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

Навчально-науковий інститут прикладного системного аналізу

Кафедра штучного інтелекту

До захисту допущено:

В. о. завідувачки кафедри

_____ Ірина ДЖИГИРЕЙ

«___»_____20__ р.

Дипломна робота

на здобуття ступеня бакалавра

за освітньо-професійною програмою «Системи і методи штучного інтелекту»

спеціальності 122 «Комп'ютерні науки»

на тему: «Моделі машинного навчання для виявлення фродових транзакцій з метою мінімізації ризиків чарджбеків»

Виконала:
студентка IV курсу, групи КІ-21
Попадюк Каріна Вадимівна _____

Керівник:
доцент кафедри штучного інтелекту, к.ф-м.н., доцент,
Пишнограєв Іван Олександрович _____

Консультант з нормоконтролю:
фахівець першої категорії кафедри штучного інтелекту,
к.т.н., доцент,
Комариста Богдана Миколаївна _____

Рецензент:
доцент кафедри економічної кібернетики ФММ, к.ф.-м.н., доцент,
Лазаренко Ірина Сергіївна _____

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

Київ – 2026 року

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

Рівень вищої освіти – перший (бакалаврський)

Спеціальність – 122 «Комп'ютерні науки»

Освітньо-професійна програма «Системи і методи штучного інтелекту»

ЗАТВЕРДЖУЮ
В. о. завідувачки кафедри
_____ Ірина ДЖИГИРЕЙ
«15» січня 2026 р.

ЗАВДАННЯ
на дипломну роботу студенту
Попадюк Каріна Вадимівна

1. Тема роботи «Розробка моделей машинного навчання для виявлення фродових транзакцій з метою мінімізації ризиків чарджбеків», керівник роботи Пишнограєв Іван Олександрович, доцент кафедри штучного інтелекту, к.ф-м.н., доцент, затверджені наказом по НН ІПСА від «27» травня 2026 р. №1944-с.

2. Термін подання студентом роботи «05» червня 2025 року.

3. Вихідні дані до роботи: набір банківських транзакцій

4. Зміст роботи: Аналіз предметної області дослідження та даних, вибір доцільних методів та моделей для роботи, розробка алгоритму обраними методами, оцінка отриманих результатів.

5. Дата видачі завдання «02» лютого 2026 року.

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1	Огляд літератури за темою	20.04.2026	Виконано
2	Підготовка першого розділу	27.04.2026	Виконано
3	Підготовка другого розділу	04.05.2026	Виконано
4	Розробка програмного продукту	11.05.2026	Виконано
5	Підготовка третього розділу	18.05.2026	Виконано
6	Оформлення розділів відповідно до нормоконтролю	25.05.2026	Виконано
7	Оформлення дипломної роботи	01.06.2026	Виконано
8	Підготовка презентації доповіді	05.06.2026	Виконано

Студент Каріна ПОПАДЮК

Керівник Іван ПИШНОГРАЄВ

РЕФЕРАТ

Дипломна робота: 93 с., 27 рис., 23 табл., 28 посилань, без додатків.

ЧАРДЖБЕК, ФРОД, ФРОДОВІ ТРАНЗАКЦІЇ, МАШИННЕ НАВЧАННЯ, БІНАРНА КЛАСИФІКАЦІЯ, НЕЗБАЛАНСОВАНІ ДАНІ, RANDOM FOREST, XGBOOST, LIGHTGBM, ВИЯВЛЕННЯ ШАХРАЙСТВА.

Об’єктом дослідження є процес виникнення фродових транзакцій у цифрових платіжних системах та їхній вплив на рівень чарджбеків.

Предметом дослідження є методи машинного навчання для виявлення фродових транзакцій і мінімізації ризиків чарджбеків для бізнесу.

Метою роботи є розроблення, аналіз та порівняння моделей машинного навчання для прогнозування фродових транзакцій з урахуванням сильного дисбалансу класів.

У дипломній роботі розглянуто чарджбек як механізм захисту споживачів і водночас як джерело фінансових, операційних та репутаційних ризиків для мерчантів. Проаналізовано основні типи чарджбеків та встановлено, що справжній фрод потребує превентивного підходу.

Роботу формалізовано як задачу бінарної класифікації з учителем для сильно незбалансованих даних. Обґрунтовано, що стандартна точність є недостатньо інформативною для цієї задачі, тому основну увагу приділено Precision, Recall, F2 та ROC-AUC. Проаналізовано логістичну регресію, дерево рішень, Random Forest, XGBoost, LightGBM, KNN, MLP та TabNet. Встановлено, що найкращу якість за F2-мірою продемонструвала модель Random Forest. Отримані результати підтверджують, що ефективна система виявлення фроду має поєднувати якісну модель класифікації, коректну роботу з дисбалансом класів і ретельне налаштування порогу прийняття рішень. Практична цінність роботи полягає у можливості використання побудованого підходу для зменшення кількості фродових транзакцій, скорочення чарджбеків і зниження фінансових ризиків бізнесу.

ABSTRACT

Thesis: 93 pages, 27 figures, 23 tables, 28 references, no appendices.

CHARGEBACK, FRAUD, FRAUDULENT TRANSACTIONS, MACHINE LEARNING, BINARY CLASSIFICATION, IMBALANCED DATA, RANDOM FOREST, XGBOOST, LIGHTGBM, FRAUD DETECTION.

The object of the study is the process of fraudulent transactions occurring in digital payment systems and their impact on the chargeback rate.

The subject of the study is machine learning methods for detecting fraudulent transactions and minimizing chargeback risks for businesses.

The aim of the thesis is to develop, analyze, and compare machine learning models for predicting fraudulent transactions while taking into account a strong class imbalance.

The thesis examines chargeback as a consumer protection mechanism and, at the same time, as a source of financial, operational, and reputational risks for merchants. The main types of chargebacks are analyzed, and it is established that true fraud requires a preventive approach.

The research problem is formalized as a supervised binary classification task for highly imbalanced data. It is substantiated that standard accuracy is not sufficiently informative for this task; therefore, the main focus is placed on Precision, Recall, F2-score, and ROC-AUC. Logistic Regression, Decision Tree, Random Forest, XGBoost, LightGBM, KNN, MLP та TabNet are analyzed. It is established that the Random Forest model demonstrated the best performance according to the F2-score. The obtained results confirm that an effective fraud detection system should combine a high-quality classification model, proper handling of class imbalance, and careful tuning of the decision threshold. The practical value of the thesis lies in the possibility of using the developed approach to reduce the number of fraudulent transactions, decrease chargebacks, and lower financial risks for businesses.

ЗМІСТ

ВСТУП

Динамічне[1] поширення цифрових платежів стало одним із визначальних чинників трансформації сучасної фінансової екосистеми. Активний розвиток електронної комерції, сервісів із регулярною оплатою, мобільних застосунків та різноманітних онлайн-платформ зумовив суттєве збільшення обсягів карткових операцій у глобальному вимірі. Для споживачів такі платежі забезпечують оперативність, зручність і можливість здійснювати фінансові операції незалежно від місця перебування. Для суб'єктів господарювання цифрові платіжні інструменти, своєю чергою, створюють передумови для масштабування діяльності, виходу на зовнішні ринки та автоматизації процесів продажу. Водночас зростання кількості електронних транзакцій супроводжується підвищенням частоти спірних операцій, випадків шахрайства та чарджбеків, що формує додаткові ризики як для мерчантів, так і для платіжної інфраструктури загалом.

Чарджбек доцільно розглядати як механізм оскарження карткової транзакції, за допомогою якого картхолдер може звернутися до банку-емітента у разі несанкціонованої, шахрайської або спірної операції. Історично цей інструмент був спрямований насамперед на захист прав споживачів та зміцнення довіри до карткових розрахунків. Його значення було особливо важливим у ситуаціях, коли покупець не мав можливості ефективно врегулювати проблему безпосередньо з продавцем. Проте в сучасному платіжному середовищі чарджбек виконує не лише захисну функцію для

[1]Тут і нижче використано такий інструмент штучного інтелекту як чат-бот з генеративним штучним інтелектом ChatGPT, виключно для корегування та редагування тексту, створеного автором цієї дипломної роботи, на основі автоматизованої перевірки граматики, структури та стилю, що відповідає Політиці використання штучного інтелекту для академічної діяльності в КПІ ім. Ігоря Сікорського (протокол №11 Вченої ради КПІ ім. Ігоря Сікорського від 11 грудня 2023 р.).

клієнта, а також став вагомим джерелом фінансових, операційних і репутаційних загроз для бізнесу.

Однією з основних причин збільшення кількості чарджбеків є фрод. У широкому розумінні фрод охоплює шахрайське або неправомірне використання платіжних інструментів, а також механізмів повернення коштів із метою отримання фінансової вигоди. Його прояви можуть бути різними. Зокрема, йдеться про справжній фрод, коли транзакція здійснюється третьою особою без згоди власника картки. Іншим поширеним різновидом є так званий дружній фрод, за якого сам власник картки оскаржує операцію, що фактично була легітимною. Обидві форми безпосередньо впливають на рівень чарджбеків, однак відрізняються за природою виникнення, ступенем контрольованості та підходами до запобігання.

Збільшення кількості чарджбеків має для торговців багатовимірні негативні наслідки. Насамперед бізнес втрачає дохід від відповідної транзакції та змушений сплачувати додаткові комісії. Крім того, компанії витрачають значні ресурси на підготовку доказової бази й участь у процедурі репрезентменту. Додатковим ризиком є можливість потрапляння до моніторингових програм платіжних систем, що може погіршити умови подальшої обробки. У довгостроковій перспективі стабільно високий рівень чарджбеків здатний призвести до підвищення тарифів від банків-еквайєрів, застосування фінансових санкцій, обмеження платіжного процесингу або навіть втрати доступу до приймання карткових платежів та блокування юридичної особи на відкриття рахунків.

За таких умов особливої актуальності набуває дослідження ролі фроду в загальній структурі чарджбеків, а також оцінювання можливостей його прогнозування. Частину випадків дружнього фроду можна мінімізувати завдяки покращенню клієнтського досвіду, підвищенню прозорості комунікації з покупцем, коректному відображенню платіжних дескрипторів і чітко сформульованим політикам повернення коштів. Натомість справжній фрод потребує переважно превентивного підходу, оскільки найбільш

ефективним є його виявлення ще на етапі авторизації транзакції. Саме тому впровадження систем попередження фроду, а також використання моделей прогнозування ризикових операцій, слід розглядати як один із ключових напрямів зниження рівня шахрайства, скорочення кількості чарджбеків і забезпечення фінансової стійкості бізнесу.

РОЗДІЛ 1 ЧАРДЖБЕКИ ЯК МЕХАНІЗМ ЗАХИСТУ СПОЖИВАЧІВ І ДЖЕРЕЛО РИЗИКІВ ДЛЯ БІЗНЕСУ

1.1 Поняття та історичне виникнення чарджбеків

Коли карткові платіжні системи були вперше впроваджені, це стало значним етапом трансформації механізмів здійснення фінансових операцій. Використання кредитних і дебетових карток дозволило споживачам здійснювати платежі без використання готівки, що суттєво підвищило швидкість та зручність транзакцій. Водночас на ранніх етапах впровадження карткових платежів спостерігався високий рівень недовіри з боку користувачів. Основні побоювання були пов’язані з питаннями безпеки, оскільки платіжні дані могли бути втрачені, викрадені або використані для несанкціонованих операцій. У результаті це суттєво стримувало масове поширення карткових платежів, особливо у випадках дистанційних транзакцій, де відсутня можливість фізичної ідентифікації сторін [1].

З метою підвищення рівня довіри та забезпечення захисту споживачів фінансовими установами було впроваджено механізм чарджбеків. Цей механізм є інструментом захисту прав власників платіжних карток і передбачає можливість оскарження транзакцій у випадках їх фродового, несанкціонованого або спірного характеру. Процедура чарджбеку реалізується через банк-емітент, до якого звертається власник картки, без безпосередньої взаємодії з торговцем. У разі підтвердження обґрунтованості звернення сума транзакції повертається клієнту, а відповідні кошти списуються з рахунку продавця. Запровадження цього механізму суттєво підвищило рівень довіри до карткових платіжних систем і стало одним із ключових чинників розвитку глобальної електронної комерції.

У сучасній фінансовій екосистемі карткові платежі залишаються одним із основних інструментів проведення транзакцій завдяки їхній швидкості, глобальній доступності та інтеграції з цифровими сервісами. Однак разом із

цим зростає кількість ризиків і спірних ситуацій, серед яких особливе місце займає процес чарджбеків, що використовується як механізм вирішення конфліктів між споживачем і торговцем.

Історично механізм чарджбеків був розроблений у період активного становлення карткових платіжних систем у 1970-х роках, коли необхідність регуляторного захисту була зумовлена високим рівнем фроду та недостатньою захищеністю споживачів. З часом ця система була формалізована в межах міжнародного фінансового регулювання, зокрема через ухвалення Fair Credit Billing Act 1974 року, який закріпив права споживачів на оскарження транзакцій та визначив обов'язки фінансових установ у процесах вирішення платіжних спорів.

У практиці платіжних систем чарджбеки зазвичай поділяються на три основні категорії.

1. Перша категорія — це помилки торговця, які виникають унаслідок неправильного оброблення транзакцій, зокрема подвійного списання, некоректного визначення суми або порушення правил платіжних систем, таких як Visa чи Mastercard.
2. Друга категорія — фродові чарджбеки, що виникають у випадках несанкціонованого використання платіжної картки, зокрема через викрадення даних, компрометацію акаунтів або використання підроблених платіжних інструментів.
3. Третя категорія — дружній фрод, який виникає тоді, коли клієнт оскаржує легітимну транзакцію. Це може бути як навмисна поведінка з метою отримання товару або послуги без оплати, так і ненавмисні ситуації, пов'язані із забутими покупками, непорозуміннями або незадоволенням якістю товару чи послуги. Дружній фрод особливо поширений у сфері e-commerce та підпискових сервісів.

1.2 Учасники та механізм роботи чарджбеку

Чарджбек ініціюється тоді, коли власник картки звертається до банку-емітента із заявою про оскарження транзакції, здійсненої за допомогою кредитної або дебетової картки. На відміну від стандартного повернення коштів, яке відбувається безпосередньо між покупцем і продавцем, чарджбек реалізується через платіжну інфраструктуру банку без участі торговця на початковому етапі. Банк здійснює перевірку обставин транзакції та, у разі підтвердження обґрунтованості звернення, ініціює зворотне списання коштів. У результаті продавець втрачає не лише дохід від операції, а й може нести додаткові фінансові та адміністративні витрати, що робить цей механізм суттєвим ризиковим фактором для бізнесу [4].

У чарджбеку зазвичай беруть участь 5 сторін.

1. Власник картки — це особа, чия платіжна картка була використана для транзакції. Вона може оскаржити транзакцію, якщо вважає її шахрайською, неавторизованою, продубльованою, якщо товар або послугу не було доставлено, вони не відповідали опису або транзакція була некоректною з іншої причини. Іноді власник картки і людина, яка здійснила покупку, — це одна й та сама особа. Однак у випадках справжнього шахрайства покупку могла здійснити інша особа, використавши викрадені дані картки.
2. Мерчант — це компанія або бізнес, що продав товар чи послугу. Після подання чарджбеку мерчант має вирішити, чи прийняти оскарження, чи боротися з ним.
3. Банк-емітент — це банк, який видав платіжну картку власнику картки. Він отримує скаргу від власника картки та вирішує, чи відповідає спір вимогам процедури чарджбеку.

4. Банк-еквайр — це банк мерчанта. Він обслуговує мерчант-рахунок і отримує повідомлення про чарджбек від карткової мережі або емітента.
5. Карткова мережа, наприклад Visa, Mastercard, American Express або Discover, встановлює правила для платіжних транзакцій і спорів. У деяких випадках мережа може долучитися, якщо спір переходить до арбітражу.
6. Платіжні процесори, шлюзи та сервіси запобігання шахрайству також можуть відігравати певну роль — залежно від того, як мерчант обробляє транзакції.

Чарджбек зазвичай починається тоді, коли власник картки помічає транзакцію, яку вважає некоректною або неавторизованою. Замість того щоб звернутися безпосередньо до мерчанта, він контактує зі своїм банком-емітентом і просить оскаржити списання.

Емітент розглядає заяву власника картки та перевіряє, чи відповідає вона затвердженій категорії оскарження. Якщо заява не відповідає вимогам, банк може відхилити спір ще до того, як він стане чарджбеком. Проте більшість банків стають на бік клієнта на початковому етапі й дозволяють спору рухатися далі, якщо причина виглядає обґрунтованою.

