наявність окремої таблиці для попереджень дозволяє відокремити критично важливі події від загального потоку транзакцій та забезпечити їх оперативне опрацювання.

Для реалізації механізму блокування рахунків використовується таблиця BLOCKED_ACCOUNTS. У ній зберігається інформація про рахунки, щодо яких було прийнято рішення про обмеження подальших фінансових операцій.

Окремо у структурі бази даних передбачено таблицю SKIPPED_ALERTS, яка використовується для зберігання інформації про попередження, що були переглянуті та відхилені користувачем без виконання додаткових дій. Подібний підхід дозволяє зберігати історію прийнятих рішень та уникати повторного відображення вже оброблених повідомлень.

Під час проєктування особлива увага приділялася забезпеченню простоти структури бази даних. Оскільки система орієнтована на виконання моніторингу транзакцій та не передбачає зберігання великої кількості взаємопов'язаних бізнес-сутностей, використання відносно компактної структури дозволяє спростити підтримку програмного забезпечення та підвищити швидкодію операцій читання і запису даних.

Крім того, обрана структура створює передумови для подальшого розширення функціональності системи. За необхідності можуть бути додані

нові таблиці для зберігання журналів подій, результатів аудиту, додаткової інформації про користувачів або розширених аналітичних показників без необхідності кардинальної зміни існуючої архітектури.

Важливим аспектом проєктування бази даних є забезпечення цілісності та узгодженості інформації. Для цього між таблицями передбачаються логічні зв'язки, що дозволяють уникнути дублювання даних та забезпечити коректність виконання операцій над інформацією. Використання первинних ключів забезпечує унікальну ідентифікацію записів, а зовнішні ключі дозволяють підтримувати зв'язки між транзакціями, попередженнями та обліковими записами користувачів.

Під час розробки структури бази даних також враховувалися вимоги до продуктивності системи. Оскільки моніторинг фінансових операцій передбачає постійне надходження нових транзакцій, важливо забезпечити швидке виконання операцій пошуку, фільтрації та оновлення даних. Для цього можуть використовуватися механізми індексування найбільш важливих полів, зокрема ідентифікаторів транзакцій, номерів платіжних карток та статусів обробки попереджень.

Додатковою перевагою розробленої структури є можливість накопичення історичних даних для подальшого аналізу та вдосконалення моделей машинного навчання. Інформація про транзакції, рівні ризику та результати перевірок може використовуватися для повторного навчання моделей, оцінювання їх ефективності та формування нових аналітичних звітів. Це створює основу для постійного підвищення якості виявлення шахрайських операцій у майбутньому.

Таким чином, у результаті проєктування було сформовано структуру реляційної бази даних, яка забезпечує зберігання всієї інформації, необхідної для функціонування системи моніторингу шахрайських транзакцій. Запропонована модель даних підтримує роботу основних функціональних компонентів системи та створює основу для подальшої реалізації програмного забезпечення.

3 РОЗРОБКА ТА РЕАЛІЗАЦІЯ ПРОГРАМНОЇ СИСТЕМИ

3.1 Реалізація моделі машинного навчання

Одним із ключових компонентів розробленої системи моніторингу шахрайських транзакцій є модуль машинного навчання, який забезпечує автоматичне визначення рівня ризику фінансових операцій. Основним призначенням даного модуля є аналіз характеристик транзакції та формування прогнозу щодо ймовірності її належності до шахрайських операцій.

Під час реалізації моделі було використано набір даних Credit Card Transactions Fraud Detection Dataset, який містить інформацію про фінансові транзакції та ознаку належності операції до шахрайських або легітимних дій. Для зменшення обсягу обчислень та прискорення процесу навчання із початкового набору було сформовано вибірку обсягом 200 000 записів. Використання достатньо великого набору даних дозволило забезпечити репрезентативність навчальної вибірки та отримати стабільні результати класифікації.

Перед початком навчання моделі було виконано попередню обробку даних. Даний етап є важливою складовою побудови систем машинного навчання, оскільки якість підготовлених даних безпосередньо впливає на точність майбутньої моделі. Загальну схему процесу навчання моделі наведено на рисунку 3.1.

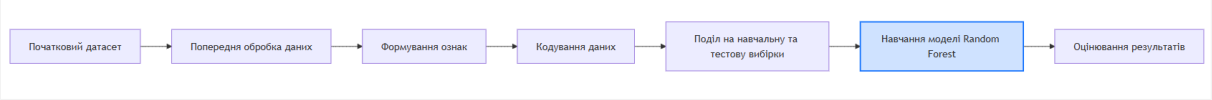


Рисунок 3.1 – Схема процесу навчання моделі машинного навчання

На першому етапі було виконано перетворення часових даних. Початковий набір містив поле trans_date_trans_time, яке зберігало дату та час проведення транзакції. Для підвищення інформативності даних на його основі було сформовано додаткові ознаки, що характеризують часові особливості здійснення операції.

У результаті виконання даного етапу для кожної транзакції було визначено годину здійснення операції, день тижня та місяць. Використання часових характеристик дозволяє моделі враховувати поведінкові особливості користувачів, оскільки шахрайські операції нерідко виконуються у нетиповий для клієнта час.

Додатково було сформовано ознаку віку користувача. Для цього використовувалися дата проведення транзакції та дата народження клієнта. Сформована ознака дозволяє враховувати демографічні характеристики клієнтів під час аналізу транзакцій.

Наступним етапом підготовки даних стало кодування категоріальних ознак. Оскільки більшість алгоритмів машинного навчання працює лише з числовими значеннями, текстові поля потребують попереднього перетворення. Для реалізації даного завдання використовувався клас LabelEncoder із бібліотеки Scikit-learn. Фрагмент коду кодування категоріальних ознак наведено у лістингу 3.1.

Лістинг 3.1 – Кодування категоріальних ознак

```
gender_encoder = LabelEncoder()
category_encoder = LabelEncoder()
df['gender'] = gender_encoder.fit_transform(
df['gender']
)
df['category'] = category_encoder.fit_transform(
df['category']
)
```

Після завершення етапу підготовки було сформовано набір ознак, який використовувався для навчання моделі. До нього увійшли характеристики транзакції, географічні координати клієнта та продавця, часові параметри, демографічні показники та категорія операції.

Для оцінювання якості моделі набір даних було розділено на навчальну та тестову вибірки у співвідношенні 80% до 20%. При цьому використовувався механізм стратифікації, що дозволив зберегти початкове співвідношення класів у кожній із вибірок.

Після підготовки даних було виконано навчання моделі Random Forest. Для побудови класифікатора використовувалася реалізація RandomForestClassifier із бібліотеки Scikit-learn. Основні параметри моделі наведено в таблиці 3.1.

Таблиця 3.1 – Параметри моделі Random Forest

Параметр	Значення
n_estimators	100
random_state	42
class_weight	balanced
n_jobs	-1

Параметр n_estimators визначає кількість дерев рішень, які входять до складу ансамблю. У межах даної роботи було використано 100 дерев, що дозволило забезпечити стабільність результатів класифікації. Використання параметра class_weight = balanced є особливо важливим для задач виявлення шахрайства, оскільки кількість шахрайських транзакцій суттєво менша за кількість легітимних операцій.

Після завершення навчання було виконано оцінювання якості моделі на тестовій вибірці. Для цього використовувалися такі показники як

Accuracy, Precision, Recall та F1-score. Отримані результати наведено в таблиці 3.2.

Таблиця 3.2 – Результати оцінювання моделі

Метрика	Значення
Accuracy	0,99765
Precision	0,96154
Recall	0,63025
F1-score	0,76142

Як видно з отриманих результатів, модель демонструє дуже високий показник Accuracy, який перевищує 99 %. Однак для задач виявлення фінансового шахрайства більш важливими є показники Precision та Recall, оскільки вони характеризують здатність системи правильно виявляти саме шахрайські операції.

Показник Precision на рівні 96 % свідчить про те, що більшість транзакцій, які модель визначає як шахрайські, дійсно належать до даного класу. Водночас значення Recall на рівні 63 % вказує на те, що частина шахрайських операцій все ще залишається невиявленою.

Для більш детального аналізу результатів класифікації було побудовано матрицю помилок, наведену на рисунку 3.2.

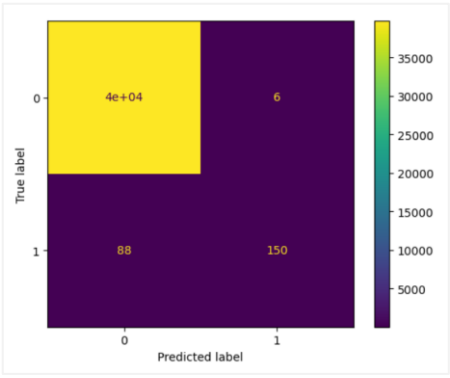


Рисунок 3.2 – Матриця помилок моделі Random Forest

Аналіз матриці помилок показує, що модель правильно класифікувала 39 756 легітимних транзакцій та 150 шахрайських операцій. Водночас було зафіксовано лише 6 хибнопозитивних спрацювань, коли звичайна операція була помилково визначена як шахрайська. Кількість пропущених шахрайських транзакцій становила 88 випадків.

Отримані результати підтверджують ефективність використання алгоритму Random Forest для задач виявлення фінансового шахрайства. Модель забезпечує високу точність класифікації та демонструє здатність виявляти більшість шахрайських операцій навіть за умов значного дисбалансу класів.

Після завершення навчання модель була збережена у файл формату PKL для подальшого використання в системі моніторингу транзакцій. Крім самої моделі також було збережено об'єкти кодування категоріальних ознак, що дозволяє забезпечити однакову обробку даних під час навчання та експлуатації системи.

Таким чином, у межах даного етапу було реалізовано модуль машинного навчання на основі алгоритму Random Forest, виконано підготовку даних, навчання моделі та оцінювання якості її роботи. Отримана модель стала основою підсистеми автоматичного виявлення шахрайських транзакцій та використовується для прогнозування рівня ризику фінансових операцій у розробленому програмному забезпеченні. Повний програмний код навчання моделі наведено у додатку А.

3.2 Реалізація бази даних

Одним із ключових компонентів розробленої системи моніторингу шахрайських транзакцій є база даних, яка забезпечує централізоване зберігання інформації та взаємодію між окремими модулями програмного забезпечення. Саме база даних використовується для накопичення результатів аналізу транзакцій, збереження повідомлень про ризикові

операції, ведення списку заблокованих рахунків та реалізації механізму авторизації користувачів.

Для реалізації сховища даних було обрано систему керування базами даних SQLite. Основною перевагою даного рішення є простота використання, відсутність необхідності встановлення окремого серверного програмного забезпечення та повна інтеграція з мовою програмування Python. Крім того, SQLite забезпечує достатню продуктивність для задачі моніторингу транзакцій у межах навчального проєкту.

Фізично база даних зберігається у файлі fraud_monitoring.db та використовується всіма компонентами системи для читання та запису інформації. Взаємодія між програмними модулями та базою даних здійснюється через окремий модуль database.py, який реалізує необхідні функції доступу до даних.

На початковому етапі реалізації було створено структуру бази даних. Для цього використовувався модуль sqlite3, який входить до стандартної бібліотеки Python. Під час запуску програми автоматично створюється файл бази даних та необхідні таблиці. Фрагмент програмного коду підключення до бази даних наведено у лістингу 3.2.

Лістинг 3.2 – Підключення до бази даних SQLite

```
import sqlite3

DB_PATH = "database/fraud_monitoring.db"

def get_connection():
    return sqlite3.connect(DB_PATH)
```

Центральною таблицею бази даних є таблиця transactions, яка використовується для зберігання інформації про всі оброблені транзакції. Саме ця таблиця накопичує результати роботи моделі машинного навчання

та є основним джерелом даних для формування статистики системи. Фрагмент створення таблиці transactions наведено у лістингу 3.3.

Лістинг 3.3 – Створення таблиці транзакцій

```
CREATE TABLE IF NOT EXISTS transactions (
    id INTEGER PRIMARY KEY AUTOINCREMENT,
    transaction_id TEXT UNIQUE,
    cc_num TEXT,
    amount REAL,
    category TEXT,
    fraud_probability REAL,
    risk_score REAL,
    risk_level TEXT,
    prediction INTEGER,
    timestamp TEXT
)
```

Після отримання результатів класифікації нова транзакція автоматично записується до таблиці transactions. Для реалізації даної функціональності було створено окрему функцію add_transaction(), яка приймає параметри транзакції та виконує їх збереження в базі даних.

Окрім інформації про фінансові операції система також зберігає дані користувачів. Для цього використовується таблиця users, яка містить ім'я користувача, пароль та роль доступу.