Після відкриття чарджбеку емітент зазвичай надає власнику картки тимчасове повернення коштів, яке часто називають тимчасовим кредитом. Потім спір передається через карткову мережу до банку-еквайра. Еквайр списує оскаржувану суму з рахунку мерчанта та може також застосувати комісію за чарджбек.

Після отримання повідомлення мерчант має відповісти до встановленого дедлайну. Є два основні варіанти.

1. Прийняти чарджбек. Мерчант погоджується зі спором або вирішує не оскаржувати його. Власник картки залишає собі повернену суму, а справу закривають.

2. Оскаржити чарджбек через репрезентмент. Мерчант подає докази, які підтверджують, що транзакція була дійсною, а заява власника картки — некоректною.

Якщо мерчант нічого не зробить до дедлайну, чарджбек зазвичай приймається автоматично. У деяких випадках з мерчанта також можуть стягнути додаткову комісію за відсутність відповіді. Блок-схему чарджбек-процесу показано на рисунку 1.1.

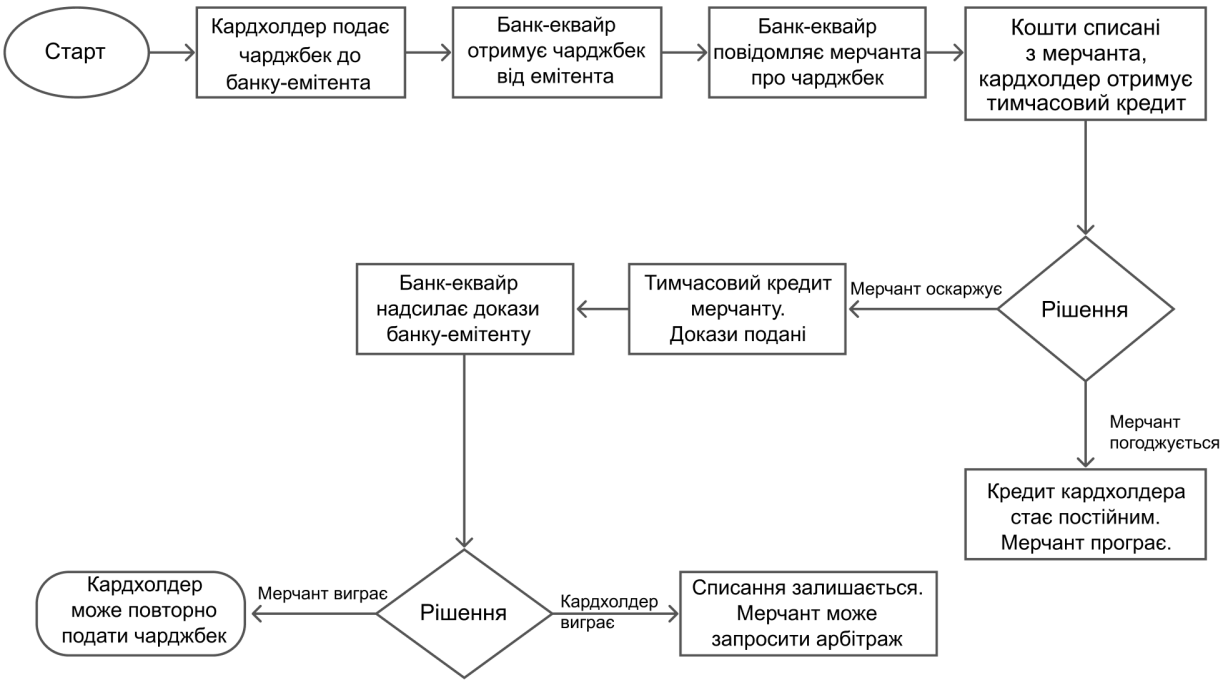

Рисунок 1.1 – Детальна блок-схема чарджбек-процесу [4]

1.3 Захисна функція чарджбеків для покупців та негативний вплив на продавців

Чарджбеки відіграють важливу роль у захисті споживачів. Цей механізм був створений, щоб надати власникам карток безпечний і відносно простий

спосіб оскаржувати підозрілі, некоректні або несправедливі списання, або у ситуаціях, коли щось іде не так із покупкою.

1. У виписці за карткою з’явилася неавторизована транзакція.
2. З клієнта двічі списали кошти за одну й ту саму покупку.
3. Мерчант списав неправильну суму.
4. Товари або послуги не були доставлені.
5. Товар прибув пошкодженим, несправним або не відповідає опису.
6. Мерчант відмовляється оформити повернення коштів, попри обґрунтовану претензію.

У таких випадках покупець не змушений покладатися лише на готовність мерчанта вирішити проблему. Він може звернутися до свого банку-емітента й попросити переглянути транзакцію. Якщо спір буде схвалено, власник картки може отримати кошти назад, що робить карткові платежі безпечнішими та надійнішими.

Цей захист є однією з причин, чому споживачі почуваються впевненіше, коли оплачують карткою, особливо онлайн або під час покупки у нового мерчанта. Чарджбеки допомагають знизити ризик фінансових втрат і дають клієнтам додатковий рівень захисту від шахрайства, помилок мерчантів і невдалих доставок.

Водночас чарджбеки не мають замінювати звичайні процеси повернення коштів. У більшості випадків покупці повинні спочатку спробувати вирішити проблему безпосередньо з мерчантом. Але коли це неможливо або коли транзакція виглядає шахрайською, чарджбек надає клієнту формальний спосіб захистити свої гроші [2].

Попри важливу роль у захисті прав споживачів, чарджбеки створюють суттєві негативні наслідки для торговців. Бізнес несе прямі фінансові втрати, включаючи вартість товарів або послуг, комісії за чарджбек та додаткові штрафні нарахування. Крім того, виникає значне операційне навантаження, пов’язане з необхідністю збору доказів, веденням спорів та взаємодією з платіжними системами. Високий рівень чарджбеків може призводити до

підвищення транзакційних витрат, посиленого моніторингу з боку платіжних мереж, а у крайніх випадках — до обмеження або припинення можливості приймання карткових платежів.

Кожен чарджбек створює прямі операційні витрати, які зазвичай становлять від 20 до 100 USD за один випадок, і ці комісії стягуються незалежно від результату розгляду спору. Це означає, що навіть у випадку успішного оскарження чарджбеку через процедуру репрезентменту, пов’язані адміністративні та операційні витрати не компенсуються. Окрім прямих витрат, торговці також зазнають непрямих втрат, серед яких — вартість повернених товарів або послуг, витрати на платіжну інфраструктуру, а також внутрішні витрати праці, пов’язані з обробкою спорів, підготовкою звітності та заходами протидії фроду. У сукупності ці витрати з часом можуть суттєво знижувати загальну рентабельність бізнесу, особливо в умовах високого обсягу транзакцій або підвищеного ризику операцій [3].

Коли рівень чарджбеків перевищує допустимі порогові значення, встановлені платіжними мережами, торговці потрапляють під дію формалізованих систем моніторингу, спрямованих на контроль системних ризиків. Наприклад, програма Mastercard Excessive Chargeback Program (ECP) оцінює торговців одночасно за абсолютною кількістю чарджбеків та відносними показниками, які виражаються у базисних пунктах. Цей показник розраховується як відношення кількості первинних чарджбеків до обсягу транзакцій попереднього місяця, помножене на 10 000. У межах цієї системи передбачено дві категорії ризику: Excessive Chargeback Merchant (ECM), яка активується при досягненні 100 чарджбеків і 150 базисних пунктів за місяць, та High Excessive Chargeback Merchant (HECM), що застосовується при 300 чарджбеках і 300 базисних пунктах або більше.

Після включення до відповідної категорії торговці зобов’язані знизити рівень диспутів у межах жорсткого моніторингового циклу. У разі недотримання вимог застосовуються поступові фінансові санкції. На другому та третьому місяцях моніторингу накладаються штрафи у розмірі 1 000 USD

щомісяця. Починаючи з четвертого по шостий місяць, сума штрафів зростає до 5 000 USD щомісяця. Додатково, з четвертого місяця Mastercard може застосовувати Issuer Recovery Assessment у розмірі 5 USD за кожен чарджбек, що перевищує 300 випадків на місяць. У разі подальшого недотримання вимог понад шість місяців санкції можуть суттєво посилюватися, включаючи обов'язкові заходи з усунення порушень за рахунок торговця та, у кінцевому підсумку, припинення обслуговування еквайрингового рахунку приблизно через дванадцять місяців систематичного порушення умов.

Аналогічну модель ризик-менеджменту застосовує Visa у межах програми Acquirer Monitoring Program (VAMP), однак вона інтегрує показники фроду та диспутів у єдину систему оцінювання ризику. У цій моделі VAMP Ratio формується як співвідношення повідомлень про фрод (TC40) та чарджбеків (TC15) до загального обсягу транзакцій, тоді як VAMP Enumeration Ratio відображає рівень карткового тестування відносно загальної кількості операцій. Хоча програма має превентивний характер і спрямована на зниження ризиків, вона водночас запроваджує багаторівневі вимоги до комплаєнсу, включаючи посилений моніторинг, збільшені резервні вимоги, додаткові операційні витрати та підвищені вимоги до обробки транзакцій для торговців, які перевищують встановлені порогові значення [5]. Обчислення VAMP рейтингу представлено у формулі 1.1

$$VAMP\ ratio = \frac{TC40+TC15}{\text{Кількість транзакцій}}, \qquad (1.1.1)$$

де $TC40$ — фродові транзакції;

$TC15$ — кількість чарджбеків;

кількість транзакцій — кількість транзакцій де оплата відбулась без фізичної присутності картки.

У межах як Mastercard, так і Visa систем, тривале недотримання встановлених вимог призводить до поступового посилення фінансових та операційних обмежень. Це може включати підвищення еквайрингових

комісій, обов’язкове впровадження додаткових механізмів протидії фроду, а також обмеження можливості обробки платежів. У крайніх випадках тривале перебування в зоні високого ризику може призвести до припинення обслуговування торговця, що фактично означає втрату можливості приймати карткові платежі.

Окрім прямих санкцій, значним є й системний ефект. Торговці, які потрапляють до категорії високого ризику, можуть бути внесені до галузевих баз даних, зокрема MATCH список (Member Alert to Control High-Risk Merchants), який використовується еквайринговими банками для оцінки ризиків під час відкриття нових рахунків. Потрапляння до таких систем суттєво ускладнює відкриття нових торгових рахунків, оскільки фінансові установи можуть відмовляти в обслуговуванні або встановлювати значно жорсткіші умови. У ряді випадків юридична особа, пов’язана з початковим мерчант-рахунком, також підпадає під довгострокові обмеження, що фактично унеможливлює відкриття нових рахунків у майбутньому. Це змушує бізнес створювати нові юридичні структури, що, у свою чергу, призводить до додаткових витрат на реєстрацію, комплаєнс-перевірки, банківське обслуговування та інтеграцію з платіжними системами.

Таким чином, чарджбеки являють собою не лише прямі витрати на рівні 20–100 USD за транзакцію, але й складний багаторівневий ризик, який може ескалювати до регуляторного моніторингу, систем штрафів, втрати можливості обробки платежів та довгострокових обмежень фінансової діяльності та безперервності бізнесу.

Чарджбек є механізмом із подвійною природою. З одного боку, він забезпечує ефективний захист споживачів і підвищує довіру до карткових платіжних систем. З іншого боку, він формує значні фінансові та операційні ризики для торговців, особливо в контексті фроду та дружнього фроду. Саме ця суперечність зумовлює необхідність розвитку сучасних систем запобігання фроду та раннього виявлення ризикових транзакцій. У цьому контексті особливу роль відіграють методи машинного навчання, які активно

застосовуються для зменшення рівня фроду та мінімізації кількості чарджбеків, що буде розглянуто у подальших розділах дослідження.

1.4 Типи чарджбеків та фрод як ключовий фактор їх виникнення

Цифрові платежі становлять значну частку активності мерчантів. 63% обсягу транзакцій мерчантів припадає на цифрові покупки, наприклад онлайн-платежі та платежі в мобільних застосунках. Із розширенням цих платіжних каналів мерчанти стикаються з більшим ризиком непорозумінь із клієнтами, шахрайської активності та платіжних спорів [6]. Основні типи чарджбеків представлено на рисунку 1.2.

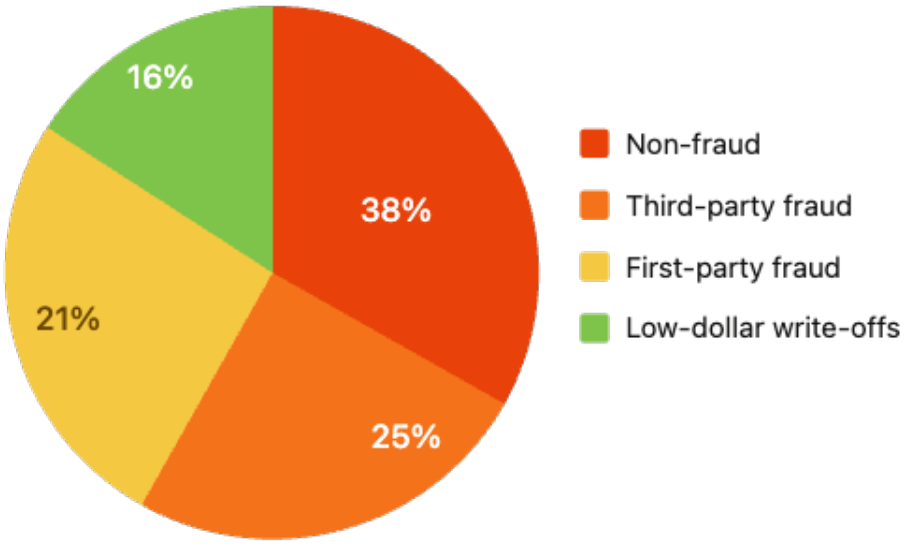

Рисунок 1.2 – Основні типи чарджбеків [6]

Шахрайство третіх осіб, також відоме як справжній фрод, відбувається тоді, коли хтось інший, а не законний власник картки, використовує викрадені дані картки або платіжні облікові дані без дозволу. У таких ситуаціях власник картки справді є жертвою шахрайства.

Поширені приклади включають онлайн-покупки, здійснені з використанням викрадених даних картки, захоплення акаунтів, неавторизовані транзакції в застосунках, а також замовлення, оформлені з використанням скомпрометованих логінів або платіжної інформації.

Цей тип шахрайства особливо поширений у транзакціях без фізичної присутності картки, оскільки покупцеві не потрібно фізично пред’являти картку. Для мерчантів такі чарджбеки може бути складно оскаржити, якщо вони не можуть надати переконливі докази, наприклад результати перевірки на шахрайство, дані автентифікації, інформацію про пристрій, записи моніторингу транзакцій або підтвердження того, що законний власник картки був причетний до покупки.

Шахрайство першої сторони, або дружній фрод, відбувається тоді, коли справжній клієнт здійснює дійсну покупку, але згодом оскаржує транзакцію як неавторизовану або шахрайську. Це може бути навмисним, наприклад коли клієнт намагається залишити товар собі й водночас повернути кошти. Також це може бути випадковим — наприклад, коли клієнт забуває про покупку, не впізнає платіжний дескриптор або неправильно розуміє умови повернення коштів. Такі спори є складними для мерчантів, оскільки початкова транзакція може виглядати цілком легітимною, але пізніше чарджбек подає справжній власник картки.

Деякі чарджбеки не пов’язані з шахрайством. Нешахрайські чарджбеки виникають через реальні проблеми з товаром, послугою, доставкою, оплатою або процесом повернення коштів. Наприклад, клієнт може оскаржити платіж, якщо товар так і не надійшов, продукт був пошкодженим або дефектним, мерчант списав неправильну суму, повернення коштів затримується або не було здійснене, або бронювання, пов’язане з подорожами, постраждало через скасування чи проблеми з обслуговуванням. Мерчанти у сфері подорожей і готельно-ресторанного бізнесу мають найвищу середню вартість чарджбеку — 120 доларів США. Одна з причин полягає в тому, що клієнти часто бронюють через сторонні платформи й можуть не розуміти, до якої компанії

звертатися, коли щось іде не так. Через нечітку логіку та непрозору комунікацію із підтримкою сервісів клієнти можуть звертатися до свого банку замість того, щоб вирішувати питання безпосередньо з мерчантом [6].

Списання невеликих сум відбувається тоді, коли мерчант або емітент вирішує, що оскаржувана транзакція не варта витрат на розслідування або оскарження. Замість того щоб витрачати час і ресурси на репрезентмент, бізнес приймає збиток.

За оцінками індустрії, приблизно 21% випадків чарджбеків пов'язані саме з справжнім фродом, що робить його значущим, хоча і не домінуючим фактором виникнення спорів [6]. Ключова особливість цього типу полягає в тому, що транзакція відбувається до моменту, коли власник картки усвідомлює факт шахрайства, що суттєво обмежує можливості її постфактум врегулювання з боку торговця. Попри наявність інструментів, таких як багатофакторна автентифікація, поведінковий аналіз і системи моніторингу транзакцій у реальному часі, основна складність полягає в обмеженій видимості реального власника платіжного засобу на момент здійснення операції. У зв'язку з цим ефективна протидія потребує превентивних механізмів, що працюють на етапі авторизації транзакції.

На противагу цьому, дружній фрод становить приблизно 25% усіх чарджбеків і останніми роками демонструє тенденцію до зростання в умовах розвитку електронної комерції [6]. На відміну від справжнього фроду, транзакція є легітимною на момент проведення, а спір виникає пізніше через поведінкові, когнітивні або ситуативні фактори. З операційної точки зору дружній фрод тісно пов'язаний не стільки з безпекою платіжної інфраструктури, скільки з якістю клієнтського досвіду та комунікації між продавцем і покупцем.

Фрод призводить до чарджбеків також тому, що платіжна система історично побудована навколо пріоритетного захисту власника картки. Якщо клієнт заявляє, що транзакція була несанкціонованою або неправомірною, банк-емітент має формальний механізм для її оскарження. У результаті ризик

часто переноситься на продавця, який повинен доводити легітимність транзакції, надання товару або послуги, дотримання умов продажу та наявність взаємодії з клієнтом.

1.5 Можливості впливу на дружній фрод

Дружній фрод є особливим типом фродових чарджбеків, оскільки він виникає не через несанкціоноване використання картки третьою особою, а через дії самого законного власника платіжного інструменту. У таких випадках клієнт фактично здійснює транзакцію, отримує товар або послугу, однак згодом оскаржує платіж через банк. Це може бути як навмисна дія з метою отримання фінансової вигоди, так і ненавмисний результат непорозуміння, незадоволення сервісом або недостатньо прозорої комунікації з боку продавця [8].

На відміну від справжнього фроду, дружній фрод є частково контрольованим з боку бізнесу. Його не завжди можна повністю усунути, оскільки частина покупців може свідомо зловживати механізмом чарджбеків. Проте значну кількість таких випадків можна зменшити шляхом покращення продукту, клієнтського досвіду, процесів підтримки та комунікації після здійснення покупки.

Одним із найважливіших способів зниження дружнього фроду є використання зрозумілих і впізнаваних платіжних дескрипторів. Часто клієнти оскаржують транзакції не тому, що вони справді були шахрайськими, а тому, що не можуть ідентифікувати продавця у банківській виписці. Якщо назва компанії у виписці відрізняється від назви бренду, сайту або продукту, клієнт може помилково вважати платіж несанкціонованим. У такому випадку правильно налаштований платіжний дескриптор може суттєво знизити кількість помилкових оскаржень.

Другим важливим напрямом є прозора комунікація умов покупки. Клієнт має чітко розуміти, за що саме він платить, яку суму буде списано, коли відбудеться списання, чи є платіж одноразовим або регулярним, які умови скасування підписки та повернення коштів. Особливо це важливо для підпискових бізнес-моделей, де значна частина чарджбеків виникає через те, що клієнт забуває про автоматичне продовження послуги або не очікує повторного списання.

Для зниження таких ризиків бізнес може впроваджувати нагадування перед повторним списанням, повідомлення про успішну оплату, простий доступ до історії платежів, а також зрозумілий механізм скасування підписки. Якщо клієнт легко може зрозуміти причину списання та самостійно керувати послугою, ймовірність звернення до банку значно зменшується.

Ще одним важливим фактором є якість опису товару або послуги. Якщо продуктова сторінка містить нечітку, неповну або перебільшену інформацію, клієнт може сформувати очікування, які не відповідають реальному досвіду. У результаті після отримання товару чи послуги він може бути незадоволений і замість звернення до продавця ініціювати чарджбек. Тому точні описи, реалістичні зображення, зрозумілі характеристики, умови доставки та обмеження використання продукту є важливими інструментами профілактики дружнього фроду.

Окрему роль відіграє політика повернення коштів. Якщо політика є складною, прихованою або незрозумілою, клієнт може сприймати звернення до банку як простіший і швидший спосіб повернути гроші. Навпаки, прозора, доступна та швидка процедура повернення коштів може зменшити кількість чарджбеків, оскільки клієнт має зрозумілий канал вирішення проблеми напряму з продавцем. Для бізнесу повернення коштів часто є менш витратним, ніж чарджбек, оскільки не призводить до штрафних списань, не погіршує рівень чаоджбеків і не створює ризику потрапляння до моніторингових програм платіжних систем.

Важливим елементом також є якість клієнтської підтримки. Якщо клієнт не може швидко отримати відповідь, знайти контактні дані продавця або вирішити проблему через підтримку, він із більшою ймовірністю звернеться до банку. Наявність доступної підтримки, швидке реагування на запити, зрозумілі шаблони відповідей, автоматичні підтвердження звернень і можливість ескалації проблеми можуть суттєво знизити ризик чарджбеків.

Також бізнес може впливати на дружній фрод через якісну доказову базу. Навіть якщо чарджбек уже виник, продавець має більше шансів успішно його оскаржити, якщо зберігає підтвердження покупки, IP-адресу, історію логінів, підтвердження доставки, переписку з клієнтом, прийняття умов коритувачем, інформацію про використання сервісу або інші докази легітимності транзакції. Це не усуває сам факт чарджбеку, але підвищує ефективність репрезентменту і може зменшити фінансові втрати.

Таким чином, дружній фрод є типом фроду, на який бізнес може впливати через внутрішні процеси. Його зниження залежить не лише від антифрод-систем, а й від якості продукту, прозорості платежів, клієнтської підтримки, політик повернення та комунікації з користувачем. Саме тому боротьба з дружнім фродом повинна розглядатися як комплексна задача, що поєднує платіжний менеджмент ризиків, досвід користувача та операційну ефективність [7].

1.6 Боротьба зі справжнім фродом

На відміну від дружнього фроду, справжній фрод має зовнішній характер і виникає внаслідок несанкціонованого використання платіжних даних власника картки. Основна проблема полягає в тому, що на момент здійснення платежу продавець не завжди може достовірно визначити, чи є особа, яка проводить транзакцію, справжнім власником картки. У цифровому

середовищі взаємодія між покупцем і продавцем часто відбувається дистанційно, без фізичної перевірки особи. Тому бізнес змушений покладатися на технічні сигнали: платіжні дані, IP-адресу, країну, пристрій, поведінкові патерни, історію попередніх покупок, результати 3DS, AVS, CVV та інші індикатори ризику. Однак жоден із цих сигналів не гарантує повної точності.

Після того як шахрайська транзакція вже проведена, можливості продавця значно зменшуються. Якщо товар був відправлений або цифрова послуга вже надана, бізнес фактично не може повністю відновити початковий стан. У випадку фізичних товарів існує ризик втрати товару, витрат на доставку та неможливості його повернення. У випадку цифрових товарів, підписок або онлайн-сервісів послуга може бути спожита одразу після оплати, що робить повернення активу практично неможливим.

Коли справжній власник картки виявляє несанкціоноване списання, він звертається до банку-емітента. Для банку така транзакція виглядає як потенційно шахрайська, а отже запускається процедура диспуту або чарджбеку. У цій ситуації продавець часто опиняється у слабкій позиції, оскільки клієнт справді не здійснював покупку. Навіть якщо продавець діяв добросовісно і виконав замовлення, сам факт несанкціонованого використання картки може бути достатньою підставою для повернення коштів власнику картки.

Репрезентмент у випадку справжнього фроду має обмежену ефективність. Якщо продавець не може довести, що транзакція була авторизована справжнім власником картки або що були виконані всі необхідні вимоги платіжної системи щодо автентифікації, шанси виграти спір зменшуються. Навіть у випадках, коли продавець має певні докази виконання замовлення, вони не завжди спростовують головний аргумент: власник картки не давав згоди на проведення транзакції.

Ще одне обмеження постфактум-боротьби полягає в часовому лагу між транзакцією та виявленням фроду. Власник картки може помітити

несанкціоноване списання через кілька днів або навіть тижнів. За цей час шахрай може вже отримати товар, використати послугу або вивести цінність із системи. Для продавця це означає, що реакція після факту не запобігає фінансовій втраті, а лише дозволяє частково управляти наслідками.

Крім того, постфактум-обробка фроду створює додаткове операційне навантаження. Бізнес повинен аналізувати чарджбек, збирати докази, готувати відповідь, взаємодіяти з еквайром або платіжною системою, оновлювати внутрішні правила ризику та документувати кейс. У великих обсягах така робота потребує окремої команди або автоматизованої системи обробки диспутів. Проте навіть ефективна операційна обробка не вирішує головної проблеми — фродова транзакція вже відбулася [6].

Таким чином, постфактум-підхід до справжнього фроду є переважно реактивним. Він дозволяє аналізувати вже здійснені інциденти, покращувати правила ризику, посилювати перевірки та формувати історичну базу для майбутнього навчання моделей. Однак він не здатний повністю запобігти первинній фінансовій втраті, якщо фрод не був виявлений до завершення транзакції.

Саме тому боротьба зі справжнім фродом має бути зосереджена не лише на обробці наслідків, а передусім на попередженні. Бізнесу необхідно виявляти ризикові транзакції до моменту їх авторизації або до надання товару чи послуги. Це може включати використання інструментів, які дозволяють оцінювати ризик у режимі реального часу.

Висновки до розділу 1

У цьому розділі було розглянуто чарджбек як важливий механізм захисту власників платіжних карток, який історично виник для підвищення довіри до карткових платежів і забезпечення можливості оскарження

несанкціонованих або спірних транзакцій. Водночас у сучасному цифровому середовищі чарджбек виконує не лише захисну функцію для споживачів, а й створює значні фінансові, операційні та репутаційні ризики для бізнесу. Для мерчантів кожен чарджбек може означати втрату доходу, додаткові комісії, витрати на обробку спорів, а в разі перевищення встановлених порогів — потрапляння до моніторингових програм платіжних систем і ризик обмеження платіжного процесингу.

Було встановлено, що причини виникнення чарджбеків є різними: вони можуть бути пов’язані з помилками торговця, реальними проблемами з товаром або послугою, дружнім фродом чи справжнім фродом. Особливу увагу в межах дослідження приділено розділенню дружнього та справжнього фроду, оскільки ці явища мають різну природу та потребують різних підходів до управління. Дружній фрод часто пов’язаний із поведінкою самого клієнта, непорозуміннями, невпізнаваними платіжними дескрипторами, недостатньо прозорими умовами оплати, складною процедурою повернення коштів або низькою якістю клієнтської підтримки. Тому значну частину таких випадків бізнес може зменшити шляхом покращення комунікації, клієнтського досвіду, політик повернення та доказової бази.

Натомість справжній фрод має зовнішній характер і виникає тоді, коли платіжні дані використовуються третьою особою без згоди власника картки. На відміну від дружнього фроду, він значно менш контрольований після проведення транзакції, оскільки на момент виявлення шахрайства товар або послуга вже можуть бути надані, а фінансова втрата відбулась. У таких випадках репрезентмент має обмежену ефективність, адже власник картки справді не авторизував операцію. Саме тому боротьба зі справжнім фродом повинна мати насамперед превентивний характер.

Отже, ключовим висновком розділу є те, що зниження рівня чарджбеків потребує диференційованого підходу. Дружній фрод доцільно мінімізувати через покращення внутрішніх бізнес-процесів, прозорість платежів, якісну підтримку клієнтів і зрозумілі правила повернення коштів. Водночас

справжній фрод потребує раннього виявлення ризикових транзакцій ще до їх завершення. Для цього бізнес має впроваджувати ефективні системи перевірки платежів, інструменти автентифікації, моніторинг поведінкових і технічних сигналів, а також моделі прогнозування фродових операцій. Саме налаштування таких превентивних механізмів є одним із найважливіших напрямів зменшення фінансових втрат, скорочення кількості чарджбеків і забезпечення стабільності платіжної діяльності мерчанта. Основне завдання бізнесу полягає не в тому, щоб ефективно реагувати на вже здійснений фрод, а в тому, щоб максимально точно виявляти й блокувати ризикові транзакції до того, як вони призведуть до чарджбеку. Таким чином, структурні відмінності між типами фроду мають принципове значення для побудови ефективних стратегій ризик-менеджменту та шахрайство третьої сторони потребує впровадження превентивних, високоточних систем виявлення в реальному часі, інтегрованих безпосередньо в платіжний процес.

РОЗДІЛ 2 МАТЕМАТИЧНІ СИСТЕМИ ТА ПІДХОДИ ВИРІШЕННЯ ПРОБЛЕМИ ПОПЕРЕДЖЕННЯ ВИНИКНЕННЯ ФРОДОВИХ ТРАНЗАКЦІЙ

Виявлення шахрайських фінансових транзакцій є одним із найскладніших і водночас найбільш практично значущих напрямів застосування методів машинного навчання у сфері фінансових технологій. Перш ніж переходити до побудови алгоритмічних рішень, необхідно сформувати чітку математичну та концептуальну основу, яка б точно описувала структуру й особливості цієї задачі. У цьому розділі розглядається формалізація проблеми з позицій бінарної класифікації з учителем, виявлення аномалій та ймовірнісного підходу, а також подаються відповідні математичні позначення і функції втрат, що використовуються під час навчання та оцінювання моделей.

2.1 Математичне формулювання задачі виявлення шахрайства в межах машинного навчання

2.1.1 Формулювання задачі виявлення фроду у парадигмі навчання з учителем

У парадигмі навчання з учителем модель навчається на розміченому наборі даних D, що складається з N пар вхід–вихід. Формально навчальна вибірка визначається відповідно формулі 2.1.1.

$$D = \{ (x^{-i}, y^{i}) \}_{i=1}^{N}, \qquad (2.1.1)$$

де D — повний розмічений навчальний набір даних, що використовується для побудови моделі;

$x^i \in \mathbb{R}^d$ — вектор ознак і-ї транзакції, що містить d числових атрибутів, таких як сума транзакції, час, що минув, та PCA-перетворені оцінки поведінки власника картки;

$y^i \in \{0,1\}$ — бінарна мітка класу і-ї транзакції: $y^i = 1$ означає шахрайську транзакцію; $y^i = 0$ означає легітимну транзакцію;

N — загальна кількість розмічених навчальних прикладів у наборі даних;

d — розмірність простору вхідних ознак (кількість ознак на транзакцію).

Мета навчання з учителем у цьому контексті полягає у наближенні невідомої цільової функції $f^*: \mathbb{R}^d \rightarrow \{0,1\}$ шляхом навчання гіпотетичної функції $h: \mathbb{R}^d \rightarrow \{0,1\}$ на навчальних даних D таким чином, щоб мінімізувати помилку узагальнення на нових тестових прикладах. Ця функція параметризована вектором ваг $\theta \in \mathbb{R}^d$, а процес навчання полягає у оптимізації θ відносно відповідної функції втрат $L(\theta)$ на навчальному розподілі.

2.1.2 Бінарна класифікація

Виявлення шахрайства з кредитними картками є по суті задачею бінарної класифікації. Для транзакції, представленої вектором ознак x, класифікатор відносить її до одного з двох взаємовиключних класів. Загальна функція прийняття рішення класифікації представлена у формулі 2.1.2.

$$f(x) = sign(g(x) - \tau), \tag{2.1.2}$$

де $f(x)$ — функція бінарного прогнозування, що відображає вектор ознак x на мітку класу {0, 1};

$g : \mathbb{R}^d \to \mathbb{R}$ — дійсна скалярна функція оцінки, що навчається моделлю та відображає міру впевненості в тому, що транзакція x є шахрайською; більші значення означають більшу імовірність шахрайства;

τ — поріг прийняття рішення — скалярне значення, вище якого транзакція класифікується як шахрайська; за замовчуванням $\tau = 0.5$ для ймовірнісних класифікаторів, але при незбалансованих наборах даних оптимальне τ визначається пошуком по валідаційній вибірці;

$sign(\cdot)$ — функція знаку, що повертає +1 якщо аргумент невід'ємний і −1 в іншому випадку.