Використання ролей дозволяє реалізувати розмежування прав доступу між звичайними користувачами та адміністраторами системи. Адміністратор отримує доступ до функцій блокування рахунків та керування повідомленнями про ризикові операції.

Для реалізації механізму контролю підозрілих транзакцій використовується таблиця alerts. Вона містить інформацію про операції, які були визначені системою як потенційно небезпечні та потребують додаткової перевірки.

У процесі роботи системи користувач може прийняти рішення про блокування рахунку або ігнорування повідомлення. Для реалізації цих можливостей були створені таблиці blocked_accounts та skipped_alerts.

Таблиця blocked_accounts використовується для зберігання інформації про рахунки, щодо яких було прийнято рішення про блокування. Для кожного запису зберігається номер картки, дата блокування, користувач, який виконав операцію, причина блокування та поточний статус.

Таблиця skipped_alerts призначена для зберігання інформації про попередження, які були переглянуті та відхилені без виконання додаткових дій. Використання даної таблиці дозволяє уникнути повторного відображення вже оброблених повідомлень. ER-діаграму реалізованої бази даних наведено на рисунку 2.5.

Після створення структури бази даних було реалізовано набір функцій для роботи з інформацією. Основними операціями є додавання транзакцій, отримання статистичних даних, формування історії транзакцій, отримання списку активних попереджень та управління заблокованими рахунками.

Для формування статистики на інформаційній панелі реалізовано функцію get_dashboard_stats(), яка виконує SQL-запити до таблиці transactions та визначає загальну кількість транзакцій, а також кількість операцій із високим, середнім та низьким рівнем ризику.

Окремо реалізовано функцію get_dashboard_extra_stats(), яка використовується для відображення кількості активних повідомлень та заблокованих рахунків.

Суттєвою перевагою використання окремого програмного модуля для роботи з базою даних є ізоляція логіки доступу до даних від інших частин системи. Завдяки цьому будь-які зміни структури таблиць або SQL-запитів можуть бути виконані без необхідності модифікації вебзастосунку чи модуля машинного навчання.

Таким чином, у межах даного етапу було реалізовано реляційну базу даних на основі SQLite, яка забезпечує зберігання інформації про користувачів, транзакції, повідомлення про ризикові операції та заблоковані рахунки. Розроблена структура даних стала основою взаємодії між усіма компонентами програмної системи та забезпечила можливість подальшої реалізації функціональних можливостей вебзастосунку.

Повний програмний код створення бази даних наведено у додатку Б.

3.3 Реалізація вебзастосунку

Для забезпечення взаємодії користувача із системою моніторингу шахрайських транзакцій було реалізовано вебзастосунок на основі фреймворку Streamlit. Даний інструмент дозволяє швидко створювати інтерактивні вебінтерфейси мовою Python без використання додаткових технологій фронтенд-розробки. Використання Streamlit забезпечило можливість реалізації зручного користувацького інтерфейсу для перегляду результатів аналізу транзакцій, моніторингу роботи системи та управління повідомленнями про потенційні шахрайські операції.

Загальна структура вебзастосунку побудована за модульним принципом. Основний файл app.py відповідає за авторизацію користувачів, формування навігаційного меню та відображення окремих сторінок системи. Кожен функціональний модуль реалізовано у вигляді окремого файлу в каталозі views, що забезпечує зручність супроводу та подальшого розширення програмного забезпечення.

Такий підхід дозволяє легко масштабувати систему шляхом додавання нових функціональних модулів без внесення суттєвих змін до основної структури застосунку. Крім того, модульна архітектура спрощує процес тестування, супроводу та подальшої модернізації програмного забезпечення.

Для керування навігацією між сторінками використовується бічна панель Streamlit. Вибір конкретного пункту меню визначає модуль, який буде відображено користувачу. Фрагмент програмного коду формування навігаційного меню наведено у лістингу 3.6.

Лістинг 3.6 – Формування навігаційного меню

```
selected_page = st.sidebar.radio(
    "Menu",
    menu_pages
)
```

Однією з основних функцій вебзастосунку є авторизація користувачів. Перед початком роботи користувач повинен пройти процедуру входу до системи, використовуючи логін та пароль. Перевірка введених облікових даних здійснюється шляхом звернення до бази даних через функцію login_user(). У разі успішної автентифікації інформація про користувача зберігається у session_state та використовується під час роботи із системою. Інтерфейс сторінки авторизації наведено на рисунку 3.3.

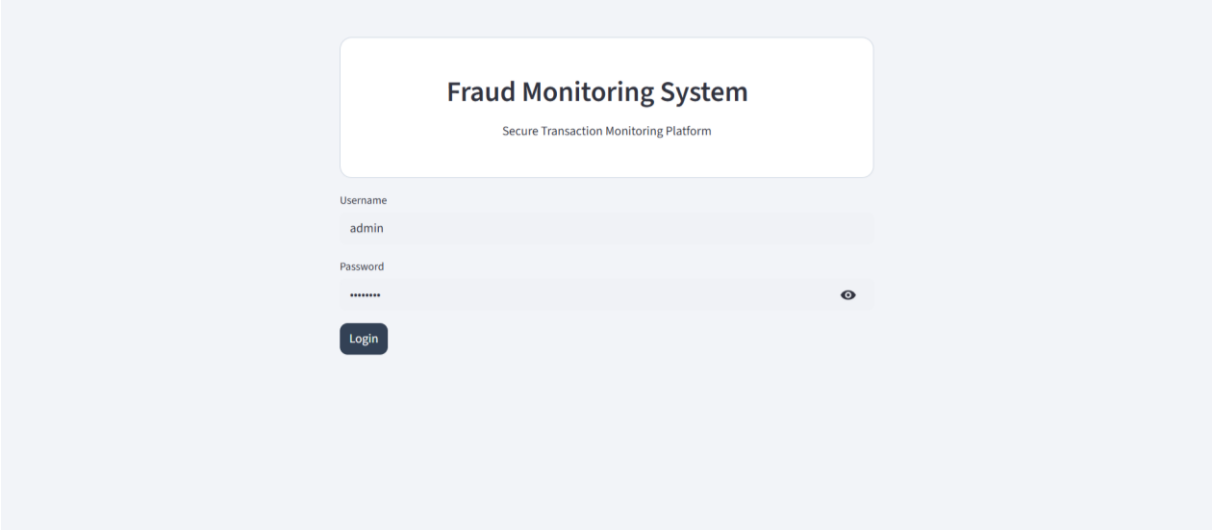


Рисунок 3.3 – Сторінка авторизації користувача

Після успішного входу користувач отримує доступ до головної інформаційної панелі Dashboard. Дана сторінка призначена для оперативного моніторингу поточного стану системи та відображення основних статистичних показників.