2.1.3 Функції втрат для класифікації фроду

Навчання моделі здійснюється шляхом мінімізації функції втрат L(θ), що кількісно оцінює помилку прогнозування. Канонічна функція втрат для бінарної ймовірнісної класифікації — бінарна крос-ентропія (log loss) представлена у формулі 2.1.3.

$$L(\theta) = -\frac{1}{N} \cdot \Sigma_{i=1}^{N} [\, y^i \cdot log(h\_\theta(x^i)) + (1 - y^i) \cdot log(1 - h\_\theta(x^i)) \,], \quad (2.1.3)$$

де $L(\theta)$ — бінарна крос-ентропія (log) функція втрат — скалярна цільова функція, що мінімізується під час навчання моделі шляхом налаштування вектора параметрів θ;

N — загальна кількість навчальних прикладів;

$y^i \in \{0,1\}$ — справжня бінарна мітка і-ї навчальної транзакції (1 = шахрайство, 0 = легітимна);

$h\_\theta(x^i)$ — прогнозована моделлю імовірність шахрайства для і-го навчального прикладу;

log(·) — натуральний логарифм; коли прогнозована імовірність $\hat{p}$ наближається до 0 для справді позитивного прикладу, $-\log(\hat{p}) \to +\infty$, накладаючи все більш суворий штраф, що спонукає оптимізатор до більшої впевненості у прикладах фроду;

$\Sigma_{i=1}^{N}$ — підсумовування по всіх N навчальних прикладах;

(1/N) — нормалізуючий множник, що перетворює загальні втрати на середні втрати на приклад.

Для методів ансамблю градієнтного бустингу (XGBoost, LightGBM, CatBoost) цільова функція включає і функцію підгонки даних, і штраф регуляризації та представлена у формулі 2.1.4 [13, 14].

$$Obj(\theta) = \Sigma_{i=1}^{N} L(y_i, \hat{y}_i) + \Sigma_k \Omega(f_k), \tag{2.1.4}$$

де *Obj(θ)* — загальна регуляризована цільова функція для ансамблю градієнтного бустингу;

$L(y_i, \hat{y}_i)$ — функція втрат на рівні прикладу (наприклад, бінарна крос-ентропія) для навчального прикладу і з справжньою міткою y_i та прогнозованою оцінкою $\hat{y}_i$;

$\Sigma_k \Omega(f_k)$ — сума штрафів регуляризації по всіх K деревах в ансамблі; контролює складність моделі для запобігання перенавчанню на малій навчальній вибірці шахрайства;

$\Omega(f_k)$ — штраф регуляризації для дерева k: $\Omega(f_k) = \gamma T_k + (1/2)\lambda \|w_k\|^2$, що штрафує як кількість листів, так і величину ваг листів;

γ — мінімальне зменшення функції втрат, необхідне для розбиття; більше γ дає більш дрібні, консервативніші дерева, що краще узагальнюють на незбалансованих даних;

T_k — кількість вузлів-листів (термінальних вузлів) у *k*-му дереві;

λ — коефіцієнт L2-регуляризації, застосований до ваг листів — зменшує оцінки листів до нуля, знижуючи ризик перенавчання на рідкісному класі шахрайства;

w_k — вектор значень ваг (оцінок) листів, призначених кожному листу k-го дерева.

2.1.4 Матриця помилок та основні метрики класифікації

Продуктивність бінарного класифікатора характеризується матрицею помилок — таблицею суміжності 2 на 2, що перераховує чотири можливі комбінації прогноз–істина. Таблиця 2.1 представляє структуру матриці помилок для бінарної класифікації шахрайства.

Таблиця 2.1 – Матриця помилок для бінарної класифікації шахрайства

	Справжній фрод (y=1)	*Справжня легітимна (y=0)*
Прогнозований фрод (ŷ=1)	Справжній позитив (TP)	Хибний позитив (FP)
Прогнозована легітимна (ŷ=0)	Хибний негатив (FN)	Справжній негатив (TN)

Справжній позитив (TP) — класифікатор правильно ідентифікує фродову транзакцію як фрод. Кожен TP представляє успішно перехоплену шахрайську операцію та запобіжений чарджбек.

Справжній негатив (TN) — класифікатор правильно ідентифікує легітимну транзакцію як легітимну. TN складає переважну більшість правильних прогнозів з огляду на дисбаланс у датасетах з транзакціями.

Хибний позитив (FP) — класифікатор помилково позначає легітимну транзакцію як фродову, що призводить до відхилення справжньої транзакції, незручностей для клієнта та збільшення операційних витрат на перевірку.

Хибний негатив (FN) — класифікатор не виявляє шахрайську транзакцію. Це найкритичніший тип помилки: фродова операція проходить, що призводить до прямих фінансових збитків, відповідальності за чарджбек, штрафних санкцій карткових мереж та репутаційної шкоди платіжній установі.

Чотири основні метрики класифікації, отримані з матриці помилок, формально визначені в поданих нижче формулах. Точність (accuracy) вимірює загальну частку правильно класифікованих транзакцій за формулою 2.1.5

$$Accuracy = (TP + TN) / (TP + TN + FP + FN), \quad (2.1.5)$$

де *TP* — справжні позитиви — фродові транзакції, правильно класифіковані як фродові;

TN — справжні негативи — легітимні транзакції, правильно класифіковані як легітимні;

FP — хибні позитиви — легітимні транзакції, помилково класифіковані як фродові;

FN — хибні негативи — фродові транзакції, помилково класифіковані як легітимні;

TP+*TN*+*FP*+*FN* — загальна кількість транзакцій у наборі для оцінювання.

Прецизійність (precision) вимірює частку прогнозів шахрайства, що є правильними, відображаючи рівень хибних тривог класифікатора за формулою 2.1.6.

$$Precision = TP / (TP + FP), \quad (2.1.6)$$

де *TP* — справді фродові транзакції, правильно позначені класифікатором;

FP — легітимні транзакції, помилково позначені як;

TP+*FP* — загальна кількість транзакцій, прогнозованих як фродові.

Повнота (recall, чутливість, true positive rate) розраховується за формулою 2.1.7 та вимірює частку справді шахрайських транзакцій, що правильно виявляються, і є основною бізнес-критичною метрикою у виявленні шахрайства.

$$Recall = TPR = TP / (TP + FN), \quad (2.1.7)$$

де *TP* — фродові транзакції, правильно виявлені класифікатором;
FN — фродові транзакції, пропущені класифікатором; кожен FN представляє потенційний чарджбек і прямі фінансові збитки;
TP+FN — загальна кількість справді фродових транзакцій у наборі для оцінювання.

F1-міра — це гармонійне середнє прецизійності та повноти, що забезпечує єдиний збалансований показник продуктивності і розраховується за формулою 2.1.8

$$F1 = 2 \cdot (Precision \times Recall) / (Precision + Recall), \quad (2.1.8)$$

2.2 Аналіз моделей машинного навчання для виявлення фродових транзакцій

У цьому розділі розглядаються основні моделі машинного навчання, що застосовуються для виявлення фродових транзакцій у платіжних системах. З огляду на складність задачі, пов’язану з високою дисбалансованістю класів, змінністю шахрайських патернів та великою кількістю ознак, особлива увага приділяється порівнянню різних підходів до класифікації та оцінюванню їхньої ефективності.

Зокрема, аналізуються як класичні методи машинного навчання, такі як логістична регресія та дерева рішень, так і більш складні ансамблеві моделі, включаючи Random Forest, Gradient Boosting та LightGBM. Окремо розглядаються їхні особливості в контексті роботи з незбалансованими даними та здатність до виявлення рідкісних, але критично важливих подій.

2.2.1 Логістична регресія

Логістична регресія є стандартною моделлю для бінарної класифікації, що будується на розрахунках за формулою 2.2.1 [9].

$$\partial L/\partial \theta_j = (1/N) \cdot \Sigma_{i=1}^{N} (h\_\theta(x_i) - y_i) \cdot x_{ij}, \quad (2.2.1)$$

де $\partial L/\partial \theta_j$ — частинна похідна крос-ентропійної функції втрат відносно j-го параметра моделі; компонента градієнта, що використовується для оновлення θ_j під час градієнтного спуску;

$h\_\theta(x_i)$ — прогнозована моделлю імовірність фроду для транзакції і при поточному векторі параметрів θ;

$h\_\theta(x_i) - y_i$ — залишок прогнозу (помилка) для транзакції і — сигнал, що керує оновленням ваги; позитивний, коли модель переоцінює імовірність шахрайства, негативний, коли недооцінює;

x_{ij} — j-те значення ознаки і-ї навчальної транзакції;

$(1/N)$ — нормалізуючий множник, що усереднює градієнт по всіх навчальних прикладах, стабілізуючи збіжність.

Логістична регресія належить до класу узагальнених лінійних моделей (GLM), які застосовуються для моделювання ризиків у різних фінансових задачах, зокрема для оцінки ймовірності настання небажаних подій [27].

Переваги логістичної регресії для виявлення шахрайства включають обчислювальну ефективність, нативний ймовірнісний вивід, високу інтерпретованість та підтримку L1/L2-регуляризації. Основним недоліком є припущення про лінійну межу рішення у вихідному просторі ознак, що обмежує її здатність вловлювати складні нелінійні шаблони фроду.

2.2.2 Дерева рішень

Дерева рішень будують ієрархічний набір бінарних правил розбиття за допомогою рекурсивного розбиття. У кожному внутрішньому вузлі n оптимальна ознака розбиття j* та поріг t* вибираються шляхом мінімізації зваженої домішковості результуючих дочірніх вузлів за формулою 2.2.2 [10, 11].

$$(j^*, t^*) = \operatorname{argmin}_{j,t} [\ (|S_L|/|S|)\cdot H(S_L) + (|S_R|/|S|)\cdot H(S_R)\ ],\quad (2.2.2)$$

де j^* — індекс оптимальної ознаки розбиття, вибраної у вузлі n;

t^* — оптимальний поріг розбиття для ознаки j*; приклади з $x_j \le t^*$ переходять до лівого дочірнього вузла, решта — до правого;

S — множина навчальних прикладів, що досягають поточного вузла n під час побудови дерева;

S_L, S_R — ліва та права дочірні підмножини, створені бінарним розбиттям на ознаці j* при порозі t*;

$|S_L|, |S_R|$ — кількість прикладів у лівому та правому дочірніх вузлах відповідно;

$H(\cdot)$ — міра домішковості вузла — або домішковість Джині, або інформаційна ентропія, розрахунок яких наведено у формулах 2.2.3 та 2.2.4 відповідно [10, 11].

$$H\_Gini(S) = 1 - \Sigma_k p_k^2, \quad (2.2.3)$$

де $H\_Gini(S)$ — домішковість Джині вузла S дорівнює 0 для чистого вузла (всі одного класу) та досягає максимуму 0.5 для ідеально збалансованого бінарного розбиття;

p_k — частка прикладів класу k серед усіх прикладів у вузлі S; для бінарної класифікації шахрайства $k \in \{0, 1\}$;

Σ_k — підсумовування по всіх класах, присутніх у вузлі.

$$H\_entropy(S) = -\Sigma_k p_k \cdot log_2(p_k), \quad (2.2.4)$$

де $H\_entropy(S)$ — інформаційна ентропія вузла S дорівнює 0 для чистого вузла та досягає максимуму 1 біт для ідеально збалансованого бінарного розбиття;

p_k — частка прикладів класу k у вузлі S;

$log_2(p_k)$ — логарифм за основою 2, що вимірює інформаційний вміст у бітах; за угодою, $0 \cdot \log_2(0) = 0$.

2.2.3 *Random Forest*

Random Forest будує ансамбль із T незалежних дерев рішень, де різноманітність між деревами досягається двома способами. По-перше, кожне дерево навчається на власній бутстреп-вибірці — підмножині навчальних даних, сформованій із поверненням. По-друге, у кожному вузлі розбиття розглядається лише випадково обрана підмножина ознак, а не весь їх набір. Ці два механізми забезпечують низьку кореляцію між деревами, завдяки чому усереднення їх прогнозів зменшує загальну дисперсію моделі порівняно з одним деревом рішень [12].

Окрім прогностичної функції, Random Forest надає оцінку важливості кожної ознаки — на основі сукупного зниження домішковості вузлів по всіх деревах ансамблю. Це робить модель частково інтерпретованою, що є цінним у контексті аналізу фродових транзакцій. Для роботи з незбалансованими даними, притаманними задачам виявлення шахрайства, модель підтримує зважування класів через параметр class_weight, що збільшує штраф за пропуск фродових транзакцій і підвищує повноту виявлення. Ансамблева оцінка імовірності фроду обчислюється за формулою 2.2.5 [12].

$$P\_RF(y=1|x) = (1/T) \cdot \Sigma_{t=1}^{T} P\_t(y=1|x), \quad (2.2.5)$$

де $P\_RF(y=1|x)$ — агрегована оцінка імовірності фроду Random Forest для транзакції x, обчислена шляхом усереднення по всіх T прогнозах окремих дерев;

T — загальна кількість дерев рішень в ансамблі;

$P\_t(y=1|x)$ — оцінка імовірності шахрайства t-го окремого дерева; рівномірно усереднюється по всіх T деревах;

$(1/T)$ — коефіцієнт усереднення — кожне дерево вносить рівний внесок до кінцевого прогнозу ансамблю;

$\Sigma_{t=1}^{T}$ — підсумовування по всіх T деревах у випадковому лісі.

2.2.4 XGBoost

XGBoost аналітично отримує оптимальні ваги листів, використовуючи апроксимацію другого порядку Тейлора функції втрат. Для кожного вузла-листа j у заданому дереві оптимальна вага обраховується за формулою 2.2.6 [14].

$$w^*_j = -(\Sigma_{i \in I_j} g_i) / (\Sigma_{i \in I_j} h_i + \lambda), \quad (2.2.6)$$

де w^*_j — аналітично оптимальна оцінка, призначена вузлу-листу j, отримана шляхом прирівнювання похідної розкладу Тейлора другого порядку Obj до нуля;

$\Sigma_{i \in I_j} g_i$ — сума градієнтів першого порядку $g_i = \partial L / \partial \hat{y}_i$ для всіх навчальних прикладів i, призначених листу j;

$\Sigma_{i \in I_j} h_i$ — сума градієнтів другого порядку $h_i = \partial^2 L / \partial \hat{y}_i^2$ для прикладів у листі j, надає інформацію про кривизну, що покращує швидкість збіжності;

λ — коефіцієнт L2-регуляризації у знаменнику для зменшення ваг листів до нуля, знижуючи ризик перенавчання на малій навчальній вибірці шахрайства;

I_j — множина навчальних прикладів, призначених листу j рішеннями про розбиття поточного дерева.

2.2.5 KNN

KNN класифікує транзакцію x шляхом голосування більшості серед k найближчих навчальних прикладів. Евклідова метрика відстані та функція прогнозування зазначені у формулах 2.2.7 та 2.2.8 відповідно [16, 17].

$$D(x_i, x_j) = \sqrt{(\Sigma_{k=1}^{n} (x_{ik} - x_{jk})^2)}, \quad (2.2.7)$$

$$\hat{y} = \text{argmax}_c \Sigma_{i \in N_k(x)} \mathbb{I}[y_i = c], \quad (2.2.8)$$

де $D(x_i, x_j)$ — евклідова відстань між транзакціями i та j у d-вимірному просторі ознак;

x_{ik}, x_{jk} — k-ті значення ознак транзакцій i та j відповідно;

n — кількість ознак (розмірність d простору ознак);

$\hat{y}$ — прогнозована мітка класу для запитуваної транзакції x;

$N_k(x)$ — множина k навчальних транзакцій, найближчих до x за евклідовою відстанню;

$\mathbb{I}[y_i = c]$ — індикаторна функція; дорівнює 1, якщо мітка сусіда i належить класу c, і 0 в іншому випадку;

argmax_c — повертає клас c, що отримав найбільше голосів серед k найближчих сусідів;

k — кількість найближчих сусідів для розгляду.

2.2.6 LightGBM

LightGBM використовує вибірку на основі однобічного градієнта (GOSS) для зосередження обчислювальних ресурсів на найбільш інформативних навчальних прикладах. Розрахунок приблизного інформаційного приросту для можливого розбиття при GOSS наведено у формулі 2.2.9 [15].

$$Gain = (1/2)\cdot[\ (\Sigma^{H} g_i + (1-a)/b\cdot\Sigma^{L} g_i)^2 \ /\ (|H| + (1-a)/b\cdot|L| + \lambda) - \ldots\ ], \quad (2.2.9)$$

де *Gain* — приблизний інформаційний приріст для можливої ознаки розбиття та порогу, обчислений лише на основі вибірки GOSS для зниження обчислювальних витрат при збереженні точності;

H — множина прикладів з великими градієнтами — всі навчальні приклади, чиї $|g_i|$ перевищують поріг, визначений верхньою a-частиною за величиною градієнтів; завжди включаються повністю;

L — випадково вибрана підмножина прикладів з малими градієнтами; лише частка b з них включається за раунд, де b зазвичай 0.1–0.2;

$\Sigma^H g_i$ — сума градієнтів по всіх прикладах з великими градієнтами — вони несуть найбільше інформації про поверхню функції втрат і завжди зберігаються;

$\Sigma^L g_i$ — сума градієнтів по вибраних прикладах з малими градієнтами;

$(1-a)/b$ — коефіцієнт підсилення, що компенсує субвибірку прикладів з малими градієнтами; зберігає статистичну узгодженість оцінки приросту;

$|H|$, $|L|$ — потужності множини прикладів з великими градієнтами та вибраної підмножини прикладів з малими градієнтами відповідно;

λ — коефіцієнт L2-регуляризації листів.

2.2.7 Багатошаровий персептрон

Багатошаровий перцептрон — це клас штучних нейронних мереж прямого поширення сигналу, що відображає вхідний вектор на вихід через послідовність навчених перетворень. Ключова ідея полягає в тому, що стекуванням кількох шарів простих нелінійних операцій та їх спільним навчанням через градієнтний спуск мережа може апроксимувати довільно складні функції — властивість, відома як теорема про універсальну апроксимацію [18].

На відміну від лінійних моделей, що можуть провести лише пряму межу між фродовими та легітимними транзакціями, MLP навчає ієрархічні представлення: ранні шари виявляють прості комбінації вхідних ознак, тоді як глибші шари об'єднують ці комбінації у більш абстрактні патерни. Кожен шар перетворює вихід попереднього, поступово вдосконалюючи представлення доти, поки кінцевий вихід не зможе розділити два класи.

Процес навчання керується зворотним поширенням помилки: мережа робить прогноз, обчислює похибку відносно істинної мітки та поширює цю похибку назад через усі шари для коригування кожної ваги на невеликий крок

у напрямку зменшення похибки. Повторення цього процесу протягом тисяч міні-пакетів транзакцій і складає процес навчання [18].

У кожному прихованому шарі ℓ мережа застосовує лінійне перетворення, за яким слідує пакетна нормалізація та нелінійна функція активації ReLU. Ця послідовність — лінійне - нормалізація - активація — є фундаментальним будівельним блоком.

$$z^{(\ell)} = W^{(\ell)}\, h^{(\ell-1)} + b^{(\ell)}, \tag{2.2.8}$$

де $z^{(\ell)}$— pre-activation vector at layer l;

$W^{(\ell)} \in \mathbb{R}^{d_\ell \times d_{\ell-1}}$ — навчені ваги матриці що поєднують шар l-1 з шаром l

x_{ik}, x_{jk} — k-ті значення ознак транзакцій і та j відповідно;

$h^{(\ell-1)}$— активація попереднього шару;

$b^{(\ell)} \in \mathbb{R}^{d_\ell}$— навчений вектор зсуву.

Матриці ваг W^(ℓ) та вектори зсуву b^(ℓ) є навченими параметрами мережі — вони починають з випадкової ініціалізації та оновлюються ітеративно в процесі навчання. Функція активації ReLU (Rectified Linear Unit) вносить нелінійність, що дозволяє мережі апроксимувати нелінійні межі прийняття рішень. Без неї будь-який стек лінійних шарів зводиться до єдиного лінійного перетворення, і мережа була б еквівалентна логістичній регресії.

Вихід останнього шару проходить через сигмоїдну функцію для отримання ймовірності фроду. Значення вище порогу рішення t* класифікуються як фродові [18].

Пакетна нормалізація усуває нестабільність навчання, відому як внутрішнє коваріаційне зміщення: коли ваги одного шару оновлюються, розподіл вхідних даних наступного шару зміщується, змушуючи його постійно адаптуватися заново [19]. Пакетна нормалізація повертає кожну попередню активацію до нульового середнього та одиничної дисперсії по поточному міні-пакету, а потім перемасштабовує з навченими параметрами γ та β.

На практиці пакетна нормалізація дозволяє використовувати вищі темпи навчання, зменшує чутливість до ініціалізації ваг та діє як слабкий регуляризатор. Під час виведення шар використовує експоненційні ковзні середні статистик міні-пакетів, накопичених під час навчання, що забезпечує детермінованість прогнозів.

Dropout — це техніка регуляризації, яка з ймовірністю p випадково обнуляє активацію кожного нейрона під час навчання [20]. Це запобігає спільній адаптації нейронів — залежності від конкретних інших нейронів, що є активними, — та ефективно навчає неявний ансамбль 2^d підмереж із спільними вагами. Під час виведення всі нейрони активні, а активації масштабуються для збереження очікуваної величини (інвертований dropout):

2.2.8 Порівняльний аналіз розглянутих моделей

У таблиці 2.2 наведено порівняння шести моделей машинного навчання за шістьма характеристиками, які визначають придатність моделей до використання в системах виявлення фродових транзакцій. Інтерпретованість відображає, наскільки легко спеціаліст може пояснити окремі прогнози моделі регуляторам, службам комплаєнсу або власникам банківських карток, що є критично важливим аспектом у фінансовій сфері. Найвищий рівень інтерпретованості демонструють логістична регресія та дерева рішень.

Нелінійність описує здатність моделі виявляти складні нелінійні закономірності шахрайства у високовимірних ознаках. Найвищу ефективність у цьому аспекті демонструють моделі градієнтного бустингу — XGBoost та LightGBM.

Швидкість навчання характеризує обчислювальну ефективність моделей під час роботи з набором даних. LightGBM завдяки гістограмному

алгоритму побудови розщеплень забезпечує найвищу швидкість серед моделей.

Стійкість до дисбалансу класів оцінюється за наявністю вбудованих механізмів компенсації дисбалансу, або коригування апріорних ймовірностей. Такі механізми дозволяють уникати домінування більшості без необхідності зовнішнього збільшення даних. Модель LightGBM отримує найвищу оцінку, оскільки під час використання вбудованого зважування класів модель автоматично збільшує вагу градієнтів для об'єктів міноритарного класу — у даному випадку фродових транзакцій. Завдяки цьому рідкісні приклади фроду мають сильніший вплив на процес навчання, навіть без застосування зовнішніх методів балансування або ресемплінгу даних.

Таблиця 2.2 – Теоретичні характеристики моделей машинного навчання для виявлення фроду

Модель	*Інтерпре-тованість*	*Не-лінійність*	*Швидкість навчання*	*Стійкість до дисбалансу*	*Масштабо-ваність*	*Важливість ознак*
Логістична регресія	Висока	Низька	Дуже швидка	Погана	Висока	Так
Дерева рішень	Висока	Помірна	Швидка	Погана	Помірна	Так
Random Forest	Низька	Висока	Помірна	Помірна	Висока	Так
XGBoost	Низька	Дуже висока	Помірна	Добра	Висока	Так
LightGBM	Низька	Дуже висока	Швидка	Дуже добра	Дуже висока	Так
KNN (k=5)	Помірна	Висока	Швидка	Низька	Низька	Ні

Масштабованість показує придатність моделей до роботи з промисловими наборами даних, які можуть містити сотні мільйонів транзакцій. Метод KNN має низьку оцінку через високі обчислювальні витрати під час навчання та прогнозування, що робить їх малопридатними для систем рівня платіжних процесингових центрів.

Важливість ознак визначає, чи може модель формувати оцінки внеску окремих ознак у результат прогнозування, що необхідно для систем пояснюваності та регуляторного аудиту. Усі моделі градієнтного бустингу підтримують такі механізми через різні підходи оцінювання важливості ознак. Метод KNN подібної функціональності не має.

Загалом модель LightGBM демонструє найбільш збалансовані результати за всіма критеріями для задачі виявлення шахрайства: високу здатність до моделювання нелінійних залежностей, швидке навчання, добру масштабованість, ефективну роботу з незбалансованими даними та вбудовані механізми оцінки важливості ознак. Саме тому вона розглядається як основний кандидат для подальшої реалізації та експериментального дослідження.

2.3 Метрики оцінювання моделей машинного навчання для проблеми виявлення фродових транзакцій

Строге та відповідне оцінювання є центральним у дослідженнях виявлення шахрайства, особливо з огляду на серйозний дисбаланс класів, що робить стандартні метрики точності оманливими. Окрім метрик, розглянутих у формулах 2.1.5, 2.1.6, 2.1.7 та 2.1.8, важливу роль відіграють ROC-AUC та Precision-Recall AUC. ROC-AUC надає міру якості ранжування, інваріантну до порогу, що представлені у формулах 2.3.1 та 2.3.2.

$$AUC\text{–}ROC = \int_0^1 TPR(FPR^{-1}(t))\,dt = P(score(x^+) > score(x^-)), \quad (2.3.1)$$

де *AUC–ROC* — площа під ROC-кривою, варіюється від 0.5 до 1.0; ймовірнісна інтерпретація робить її інваріантною до масштабу та незалежною від порогу;

$TPR(FPR^{-1}(t))$ — частота справжніх позитивів, що відповідає конкретному рівню FPR t, відтворюючи ROC-криву при зміні t від 0 до 1;

$score(x^+)$ — оцінка шахрайства моделі для випадково взятої справді шахрайської транзакції x^+;

$score(x^-)$ — оцінка шахрайства моделі для випадкової легітимної транзакції x^-;

$P(...)$ — імовірність того, що модель правильно ранжує шахрайський приклад вище легітимного по всій тестовій вибірці.

AP більш інформативна, ніж ROC-AUC при дисбалансі класів, її базова лінія дорівнює апріорній імовірності класу представлена у формулі 2.3.2.

$$AP = \Sigma_n (R_n - R_{n-1}) \cdot P_n,, \quad (2.3.2)$$

де *AP* — середня прецизійність; еквівалентна площі під кривою Prec-Rec;

P_n — прецизійність при n-му пороговому значенні;

R_n — повнота при n-му пороговому значенні;

$R_n - R_{n-1}$ — приріст повноти на кроці порогу n, що використовується як вага і гарантує, що покращення прецизійності при операційних точках з великою повнотою вносять пропорційно більший внесок до метрики.

Висновки до розділу 2

Аналітичне дослідження існуючих моделей, сформувало цілісну основу для подальшої експериментальної реалізації системи виявлення фродових транзакцій на основі методів машинного навчання.

Виявлення шахрайства формалізується як задача бінарної класифікації з учителем для сильно незбалансованого набору даних. Для оцінювання якості моделей використовується матриця помилок та похідні метрики класифікації які разом забезпечують коректний аналіз в умовах дисбалансу класів.

Парадокс точності демонструє, що модель, яка класифікує всі транзакції як легітимні, може досягати надзвичайно високої точності, не виявляючи жодної шахрайської операції. Це підтверджує непридатність accuracy як основного критерію оцінювання у задачах фрод-детекшну. Метрика ROC-AUC, хоча і не залежить від порога класифікації, схильна переоцінювати якість моделей при сильному дисбалансі, оскільки велика кількість справжньо негативних результатів приховує практичний вплив хибнопозитивних прогнозів. Тому основним критерієм вибору моделей визначено PR-AUC (Average Precision. Додатково використовуються метрики Recall та F1-Score.

Теоретичне порівняння моделей, свідчать про те, що найбільш придатними моделями для задачі виявлення шахрайства є LightGBM та XGBoost. Вони поєднують високу здатність до моделювання складних нелінійних залежностей у PCA-перетворених ознаках транзакцій, вбудовані механізми компенсації дисбалансу через параметри scale_pos_weight та is_unbalance, надзвичайно швидке прогнозування, сумісне з часовими обмеженнями реальних платіжних систем, а також високу масштабованість для наборів даних із сотнями мільйонів транзакцій.

РОЗДІЛ 3 ЗАСТОСУВАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ПРАКТИЧНОГО ВИКОРИСТАННЯ В ВИРІШЕННЯ ПРОБЛЕМИ ВИЯВЛЕННЯ ФРОДОВИХ ТРАНЗАКЦІЙ

3.1 Опис датасету та формування підходу

Експерименти проводились на наборі даних ULB Credit Card Fraud Detection [28]. Набір містить 284 807 транзакцій, здійснених власниками карток із Європи у вересні 2013 р. З них 492 є фродовими, що становить 0,173% від усіх транзакцій — типовий сильний дисбаланс класів у реальному платіжному середовищі. Графік представлено на рисунку 3.1.

Набір ознак складається з 30 вхідних змінних та однієї цільової змінної — загалом 31 стовпець. Список ознак представлено у таблиці 3.1.

Таблиця 3.1 – Опис ознак набору даних ULB Credit Card Fraud Detection

Ознака	*Тип*	*Опис*	*Діапазон / Примітки*
Time	Неперервна	Кількість секунд від початку спостереження	0–172 792 с (≈ 48 год)
Amount	Неперервна	Вартість транзакції в євро	0,00–25 691,16 €; середнє 88,35 €, медіана 22,00 €
V1–V28	Неперервна	28 головних компонент PCA вихідних ознак транзакції; ідентифікатори приховано авторами набору з міркувань конфіденційності	Стандартизовано; центровано відносно нуля для легітимних транзакцій
Class	Бінарна	Цільова змінна: 0 = легітимна, 1 = фрод	284 315 легітимних / 492 шахрайські

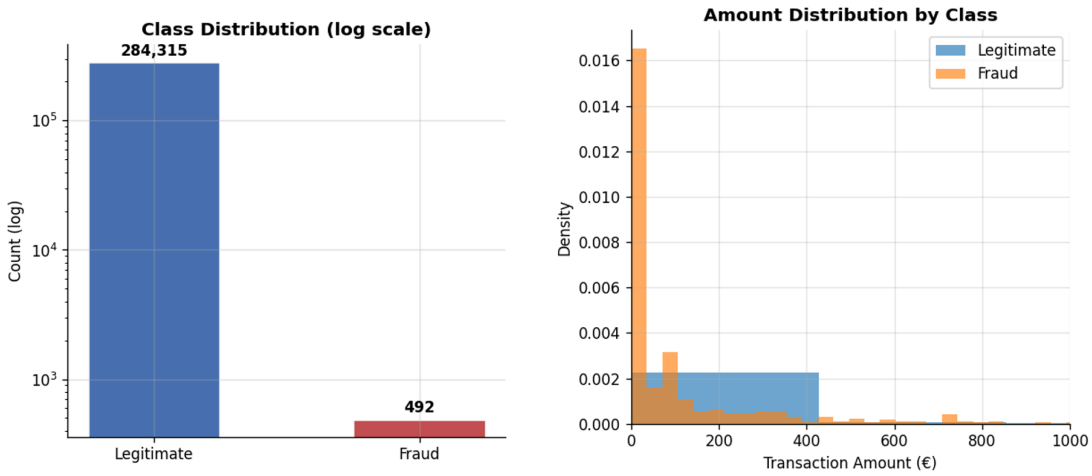

Рисунок 3.1 – Загальний огляд набору даних: розподіл класів, розподіл сум транзакцій та часові характеристики за 48-годинне вікно спостереження

Ознака Time охоплює рівно 48 годин спостереження (172 792 секунди). Обсяг транзакцій демонструє чіткий добовий патерн — два піки активності, розділені нічним спадом, — що помітно як для легітимних, так і для фродових транзакцій.

Ознака Amount демонструє сильно скошений розподіл. Фродові транзакції мають помітно нижчу медіанну суму (9,25 €) порівняно з легітимними (22,00 €), попри дещо вище середнє значення (122,21 € проти 88,29 €). Цей патерн свідчить про те, що значна частина фродових транзакцій є дрібними зондувальними списаннями, тоді як окремі великі операції завищують середнє.

28 V-ознак є головними компонентами, обчисленими авторами набору даних. Незважаючи на анонімізацію, ознаки зберігають виражений розпізнавальний сигнал. Різниця середніх значень суттєво варіюється між ознаками та 10 основних ознак представлено у таблиці 3.2.

V3, V14, V17, V12, V10 та V7 є сильно від'ємними для фроду відносно легітимних транзакцій, тоді як V4 та V11 — сильно додатними. Цей асиметричний патерн відображає ортогональне перетворення PCA, застосоване авторами набору.

Таблиця 3.2 – Топ-10 ознак за різницею середніх між фродовими та легітимними транзакціями

Ранг	*Ознака*	*Сер. (фрод.)*	*Сер. (легіт.)*	*|ΔСер.|*	*Напрямок*
1	V3	−7,033	+0,012	7,045	Фрод — сильно від'ємне
2	V14	−6,972	+0,012	6,984	Фрод — сильно від'ємне
3	V17	−6,666	+0,012	6,677	Фрод — сильно від'ємне
4	V12	−6,259	+0,011	6,270	Фрод — сильно від'ємне
5	V10	−5,677	+0,010	5,687	Фрод — сильно від'ємне
6	V7	−5,569	+0,010	5,578	Фрод — сильно від'ємне
7	V1	−4,772	+0,008	4,780	Фрод — сильно від'ємне
8	V4	+4,542	−0,008	4,550	Фрод — сильно додатне
9	V16	−4,140	+0,007	4,147	Фрод — сильно від'ємне
10	V11	+3,800	−0,007	3,807	Фрод — сильно додатне

Розподіл сум транзакцій у логарифмічній шкалі свідчить про те, що фродові транзакції концентруються в діапазоні менших сум — щільність фроду найвища в межах 1–100 €, тоді як легітимні транзакції демонструють ширший розподіл. Транзакції з найвищими сумами (>10 000 €) майже виключно легітимні, що вказує на рідкість фродових транзакцій з високою.