На Dashboard відображаються:

– загальна кількість оброблених транзакцій;

– кількість активних попереджень;

– кількість заблокованих рахунків;

– кількість операцій із високим, середнім та низьким рівнем ризику;

– кругова діаграма розподілу транзакцій за рівнем ризику;

– таблиця останніх оброблених транзакцій.

Для візуалізації статистичної інформації використовуються компоненти Metric та інтерактивні графіки бібліотеки Plotly. Інтерфейс дашборду наведено на рисунку 3.4.

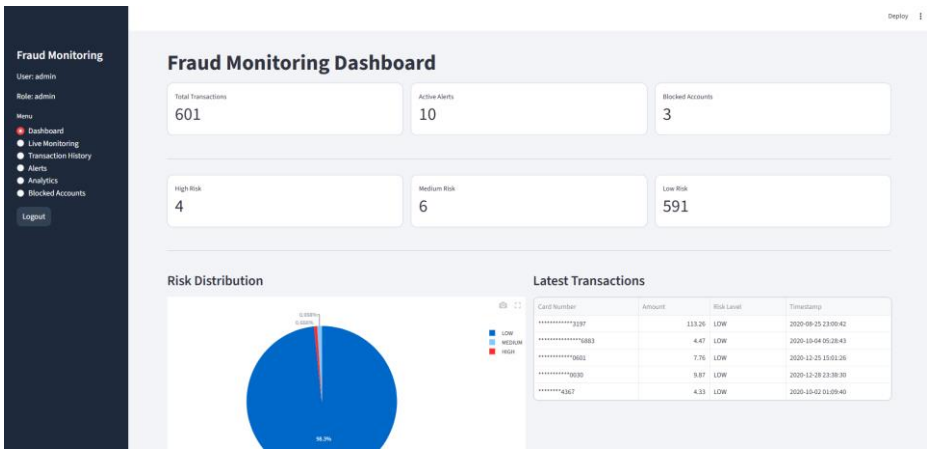


Рисунок 3.4 – Головна інформаційна панель Dashboard

Наступним компонентом системи є сторінка Live Monitoring, яка забезпечує керування фоновим сервісом моніторингу транзакцій. Дана сторінка дозволяє запускати та зупиняти процес автоматичного аналізу фінансових операцій без необхідності використання командного рядка.

Під час роботи сервіс у фоновому режимі здійснює обробку нових транзакцій, передає дані до моделі машинного навчання та формує результати оцінювання ризику. Такий підхід дозволяє забезпечити безперервний контроль фінансових операцій у режимі реального часу.

Користувач може контролювати поточний стан сервісу та отримувати інформацію про його роботу безпосередньо через вебінтерфейс. Для цього реалізовано функції запуску, зупинки та перевірки активності сервісу. Інтерфейс сторінки моніторингу наведено на рисунку 3.5.

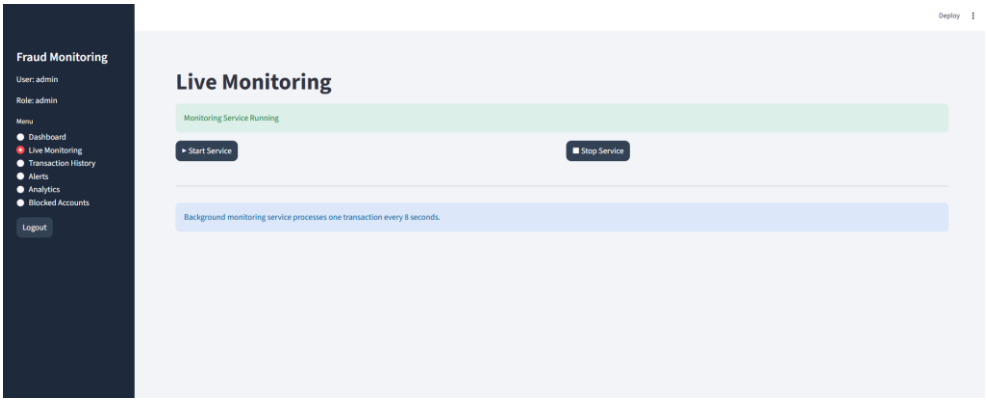


Рисунок 3.5 – Сторінка керування сервісом моніторингу

Для перегляду накопиченої інформації про оброблені транзакції реалізовано сторінку Transaction History. На даній сторінці відображається повна історія фінансових операцій, що були проаналізовані системою. Ця інформація періодично оновлюється автоматично і є актуальною. Для кожної транзакції відображається основна інформація, необхідна для проведення аналізу, зокрема дата та час виконання операції, сума платежу, категорія транзакції, результат класифікації моделі та присвоєний рівень ризику. Наявність цих даних дозволяє користувачу швидко оцінити характер фінансової операції. Також можна побачити загальну кількість транзакцій, здійснених на момент перегляду сторінки. Автоматичне оновлення дозволяє цій інформації залишатися актуальною впродовж перебування на цій сторінці

З метою підвищення рівня конфіденційності номери платіжних карток автоматично маскуються, а користувачу відображаються лише останні чотири цифри номера. Додатково реалізовано можливість фільтрації записів за рівнем ризику, що значно спрощує пошук підозрілих транзакцій. Інтерфейс сторінки історії транзакцій наведено на рисунку 3.6.