Розподіл у часі демонструє два чітких цикли активності, відповідних денно-нічним патернам за 48-годинне вікно спостереження. Фродові транзакції відбуваються протягом обох діб, але дещо підвищена щільність спостерігається у пізні нічні години (0–6 год) — патерн, характерний для автоматизованих шахрайських атак у період зниженого моніторингу. Дані візуалізовано на рисунку 3.2

Коефіцієнт дисбалансу класів приблизно 577,9 : 1 (легітимні : фродові) робить стандартну точність ненадійним показником ефективності — класифікатор, що завжди прогнозує «легітимна», досягне точності 99,83%, не виявивши жодного шахрайства. Тому в усіх експериментах використовуються

метрики з акцентом на повноту: F2-оцінка (яка надає повноті вдвічі більшу вагу порівняно з точністю), F1-оцінка, площа під кривою Precision-Recall (*Average Precision*) та площа під ROC-кривою (*ROC-AUC*). Також відстежується частка хибнопозитивних спрацювань (FPR), оскільки надмірна кількість помилкових тривог ускладнює обслуговування легітимних клієнтів.

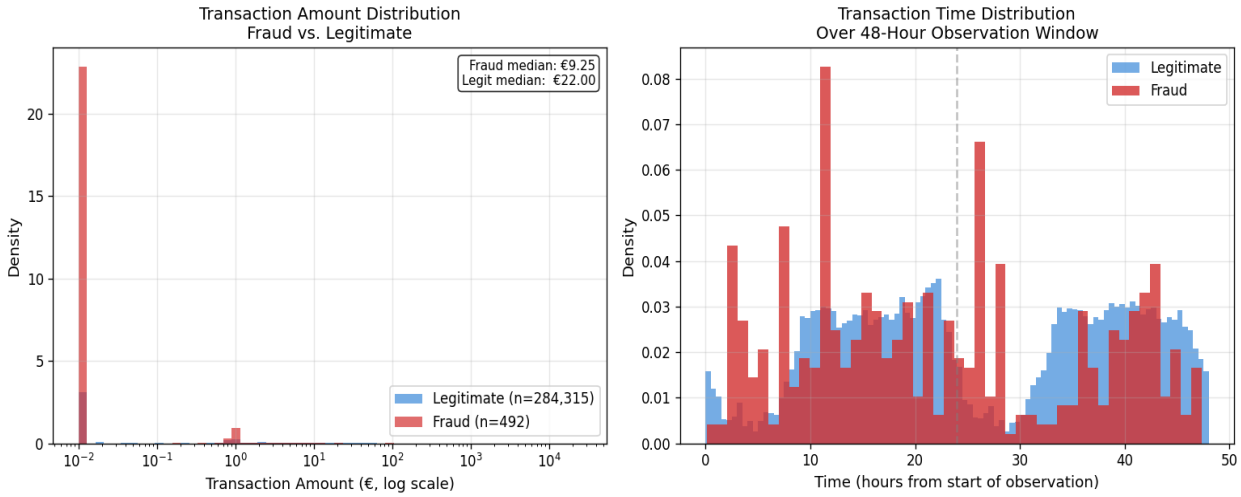

Рисунок 3.2 – Розподіл суми транзакцій (у логарифмічній шкалі) та часових характеристик для фродових та легітимних транзакцій

3.2 Попередня обробка даних

Набір даних було розбито на три непересічні підмножини за допомогою стратифікованого відбору зі збереженням вихідної частки шахрайства 0,173% у кожній частині та представлено у таблиці 3.3.

Усі кроки попередньої обробки — масштабування ознак, перевибірка методом SMOTE та навчання автоенкодера — виконувались виключно на тренувальній вибірці й застосовувались без повторного підбору до валідаційної та тестової. Цей підхід гарантує, що результати тесту відображають узагальнювальну здатність на нових даних.

Числові ознаки мають значно різні масштаби: *Amount* варіюється від 0 до 25 691 €, тоді як ознаки PCA вже стандартизовано. Всі ознаки нормалізовано за допомогою *StandardScaler* (нульове середнє, одинична дисперсія), підбір якого виконувався лише на тренувальній вибірці. Це масштабування є необхідним для моделей як логістична регресія та KNN і покращує збіжність градієнта у методах бустингу на деревах.

Таблиця 3.3 – Поділ набору даних

Підмножина	*Кількість зразків*	*Фродові*	*Частка фроду*
Тренувальна	199 364 (70%)	344	0,173%
Валідаційна	42 721 (15%)	74	0,173%
Тестова	42 722 (15%)	74	0,173%

Для вирішення проблеми дисбалансу класів у різних експериментах застосовувались різні стратегії для покращення отриманих результатів.

1. Коригування ваг класів використовувалось для логістичної регресії, Random Forest та Decision Tree. Моделі навчались із параметром class_weight='balanced', який автоматично підвищує вагу меншини у функції втрат на коефіцієнт, обернено пропорційний частоті класу.
2. Навчання з урахуванням вартості помилок використовувалось для XGBoost, безпосередньо масштабуючи градієнт для фродових зразків.
3. Перевибірка SMOTE для KNN метод SMOTE (Synthetic Minority Over-sampling Technique) підвищив частку шахрайства з 0,173% до приблизно 9,09% (218 922 навчальних зразки після перевибірки), генеруючи синтетичні шахрайські зразки шляхом інтерполяції між наявними, зміни візуалізовано на рисунку 3.3.
4. SMOTE + Tomek Links для LightGBM очищав неоднозначні регіони та зменшував хибнопозитивні спрацювання без втрати повноти.

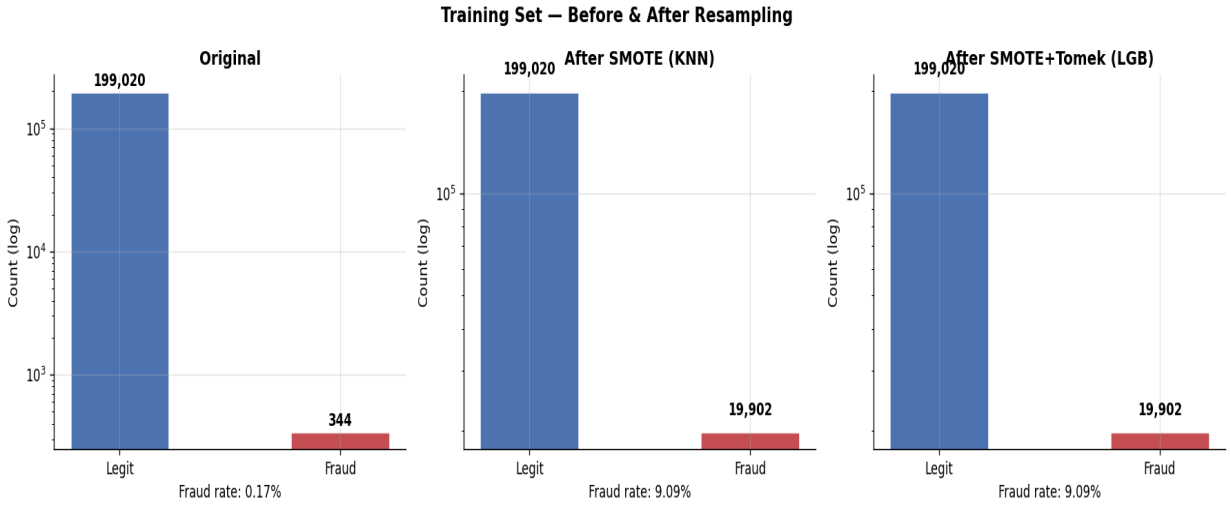

Рисунок 3.3 – Вплив перевибірки SMOTE на розподіл класів у тренувальній вибірці

Також важливим було оптимізувати поріг прийняття рішень, оскільки стандартний поріг прийняття рішення 0,5 є неприйнятним для сильно незбалансованих даних. Для вибору порогу була використана максимізація F2. Перебір 150 порогів від 0,01 до 0,99 виконувався на валідаційній вибірці; поріг, що максимізує F2-оцінку, застосовувався до тестової вибірки.

3.3 Моделі машинного навчання, нейронні мережі та їх навчання

Шість моделей та 2 нейронні мережі з різними архітектурами навчено та оцінено на однаковому стратифікованому поділі набору даних. Кожна модель використовує спеціальну стратегію обробки дисбалансу та поріг, відкалібрований на валідаційній вибірці для максимізації F2. У дослідженні порівнювались методи чотирьох груп. А саме лінійні моделі (логістична регресія як базова інтерпретована модель), деревоподібні методи (дерево рішень та Random Forest), градієнтний бустинг (XGBoost та LightGBM) і нейронні мережі (багатошаровий перцептрон та TabNet).

3.3.1 Логістична регресія

Логістична регресія є лінійною базовою моделлю дослідження. Вона оцінює ймовірність фродовості транзакції як зважену суму вхідних ознак, пропущену через сигмоїдну функцію. Позитивна вага ознаки означає, що більші значення зміщують прогноз до фроду; від'ємна вага — до легітимної. Незважаючи на свою простоту, модель показує хороші результати, оскільки кілька PCA-похідних V-ознак майже лінійно розділяють два класи після стандартизації.

Основна проблема — дисбаланс класів 577 : 1. Без коригування модель навчиться завжди прогнозувати «легітимна» — досягаючи 99,8% точності при нульовому виявленні фроду. Параметр class_weight=balanced виправляє це, збільшуючи штраф за пропущені фродові випадки пропорційно до відношення дисбалансу.

Найбільш впливовим кроком є калібрування порогу. При стандартному t=0,50 модель генерує 903 хибних тривоги — абсолютно непридатно для виробничого середовища. Підвищення порогу до t*=0,99 скорочує хибні тривоги на 97%, зберігаючи повноту вище 0,81. Конфігурацію моделі логістичної регресії наведено у таблиці 3.4.

Криві точності, повноти та F2 залежно від порогу рішення, а також матриця помилок при t*=0,99 представлено на рисунку 3.4. Різке зростання точності вище t=0,95 відповідає стрімкому скороченню хибнопозитивних результатів.

Результати логістичної регресії на тестовій вибірці наведено у таблиці 3.5. Налаштування порогу скорочує хибнопозитивні результати з 903 до 29.

Ключовий висновок: лише калібрування порогу перетворює операційно непридатну модель (F2=0,257) на конкурентоспроможну базову (F2=0,779). ROC-AUC залишається незмінним, оскільки є незалежним від порогу —

ранжувальна здатність моделі є сильною, необхідним було лише коригування робочої точки.

Таблиця 3.4 – Логістична регресія, конфігурація моделі

Параметр	*Значення*	*Призначення*
Солвер	L-BFGS	Ефективний квазіньютонівський оптимізатор для невеликих і середніх наборів даних
Регуляризація	L2 (C=1,0)	Запобігає перенавчанню шляхом штрафування великих значень ваг
Вага класів	balanced	Збільшує вагу фродових зразків у ~578 разів для нейтралізації дисбалансу
Макс. ітерацій	1 000	Достатньо для збіжності на цьому просторі ознак
Поріг рішення	t*=0,99	Налаштовано на валідаційній вибірці для максимізації F2-міри

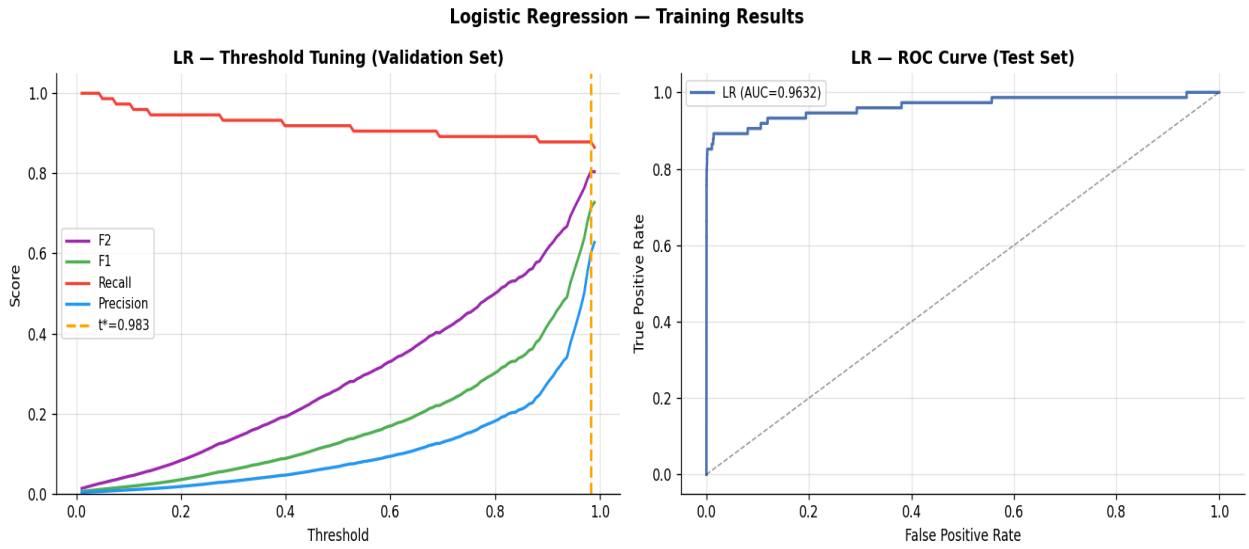

Рисунок 3.4 – Логістична регресія: точність, повнота та F2 залежно від порогу рішення (ліворуч) та ROC крива

Таблиця 3.5 – Логістична регресія – результати на тестовій вибірці

Налаштування	*t**	*TP*	*FP*	*FN*	*Precision*	*Recall*	*F2*	*ROC-AUC*
Стандартний (t=0,50)	0,50	65	903	9	0,067	0,878	0,257	0,968
Налаштований (t*=0,99)	0,99	60	29	14	0,674	0,811	0,779	0,968

3.3.2 Дерево рішень

Дерево рішень будує блок-схему запитань «так/ні» щодо ознак, рекурсивно розділяючи дані у кожному вузлі для якнайкращого розрізнення фродових і легітимних транзакцій. Кожне рішення можна простежити та перевірити — що є особливо цінним у регульованих середовищах, де людина-аудитор повинна мати можливість пояснити, чому транзакцію було позначено.

Глибина дерева визначає складність правил. Мілке дерево пропускає тонкі патерни; глибоке — запам'ятовує навчальні дані та погано узагальнює. Пошук по сітці серед глибин {3, 5, 7, 10, 15} визначив оптимальною глибину=5 — точку, в якій F2-міра на валідаційній вибірці досягає піку перед початком домінування перенавчання.

Аналіз важливості ознак виявляє разючу концентрацію: V14 самостійно забезпечує 83,1% усього скорочення нечистоти за Джині. Ця єдина PCA-компонента несе основний фрод-сигнал, а V4 (5,3%) та V12 (3,6%) є далекими вторинними внесками. Конфігурацію дерева рішень наведено у таблиці 3.6.

Таблиця 3.6 – Дерево рішень, конфігурація моделі

Параметр	*Значення*	*Призначення*
Алгоритм	CART (критерій Джині)	Стандартне дерево класифікації — мінімізує змішування класів у кожному розбитті
Макс. глибина	5 (пошук по сітці)	Запобігає перенавчанню; більші глибини демонстрували зниження F2 на валідаційній вибірці
Мін. зразків у листі	5	Кожне правило листа застосовується щонайменше до 5 навчальних прикладів
Вага класів	balanced	Компенсує дисбаланс 577:1 у критерії розбиття
Поріг рішення	t*=0,99	Дуже високий поріг необхідний — дерево надмірно призначає ймовірність фроду при глибині=5

Залежність F2 та повноти від глибини дерева, а також перебір порогів при оптимальній глибині проілюстровано на рисунку 3.5.