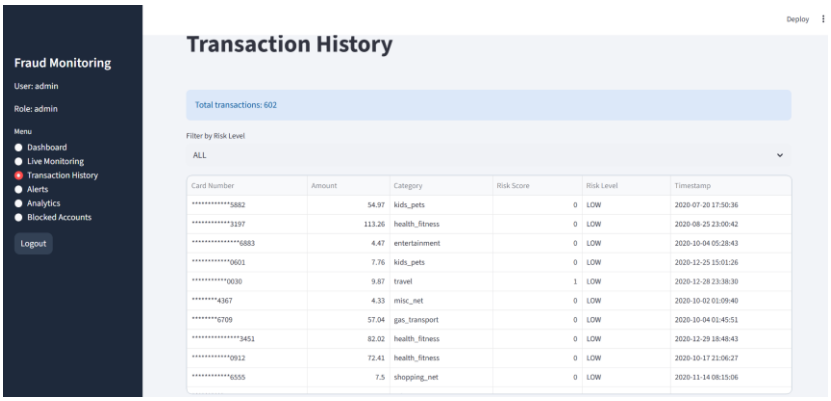


Рисунок 3.6 – Сторінка перегляду історії транзакцій

Однією з найважливіших сторінок системи є модуль Alerts, призначений для роботи з транзакціями, які були класифіковані як потенційно небезпечні. Система автоматично відображає всі операції із середнім та високим рівнем ризику. Для кожної транзакції користувач може переглянути основні характеристики операції, значення ризикового показника та час її виконання. Відображення транзакцій за рівнем ризику дозволяє користувачу швидко зосередитися на операціях, які потребують додаткової уваги. Для зручності аналізу інформація подається у структурованому вигляді, що спрощує перегляд основних параметрів транзакції та оцінку потенційної загрози.

Додатковою перевагою сторінки є можливість оперативного реагування на виявлені ризикові операції. Адміністратор може безпосередньо через інтерфейс системи виконувати необхідні дії, пов'язані з контролем підозрілих транзакцій, що сприяє підвищенню загального рівня безпеки фінансового моніторингу.

Оновлення інформації здійснюється автоматично кожні п'ять секунд, що забезпечує наближений до реального часу режим роботи системи. Інтерфейс сторінки Alerts наведено на рисунку 3.7.

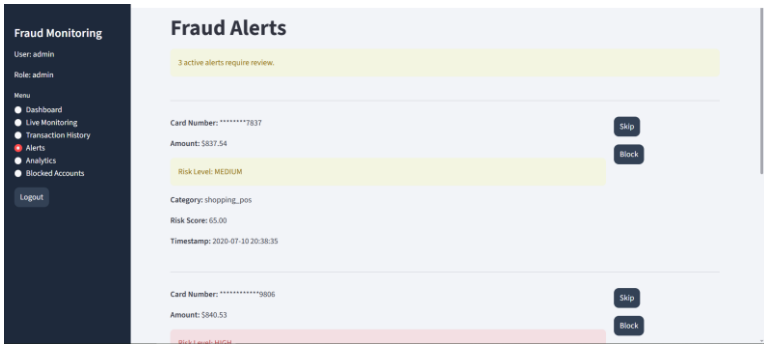


Рисунок 3.7 – Сторінка активних попереджень

Для аналізу накопиченої статистики реалізовано сторінку Analytics. На ній відображаються інтерактивні візуалізації, побудовані на основі даних із бази даних. Користувач може переглядати розподіл транзакцій за рівнем ризику, найбільш популярні категорії транзакцій, розподіл значень ризикового показника.

Використання графічних методів подання інформації дозволяє швидко оцінити загальний стан системи та виявити можливі аномалії у фінансових операціях. Інтерфейс (масштабований) сторінки Analytics наведено на рисунку 3.8.

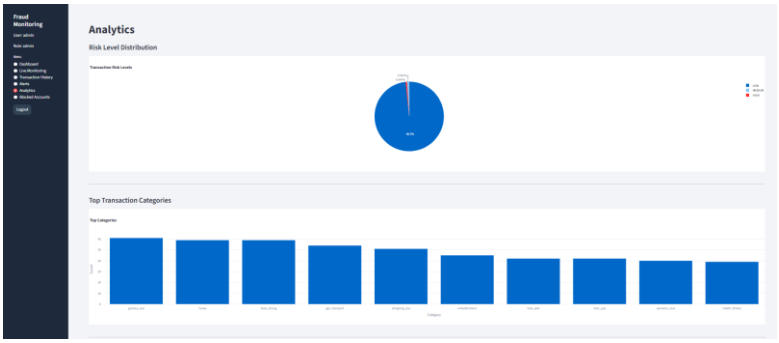


Рисунок 3.8 – Сторінка аналітики транзакцій

Для адміністраторів системи реалізовано окремий модуль Blocked Accounts. Доступ до даної сторінки мають лише користувачі з відповідною роллю. Сторінка відображає список усіх заблокованих рахунків із зазначенням часу блокування, причини та статусу.

Наявність централізованого списку заблокованих рахунків дозволяє адміністраторам оперативно контролювати застосовані обмеження та відстежувати історію прийнятих рішень. Це сприяє підвищенню рівня безпеки системи та забезпечує можливість швидкого реагування у випадках помилкового або тимчасового блокування рахунків.

За необхідності адміністратор може виконати розблокування рахунку безпосередньо через вебінтерфейс системи. Інтерфейс сторінки наведено на рисунку 3.9.

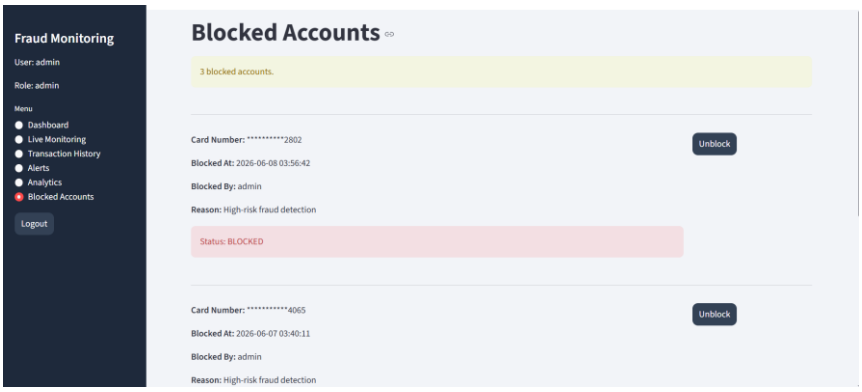


Рисунок 3.9 – Сторінка заблокованих рахунків

Особливістю реалізованого вебзастосунку є автоматичне оновлення інформації на більшості сторінок. Для цього використовується бібліотека streamlit-autorefresh, яка дозволяє оновлювати дані без необхідності ручного перезавантаження сторінки. Такий підхід забезпечує актуальність інформації та покращує зручність роботи користувачів.

Таким чином, у межах даного етапу було реалізовано повнофункціональний вебзастосунок для моніторингу шахрайських транзакцій. Розроблений інтерфейс забезпечує зручну взаємодію з усіма

компонентами системи, підтримує розмежування прав доступу та надає користувачу інструменти для оперативного аналізу фінансових операцій і реагування на потенційні загрози.

3.4 Реалізація фонового моніторингу

Однією з ключових особливостей розробленої системи є можливість автоматичного аналізу фінансових транзакцій у фоновому режимі. Для реалізації даної функціональності було створено окремий сервіс моніторингу, який незалежно від вебінтерфейсу виконує обробку транзакцій та передає результати аналізу до бази даних.

Використання окремого фонового процесу дозволяє наблизити роботу системи до реальних умов експлуатації, у яких фінансові операції надходять безперервним потоком та повинні аналізуватися в режимі реального часу. Завдяки такому підходу вебзастосунок відповідає лише за відображення інформації, тоді як усі операції обробки даних виконуються окремим сервісом.

Основним завданням сервісу моніторингу є автоматичне отримання транзакцій із тестового набору даних, виконання прогнозування за допомогою моделі машинного навчання та збереження результатів до бази даних. Реалізація даного функціоналу здійснюється у модулі monitoring_service.py.

Під час запуску сервісу виконується завантаження набору даних fraudTest.csv, який використовується як джерело фінансових транзакцій для подальшого аналізу. Використання окремого тестового набору дозволяє моделювати надходження нових операцій після завершення навчання моделі. Фрагмент програмного коду завантаження даних наведено у лістингу 3.7.

Лістинг 3.7 – Завантаження тестового набору даних

```
def load_data():

    return pd.read_csv(
        "data/fraudTest.csv"
    )
```

Після завантаження даних система переходить до циклічної обробки транзакцій. Для імітації потоку фінансових операцій випадковим чином вибирається одна транзакція з набору даних. Надалі вона передається до модуля прогнозування, який визначає ймовірність шахрайства та рівень ризику.

Функція process_transaction() виконує повний цикл аналізу однієї транзакції. На даному етапі формується прогноз моделі машинного навчання, після чого результати передаються до бази даних для подальшого використання вебзастосунком.

Особливістю реалізованого рішення є використання безперервного циклу роботи сервісу. Після завершення аналізу поточної транзакції система очікує певний проміжок часу та переходить до обробки наступної операції. Фрагмент основного циклу моніторингу наведено у лістингу 3.8.

Лістинг 3.8 – Основний цикл роботи сервісу

```
while True:

    process_transaction(df)

    time.sleep(8)
```

Використання затримки тривалістю вісім секунд дозволяє імітувати поступове надходження нових фінансових операцій та спостерігати зміну статистичних показників системи в реальному часі. Після обробки кожної

транзакції інформація автоматично зберігається до таблиці transactions бази даних.

Для зручності керування сервісом моніторингу було реалізовано окремий програмний модуль service_manager.py. Його основним призначенням є запуск, зупинка та перевірка стану фонового процесу.

У межах даного модуля реалізовано три основні функції:

– start_service() – запуск сервісу моніторингу;

– stop_service() – завершення роботи сервісу;

– is_service_running() – перевірка поточного стану процесу.

Для запуску сервісу використовується механізм створення окремого процесу операційної системи за допомогою модуля subprocess. Завдяки цьому сервіс моніторингу працює незалежно від вебінтерфейсу та не блокує виконання інших компонентів системи.

Фрагмент коду запуску сервісу наведено у лістингу 3.9.

Лістинг 3.9 – Запуск сервісу моніторингу

```
service_process = subprocess.Popen(
    [
        "py",
        "services/monitoring_service.py"
    ],
    cwd=os.getcwd()
)
```

Для зупинки сервісу використовується метод terminate(), який завершує виконання фонового процесу. Завдяки цьому користувач може в будь-який момент припинити обробку транзакцій через вебінтерфейс системи.

Взаємодія між вебзастосунком та сервісом моніторингу реалізована через сторінку Live Monitoring. Дана сторінка відображає поточний стан

сервісу та надає користувачу можливість запускати або зупиняти процес аналізу транзакцій.

Під час роботи сервісу кожна нова транзакція автоматично проходить через модель машинного навчання. Після визначення рівня ризику результати зберігаються до бази даних та стають доступними на сторінках Dashboard, Alerts, Transaction History та Analytics. Завдяки автоматичному оновленню вебінтерфейсу користувач може спостерігати зміни практично в режимі реального часу.

Таким чином, у межах даного етапу було реалізовано окремий сервіс фонового моніторингу транзакцій, який забезпечує автоматичне надходження нових даних, виконання прогнозування за допомогою моделі машинного навчання та передачу результатів до бази даних. Реалізований підхід дозволив створити повноцінну систему моніторингу фінансових операцій із підтримкою безперервного аналізу транзакцій та оперативного відображення результатів користувачу.

3.5 Тестування системи

Після завершення розробки програмної системи було проведено комплексне тестування її функціональних можливостей. Основною метою тестування була перевірка коректності роботи всіх програмних компонентів, оцінювання стабільності функціонування системи та підтвердження правильності взаємодії між модулем машинного навчання, базою даних, вебзастосунком та сервісом фонового моніторингу. Процес тестування охоплював перевірку як окремих функціональних модулів, так і роботи системи в цілому.

Тестування виконувалося на персональному комп'ютері під керуванням операційної системи Windows із використанням середовища Python та вебфреймворку Streamlit. Під час перевірки використовувалися транзакції з набору даних fraudTest.csv, які послідовно оброблялися

сервісом моніторингу та передавалися до інших компонентів системи. Особливу увагу було приділено перевірці роботи сервісу фонового моніторингу. Даний компонент відповідає за автоматичне надходження транзакцій до системи, їх аналіз за допомогою моделі машинного навчання та подальше збереження результатів до бази даних.

На першому етапі було проведено функціональне тестування основних можливостей програмного забезпечення. Результати перевірки наведено у таблиці 3.3.

Таблиця 3.3 – Результати функціонального тестування системи

Функціональна можливість	Очікуваний результат	Результат
Авторизація користувача	Успішний вхід до системи	Успішно
Запуск сервісу моніторингу	Початок аналізу транзакцій	Успішно
Зупинка сервісу моніторингу	Завершення роботи сервісу	Успішно
Аналіз транзакцій моделлю	Визначення рівня ризику	Успішно
Запис транзакцій у базу даних	Створення нового запису	Успішно
Відображення історії транзакцій	Формування таблиці транзакцій	Успішно
Формування попереджень	Відображення ризикових операцій	Успішно
Блокування рахунку	Додавання до списку блокування	Успішно
Розблокування рахунку	Видалення зі списку блокування	Успішно
Побудова аналітики	Формування графіків та статистики	Успішно

Аналіз результатів, наведених у таблиці 3.3, свідчить про коректну роботу всіх основних функціональних компонентів системи. Під час тестування не було виявлено критичних помилок, які могли б вплинути на працездатність програмного забезпечення або призвести до втрати даних.