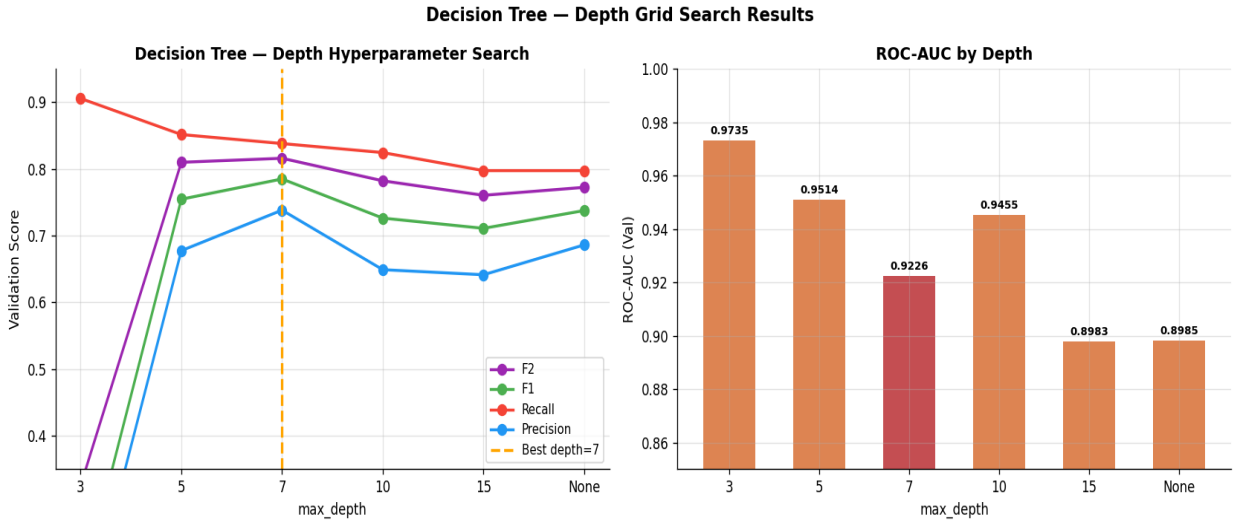

Рисунок 3.5 – Ліворуч зображено графік F2 та повнота на валідаційній вибірці залежно від глибини дерева. Праворуч зображено перебір порогів при найкращій глибині

Залежності точності, повноти та F2 від порогу на валідаційній вибірці представлено на рисунку 3.6.

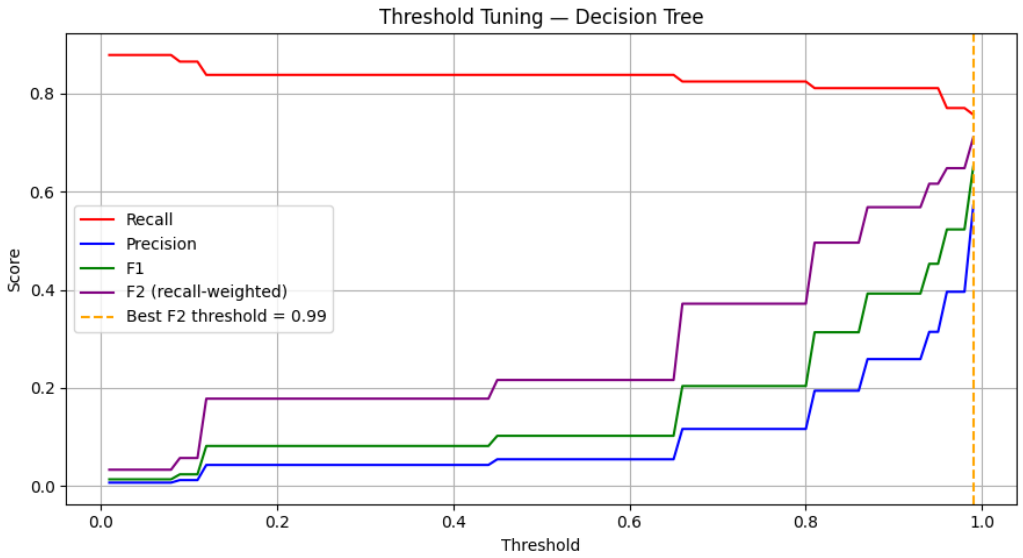

Рисунок 3.6 – Дерево рішень: точність, повнота та F2 залежно від порогу

Криву Точність–Повнота та ROC-криву на валідаційній вибірці представлено на рисунку 3.7. Точність–Повнота (ліворуч) та ROC-крива (праворуч) на валідаційній вибірці.

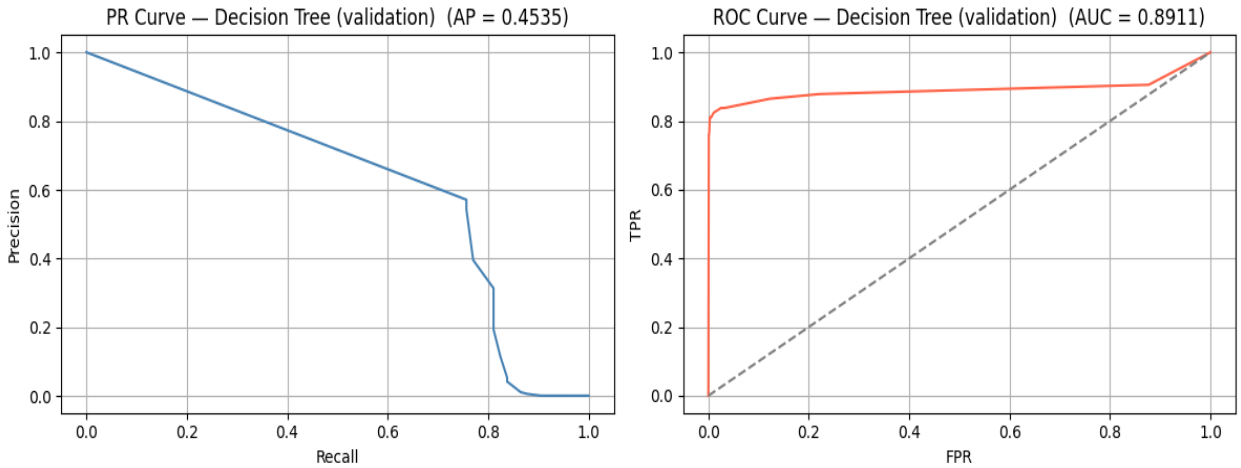

Рисунок 3.7 – Дерево рішень: крива Точність–Повнота (ліворуч) та ROC-крива (праворуч) на валідаційній вибірці

Важливість ознак за скороченням нечистоти за Джині наведено на рисунку 3.8. V14 домінує з 83,1% від загальної важливості — дерево фактично навчає одне основне правило розбиття.

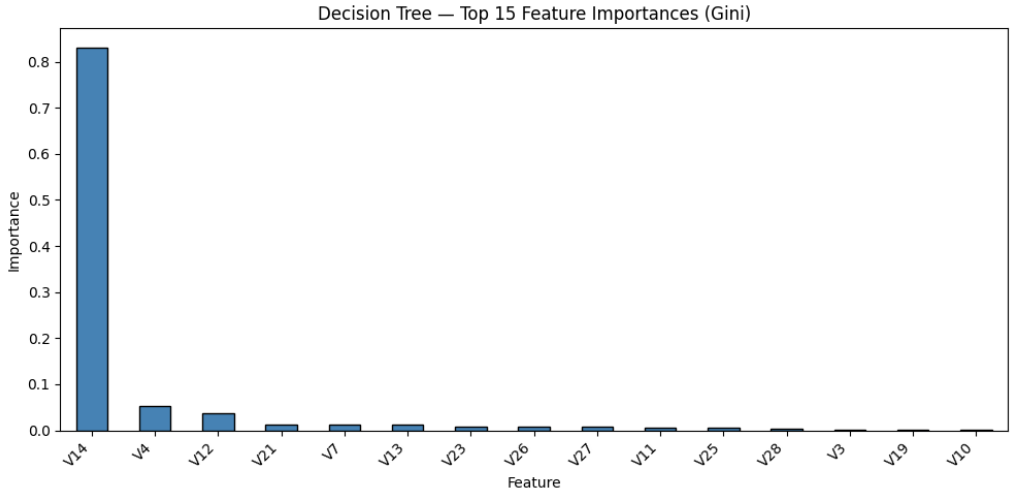

Рисунок 3.8 – Важливість ознак за Джині

Перші 3 рівні дерева глибини=5 представлено на рисунку 3.9. Кореневе розбиття за V14 ≤ −4,56 негайно відокремлює переважну більшість фродових випадків, що узгоджується з аналізом важливості ознак.

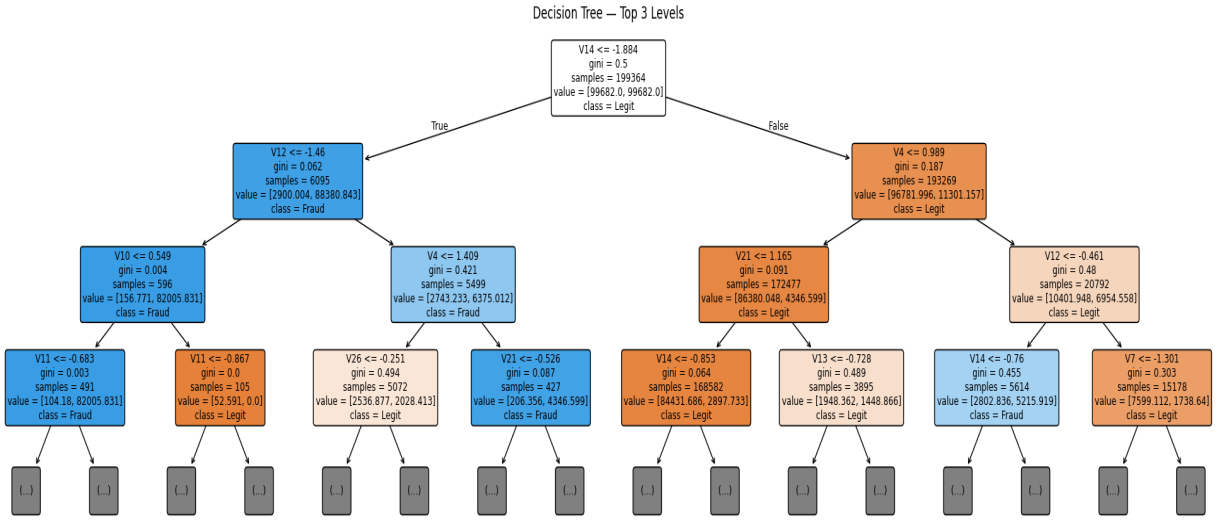

Рисунок 3.9 – Перші 3 рівні дерева глибини=5

Матриці помилок дерева рішень на тестовій вибірці при t=0,50 та t*=0,99 представлено на рисунку 3.10. Різке скорочення хибнопозитивних результатів з 1 109 до 36 є основною перевагою налаштування порогу.

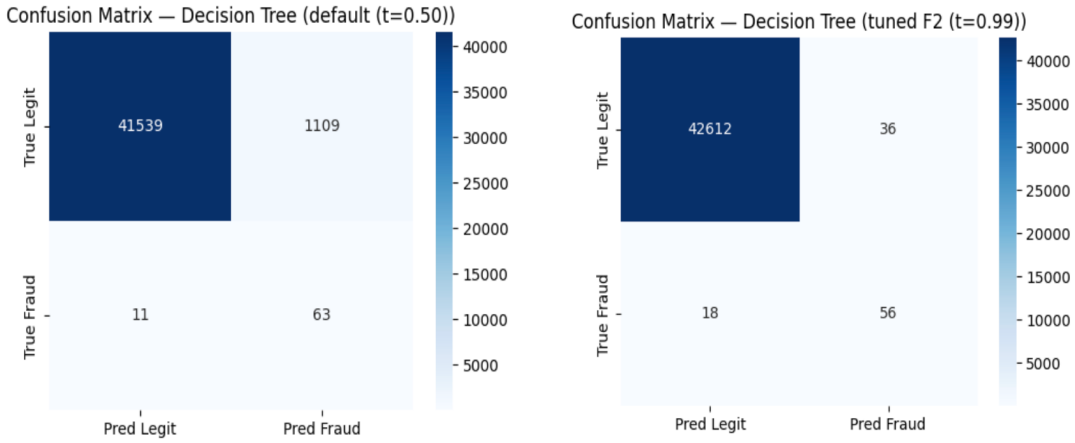

Рисунок 3.10 – Матриці помилок дерева рішень на тестовій вибірці при t=0,50 (ліворуч) та t*=0,99 (праворуч)

Результати дерева рішень на тестовій вибірці наведено у таблиці 3.7. Ключовий висновок: дерево рішень є найслабшим класифікатором за F2 (0,722), але його 5-рівнева структура з 28 листами може бути виведена та зрозуміла. Жодна інша модель у цьому дослідженні не пропонує такої властивості.

Таблиця 3.7 – Дерево рішень, результати на тестовій вибірці

Налаштування	*t**	*TP*	*FP*	*FN*	*Precision*	*Recall*	*F2*	*ROC-AUC*
Стандартний (t=0,50)	0,50	56	1109	18	0,048	0,757	0,215	0,891
Налаштований (t*=0,99)	0,99	56	36	18	0,609	0,757	0,722	0,891

3.3.3 Random Forest

Random Forest розширює одиночне дерево рішень, навчаючи 200 дерев, кожне на різній випадковій бутстреп-вибірці навчальних даних. Кінцева ймовірність фроду є середнім значенням по всіх 200 деревах. Ця стратегія — що називається бегінгом (bagging) — різко зменшує дисперсію: окремі дерева можуть перенавчатись на своїй конкретній бутстреп-вибірці, але їхні помилки значною мірою некорельовані та усереднюються.

Важливим спостереженням є дуже низький оптимальний поріг: t*=0,17, порівняно з t*=0,99 для дерева рішень. Це відображає, як ансамблеве усереднення зміщує розподіл ймовірностей. Окремі дерева, навчені на незбалансованих даних, призначають помірні ймовірності фроду прикордонним транзакціям. Коли 200 таких виходів усереднюються, результуючі ймовірності стискаються до середини діапазону [0,1]. Тому необхідний нижчий поріг для відновлення повноти — а ансамблеве усереднення є тим, що утримує точність на рівні навіть при такому дозволяючому порозі. Конфігурацію випадкового лісу наведено у таблиці 3.8. Перебір порогів та ROC-крива при t*=0,17 представлено на рисунку 3.11. Широка рівна область високого F2 між t=0,10 та t=0,25 демонструє стійкість до вибору порогу — невеликі зміни t не впливають суттєво на продуктивність.

Результати випадкового лісу на тестовій вибірці наведено у таблиці 3.9. Модель за F2 на тестовій вибірці показує 0,847 лише з 8 хибними тривогами на 42 648 легітимних транзакцій — частка хибнопозитивних 0,019%. Поєднання ансамблевого усереднення, збалансованих ваг класів та налаштування порогу дає модель, що є водночас надзвичайно чутливою та операційно практичною.

Таблиця 3.8 – Random Forest, конфігурація моделі

Параметр	*Значення*	*Призначення*
Кількість дерев	200	Достатньо для стабільних оцінок ймовірностей; ефект від збільшення кількості незначний
Вибір ознак	sqrt(n_ознак) / розбиття	Декорелює окремі дерева — кожне бачить випадкову підмножину ознак
Макс. глибина	Необмежена	Окремим деревам дозволено повне перенавчання; ансамблеве усереднення виправляє це
Вага класів	balanced	Кожне дерево застосовує збалансовані ваги класів до своєї бутстреп-вибірки
Поріг рішення	t*=0,17	Низький поріг необхідний через стиснення ймовірностей після усереднення

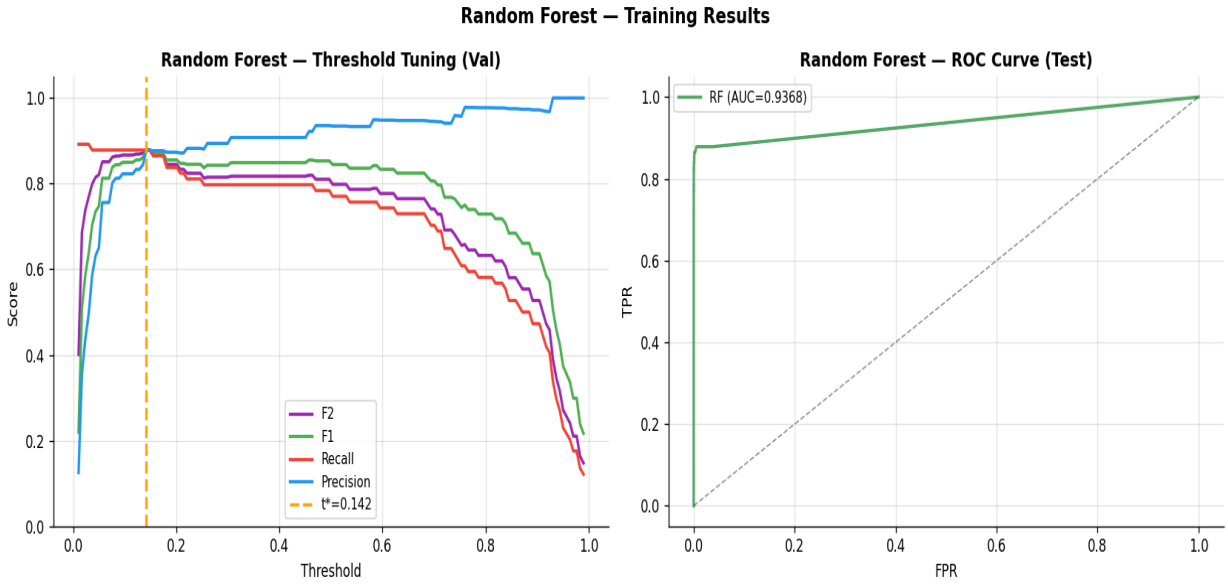

Рисунок 3.11 – Random Forest: перебір порогів (ліворуч) ROC-крива (праворуч)

Таблиця 3.9 – Random Forest, результати на тестовій вибірці. Найкращий F2 на тестовій вибірці з найменшою кількістю хибних тривог

Налаштування	*t**	*TP*	*FP*	*FN*	*Precision*	*Recall*	*F2*	*ROC-AUC*
Стандартний (t=0,50)	0,50	62	34	12	0,646	0,838	0,755	0,937
Налаштований (t*=0,17)	0,17	62	8	12	0,886	0,838	0,847	0,937

3.3.4 XGBoost

XGBoost (Extreme Gradient Boosting) будує дерева послідовно: кожне нове дерево зосереджується на виправленні помилок прогнозування поточного ансамблю. Тоді як Random Forest навчає всі дерева незалежно та усереднює їх, XGBoost навчає кожне дерево на залишкових помилках своїх попередників — процес, що називається бустингом. Це робить його особливо ефективним для захоплення складних нелінійних взаємодій, які попередні дерева не виявили.