Під час роботи сервісу в консольному режимі відображаються повідомлення про обробку нових транзакцій та визначений рівень ризику. Приклад роботи сервісу наведено на рисунку 3.10.

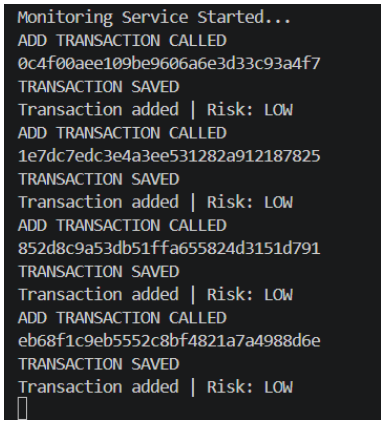


Рисунок 3.10 – Робота сервісу фонового моніторингу

Отримані результати підтверджують, що сервіс коректно здійснює аналіз нових транзакцій та забезпечує безперервне надходження даних до системи моніторингу.

Використання незалежної тестової вибірки дозволяє об'єктивно оцінити здатність моделі узагальнювати отримані під час навчання знання та коректно працювати з новими даними. Такий підхід забезпечує більш достовірне визначення якості класифікації та дозволяє оцінити можливість практичного застосування моделі в умовах реального функціонування системи.

Наступним етапом стала перевірка роботи моделі машинного навчання. Оцінювання виконувалося на тестовій вибірці, яка не використовувалася під час навчання моделі. Для аналізу якості класифікації

використовувалися метрики Accuracy, Precision, Recall та F1-score. Результати оцінювання наведено у таблиці 3.4.

Таблиця 3.4 – Основні метрики моделі машинного навчання

Метрика	Значення
Accuracy	99,76 %
Precision	96,15 %
Recall	63,03 %
F1-score	76,14 %

Як видно з таблиці 3.4, модель демонструє високий рівень точності класифікації. Особливо важливим для задач виявлення шахрайства є показник Precision, значення якого перевищує 96%. Це означає, що більшість транзакцій, позначених системою як шахрайські, дійсно належать до даного класу.

Показник Recall є нижчим та становить близько 63%, що пояснюється значною незбалансованістю набору даних. Кількість шахрайських транзакцій у вибірці є суттєво меншою за кількість легітимних операцій, що характерно для реальних фінансових систем. Незважаючи на це, отримані результати є достатніми для використання моделі як основи системи моніторингу фінансових операцій.

Додатковий аналіз якості класифікації було виконано за допомогою матриці помилок, наведеної у підрозділі 3.1. Аналіз матриці показав, що модель успішно виявляє більшість шахрайських транзакцій та характеризується низькою кількістю хибнопозитивних спрацювань.

Для перевірки роботи механізму оцінювання ризику було проаналізовано результати класифікації окремих транзакцій. Приклад отриманих результатів наведено у таблиці 3.5.

Таблиця 3.5 – Приклад результатів класифікації транзакцій

Рівень ризику	Характеристика
LOW	Ймовірність шахрайства менше 40 %
MEDIUM	Ймовірність шахрайства від 40 % до 70 %
HIGH	Ймовірність шахрайства понад 70 %

Використання трирівневої системи оцінювання ризику дозволяє більш гнучко реагувати на потенційно небезпечні фінансові операції. Транзакції з високим рівнем ризику автоматично потрапляють до списку попереджень та можуть бути додатково перевірені адміністратором системи.

Під час тестування також було перевірено механізм роботи з попередженнями та заблокованими рахунками. Було підтверджено, що після блокування рахунку інформація коректно відображається на сторінці Blocked Accounts, а після розблокування відповідний запис успішно видаляється зі списку заблокованих карток.

Окремо було протестовано роботу аналітичного модуля. Система успішно формує графіки розподілу транзакцій за рівнем ризику, статистику категорій платежів та гістограми значень ризикового показника. Отримані результати дозволяють оперативно оцінювати поточний стан фінансових операцій та виявляти потенційно небезпечні тенденції.

Таким чином, проведене тестування підтвердило працездатність усіх компонентів розробленої програмної системи. Було встановлено коректність роботи моделі машинного навчання, сервісу фонового моніторингу, бази даних та вебінтерфейсу. Отримані результати свідчать про можливість використання розробленого програмного забезпечення для автоматизованого моніторингу фінансових транзакцій та виявлення потенційно шахрайських операцій.

ВИСНОВКИ

У кваліфікаційній роботі розглянуто актуальну задачу виявлення шахрайських фінансових транзакцій із використанням методів машинного навчання. Зростання кількості електронних платежів та розвиток цифрових фінансових сервісів призводять до збільшення ризиків фінансового шахрайства, що обумовлює необхідність створення сучасних автоматизованих систем моніторингу та аналізу транзакцій.

У процесі виконання роботи було проведено аналіз предметної галузі, досліджено основні види фінансового шахрайства, сучасні методи його виявлення та особливості застосування алгоритмів машинного навчання для розв’язання задач класифікації фінансових операцій. Також виконано огляд існуючих програмних рішень і наукових підходів до побудови антифрод-систем, що дозволило визначити основні вимоги до розроблюваної системи.

У результаті роботи спроєктовано систему моніторингу шахрайських транзакцій, виконано аналіз і підготовку набору даних, обрано алгоритм Random Forest для побудови моделі машинного навчання, розроблено архітектуру програмної системи та структуру бази даних. Реалізовано вебзастосунок для моніторингу фінансових операцій, перегляду історії транзакцій, формування попереджень та аналізу статистичних показників роботи системи.

Практичним результатом роботи стала розроблена програмна система моніторингу фінансових транзакцій, яка забезпечує автоматичне оцінювання ризику фінансових операцій та підтримує прийняття рішень щодо виявлення потенційно шахрайської активності. Для реалізації системи використано мову програмування Python, бібліотеки Pandas, Scikit-learn, Streamlit, Plotly та систему керування базами даних SQLite.

Під час тестування отримано такі показники якості моделі: Accuracy – 99,76 %, Precision – 96,15 %, Recall – 63,03 %, F1-score – 76,14 %. Отримані результати свідчать про високу ефективність використання

алгоритму Random Forest для задачі виявлення шахрайських транзакцій та підтверджують доцільність його застосування в системах фінансового моніторингу.

Розроблена система відповідає сучасним тенденціям розвитку антифрод-рішень та за своїм функціональним призначенням узгоджується з підходами, які використовуються у вітчизняних і міжнародних системах моніторингу фінансових операцій. Водночас запропоноване рішення має навчальний та дослідницький характер і може бути використане як основа для подальшого розвитку більш складних систем аналізу транзакцій.

Результати роботи можуть бути використані в навчальному процесі під час вивчення дисциплін, пов’язаних із машинним навчанням, аналізом даних, проєктуванням інформаційних систем та розробкою програмного забезпечення. Окремі рішення, реалізовані в межах роботи, можуть бути використані як основа для подальших наукових досліджень у сфері виявлення фінансового шахрайства.

Подальший розвиток роботи доцільно спрямувати на використання більш складних моделей машинного навчання та глибокого навчання, розширення набору ознак транзакцій, інтеграцію системи з реальними потоками фінансових даних та підвищення точності виявлення шахрайських операцій у режимі реального часу.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Mohan A. Banking fraud detection with machine learning and real-time analytics on AWS. *AWS*. URL: https://aws.amazon.com/ru/blogs/industries/banking-fraud-detection-with-machine-learning-and-real-time-analytics-on-aws (date of access: 09.06.2026).

2. Fraud detection using machine learning in banking. *Tookitaki*. URL: https://www.tookitaki.com/compliance-hub/fraud-detection-using-machine-learning-in-banking (date of access: 09.06.2026).

3. Learned lessons in credit card fraud detection from a practitioner perspective / A. Dal Pozzolo et al. *Expert systems with applications*. 2014. Vol. 41, no. 10. P. 4915–4928. URL: https://doi.org/10.1016/j.eswa.2014.02.026 (date of access: 09.06.2026).

4. Credit card fraud detection: a realistic modeling and a novel learning strategy. *IEEE transactions on neural networks and learning systems*. 2018. Vol. 29, no. 8. P. 3784–3797. URL: https://doi.org/10.1109/tnnls.2017.2736643 (date of access: 10.06.2026).

5. Real-time credit card fraud detection using machine learning / A. Thennakoon et al. *2019 9th international conference on cloud computing, data science & engineering (confluence)*, Noida, India, 10–11 January 2019. 2019. URL: https://doi.org/10.1109/confluence.2019.8776942 (date of access: 10.06.2026).

6. Financial fraud detection using machine learning : credit card fraud / O. Israel et al. *International journal of recent engineering science*. 2023. Vol. 10, no. 3. P. 23–32. URL: https://doi.org/10.14445/23497157/ijres-v10i3p104 (date of access: 10.06.2026).

7. Flinders M., Smalley I., Schneider J. AI fraud detection in banking. *IBM*. URL: https://www.ibm.com/think/topics/ai-fraud-detection-in-banking (date of access: 10.06.2026).

8. Dornadula V. N., Geetha S. Credit card fraud detection using machine learning algorithms. *Procedia computer science*. 2019. Vol. 165. P. 631–641. URL: https://doi.org/10.1016/j.procs.2020.01.057 (date of access: 10.06.2026).

9. George M. Z. H., Alam M. K., Hasan M. T. Machine learning for fraud detection in digital banking: a systematic literature review. *ASRC procedia: global perspectives in science and scholarship*. 2023. Vol. 03, no. 01. P. 37–61. URL: https://doi.org/10.63125/913ksy63 (date of access: 11.06.2026).

10. MacDougall J. Machine learning for fraud detection. *Ravelin*. URL: https://www.ravelin.com/insights/machine-learning-for-fraud-detection (date of access: 11.06.2026).

11. Bhanusri R., Ramana Reddy N. Fraud detection in banking transactions using machine learning. *International journal of advance research and innovation*. 2025. Vol. 13, no. 3. P. 27–36. URL: https://doi.org/10.69996/ijari.2025016 (date of access: 11.06.2026).

12. Tandel M. S., Bhati M. B. Credit card fraud detection using machine learning. *Interantional journal of scientific research in engineering and management*. 2024. Vol. 08, no. 008. P. 1–9. URL: https://doi.org/10.55041/ijsrem37368 (date of access: 12.06.2026).

13. Credit card fraud detection: a realistic modeling and a novel learning strategy. *IEEE transactions on neural networks and learning systems*. 2018. Vol. 29, no. 8. P. 3784–3797. URL: https://doi.org/10.1109/tnnls.2017.2736643 (date of access: 12.06.2026).

14. Le Borgne Y.-A., Bontempi G. Reproducible machine learning for credit card fraud detection - practical handbook. 2021.

15. Anomaly detection in financial transactions using convolutional neural networks / Md. Tanvir Rahman Mazumder et al. *Journal of economics, finance and accounting studies*. 2025. Vol. 7, no. 2. P. 195–207. URL: https://doi.org/10.32996/jefas.2025.7.2.17 (date of access: 12.06.2026).

16. RandomForestClassifier. *Scikit learn*. URL: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html (date of access: 11.06.2026).

17. Subramaniam A. Handling imbalanced datasets in machine learning: SMOTE, oversampling & undersampling explained. *Medium*. URL: https://pub.towardsai.net/handling-imbalanced-datasets-in-machine-learning-smote-oversampling-undersampling-explained-6a86116c1f41 (date of access: 12.06.2026).

18. Shenoy K. Credit card transactions fraud detection dataset. *Kaggle*. URL: https://www.kaggle.com/datasets/kartik2112/fraud-detection (date of access: 12.06.2026).

19. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, Incorporated, 2022.

20. Nasteski V. An overview of the supervised machine learning methods. *Horizons.b*. 2017. Vol. 4. P. 51–62. URL: https://doi.org/10.20544/horizons.b.04.1.17.p05 (date of access: 13.06.2026).

21. Rokach L., Maimon O. Decision trees. *Data mining and knowledge discovery handbook*. New York. P. 165–192. URL: https://doi.org/10.1007/0-387-25465-x_9 (date of access: 13.06.2026).

22. Breiman L. *Machine learning*. 2001. Vol. 45, no. 1. P. 5–32. URL: https://doi.org/10.1023/a:1010933404324 (date of access: 13.06.2026).