Було протестовано дві конфігурації. Базова використовує оригінальні 30 ознак із навчанням з урахуванням вартості (пропущений фрод-випадок важливіший за хибну тривогу). Розширена конфігурація доповнює набір ознак двома невідстеженими аномалійними сигналами (помилка реконструкції автоенкодера та оцінка Isolation Forest [21]) і застосовує ресемплінг SMOTE+Tomek. Порівняльну конфігурацію XGBoost наведено у таблиці 3.10.

Перебір порогів та та log loss при t*=0,29 для базової конфігурації XGBoost наведено на рисунку 3.12. Точність і повнота добре збалансовані у широкому діапазоні порогів, що свідчить про добре відкалібровані виходи ймовірностей.

Таблиця 3.10 – XGBoost, порівняння конфігурацій (базова та розширена)

Параметр	*Базова*	*SMOTE+AE+IF*
Ознаки	30 (оригінальні)	32 (+ помилка AE + оцінка IF)
Обробка дисбалансу	scale_pos_weight=578	Ресемплінг SMOTE+Tomek
Кількість дерев (n_estimators)	200	200
Макс. глибина	5	5
Темп навчання	0,1	0,1
Subsample	0,8	0,8
colsample_bytree	0,8	0,8
Поріг рішення	t*=0,29	t*=0,39

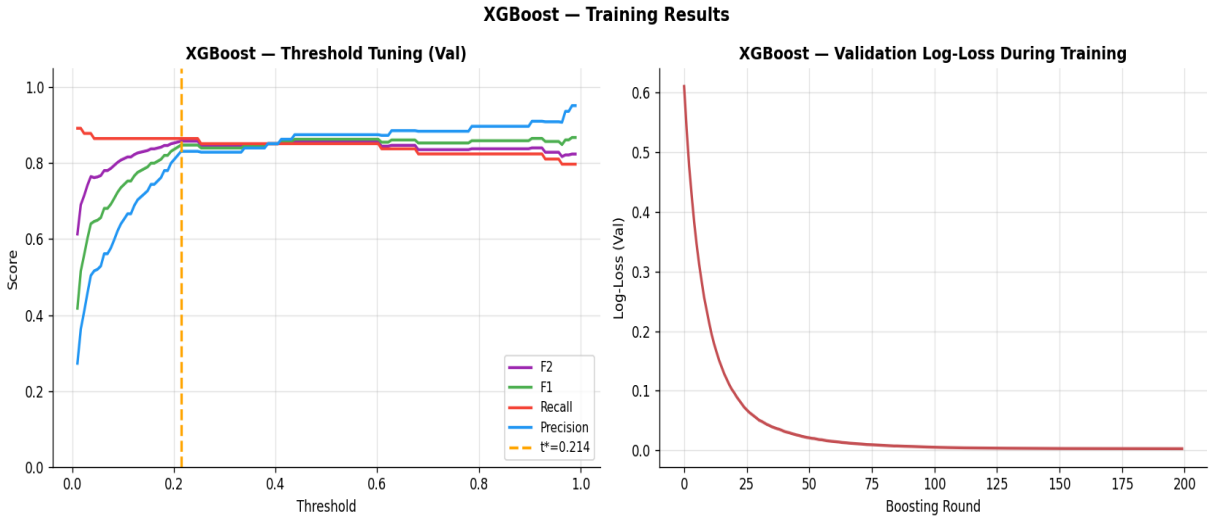

Рисунок 3.12 – XGBoost (базова): перебір порогів (ліворуч) та log loss при t*=0,29 (праворуч)

Результати XGBoost на тестовій вибірці наведено у таблиці 3.11. Розширена конфігурація досягає найвищого ROC-AUC (0,984) серед усіх моделей.

Таблиця 3.11 – XGBoost, результати на тестовій вибірці

Конфігурація	*t**	*TP*	*FP*	*FN*	*Precision*	*Recall*	*F2*	*ROC-AUC*
Базова	0,29	61	11	13	0,847	0,824	0,829	0,966
SMOTE+AE+IF	0,39	—	—	—	0,768	0,851	0,833	0,984

XGBoost досягає сильної точності 0,847 і є найбільш збалансованою одиночною моделлю за балансом точності–повноти. Додавання розроблених аномалійних ознак підвищує ROC-AUC з 0,966 до 0,984.

3.3.5 LightGBM

LightGBM — це фреймворк градієнтного бустингу, який відрізняється від XGBoost двома ключовими аспектами: дерева зростають поаркушево (завжди розбиваючи аркуш із найбільшим скороченням втрат, незалежно від рівня глибини), а не порівневово, та використовується Gradient-based One-Side Sampling (GOSS) для пропуску навчальних прикладів із низьким градієнтом. Ці конструктивні рішення роблять його швидшим на великих наборах даних і здатним навчати глибші, більш спеціалізовані правила рішень.

Модель навчена на 32 ознаках: 30 оригінальних масштабованих ознак плюс два несупервізовані аномалійні сигнали, розроблені виключно на навчальних даних. Автоенкодер — це нейронна мережа, навчена виключно на легітимних транзакціях. Вона навчає стиснуте представлення нормальної поведінки. Фродові транзакції, що не відповідають тим самим патернам, виробляють набагато вищі помилки реконструкції — середній розрив MSE становить 87 разів (20,02 для фродових проти 0,23 для легітимних). Isolation Forest незалежно ідентифікує аномалії, вимірюючи, наскільки легко кожну транзакцію можна ізолювати через випадкові розбиття.

Гіперпараметри знайдено автоматично за допомогою Optuna — фреймворку байєсівської оптимізації, що інтелектуально досліджує простір конфігурацій протягом 50 спроб [22]. Замість перебору всіх комбінацій (пошук по сітці) або випадкових вибірок, Optuna навчається, які діапазони гіперпараметрів дають хороші результати. Гіперпараметри, знайдені Optuna, порівняно зі стандартними значеннями, наведено у таблиці 3.12.

Результати LightGBM на тестовій вибірці наведено у таблиці 3.13. Стандартний поріг дає 236 хибних тривог, а налаштування скорочує їх до 19.

Таблиця 3.12 – LightGBM, гіперпараметри, знайдені Optuna, порівняно зі стандартними

Гіперпараметр	*Значення Optuna*	*Стандартне / Типове*
n_estimators	474	100
num_leaves	114	31
max_depth	9	-1 (необмежена)
learning_rate (темп навчання)	0,123	0,1
subsample	0,705	1,0
colsample_bytree	0,907	1,0
reg_alpha (L1)	0,543	0,0
reg_lambda (L2)	0,0004	0,0

Таблиця 3.13 – LightGBM, результати на тестовій вибірці

Налаштування	*t**	*TP*	*FP*	*FN*	*Precision*	*Recall*	*F2*	*ROC-AUC*
Стандартний (t=0,50)	0,50	62	236	12	0,208	0,838	0,509	0,974
Налаштований (t*=0,457)	0,457	62	19	12	0,765	0,838	0,822	0,974

Перебір порогів та ROC крива для LightGBM (налаштованого Optuna) представлено на рисунку 3.13.

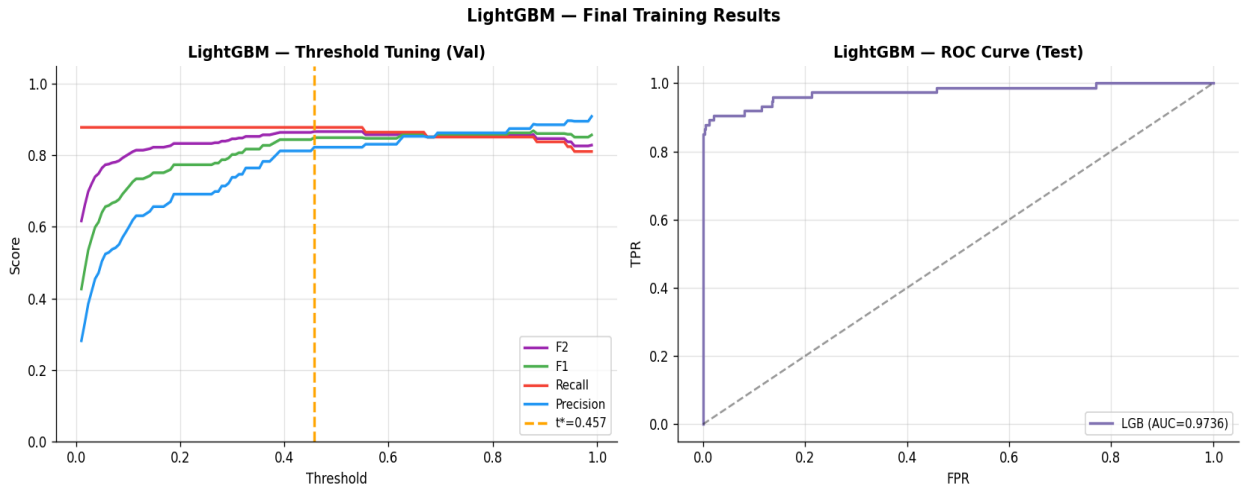

Рисунок 3.13 – Зображення LightGBM, перебір порогів та ROC крива

3.3.6 Метод k-найближчих сусідів

Метод k-найближчих сусідів (KNN) фундаментально відрізняється від усіх інших моделей у цьому дослідженні — він не навчає жодних параметрів. Натомість весь навчальний набір зберігається, і під час прогнозування кожна нова транзакція порівнюється з усіма збереженими прикладами. Ймовірність фроду — це просто частка з k найближчих навчальних прикладів, що були фродовими.

Ця простота має два важливих наслідки. По-перше, KNN не має вбудованого механізму обробки дисбалансу класів. Якщо сусідство домінованe легітимними транзакціями у співвідношенні 577:1, кожен прогноз буде «легітимна». Ресемплінг SMOTE вирішує цю проблему, генеруючи синтетичні фродові приклади до тих пір, поки фрод не становитиме 9,09% навчального набору — надаючи кожному сусідству реалістичний шанс

містити фродових сусідів. По-друге, оскільки кожен прогноз вимагає пошуку серед 218 922 збережених навчальних прикладів, виведення є обчислювально витратним.

Центральним конструктивним вибором є k — кількість сусідів, що беруть участь у голосуванні. Малі k виробляють рваний, зашумлений кордон рішень, що перенавчається до окремих навчальних точок. Більші k згладжують кордон, підвищуючи точність, але потенційно втрачаючи повноту. Аналіз чутливості по 11 значеннях від k=1 до k=101 точно відображає цей компроміс. Конфігурацію KNN наведено у таблиці 3.14.

Таблиця 3.14 – KNN, конфігурація моделі та конструктивні вибори

Параметр	*Значення*	*Примітки*
Метрика відстані	Евклідова	Стандартна L2-відстань у стандартизованому просторі ознак
Індексна структура	Ball tree	Ефективний точний пошук найближчих сусідів для даної розмірності
Протестовані k	1, 3, 5, 7, 10, 15, 21, 31, 51, 75, 101	11 значень, що охоплюють широкий діапазон
Вибране k	21 (test F2=0,813)	—
Розмір навчальної вибірки	218 922 зразки	Після SMOTE: 9,09% фроду (порівняно з 0,173%)
Поріг рішення	t*=0,957 (k=21), t*=0,937 (k=15)	Високий поріг необхідний — SMOTE зміщує розподіл ймовірностей угору

Чутливість KNN до k на валідаційній вибірці (SMOTE навчання, перебір порогів) наведено у таблиці 3.15.

Таблиця 3.15 – KNN: чутливість до k на валідаційній вибірці

k	*F2*	*Precision*	*Recall*	*ROC-AUC*
1	0,810	0,677	0,851	0,925
3	0,818	0,708	0,865	0,932
5	0,818	0,747	0,878	0,939
7	0,827	0,785	0,878	0,939
10	0,840	0,849	0,878	0,939
15	0,849	0,899	0,892	0,945
21	**0,840**	**0,910**	**0,892**	**0,945**
31	0,833	0,938	0,892	0,945
51	0,826	0,967	0,905	0,951
75	0,833	0,816	0,905	0,951
101	0,836	0,827	0,919	0,957

Аналіз чутливості KNN по всіх 11 значеннях k проілюстровано на рисунку 3.14.

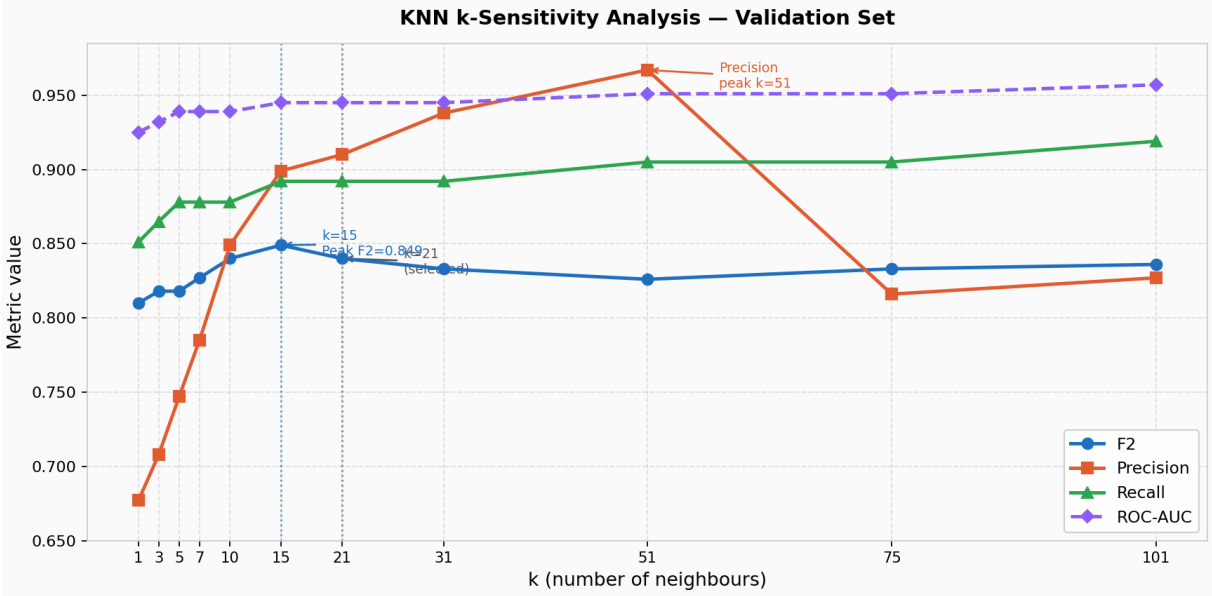

Рисунок 3.14 – Аналіз чутливості KNN по всіх 11 значеннях k

Матриці помилок KNN k=21 та k=15 на тестовій вибірці при стандартному та F2-оптимальному порогах представлено на рисунках 3.15 та 3.16. При t=0,50 модель k=21 виявляє 63 фрод-випадки, але генерує 128 хибних тривог. Налаштування порогу зберігає 59 виявлень фроду, скорочуючи хибні тривоги до 8. Для KNN k=15 такий ж патерн, що й у k=21, але з 11 хибними тривогами при проти 8 для k=21 — що надає k=21 перевагу.

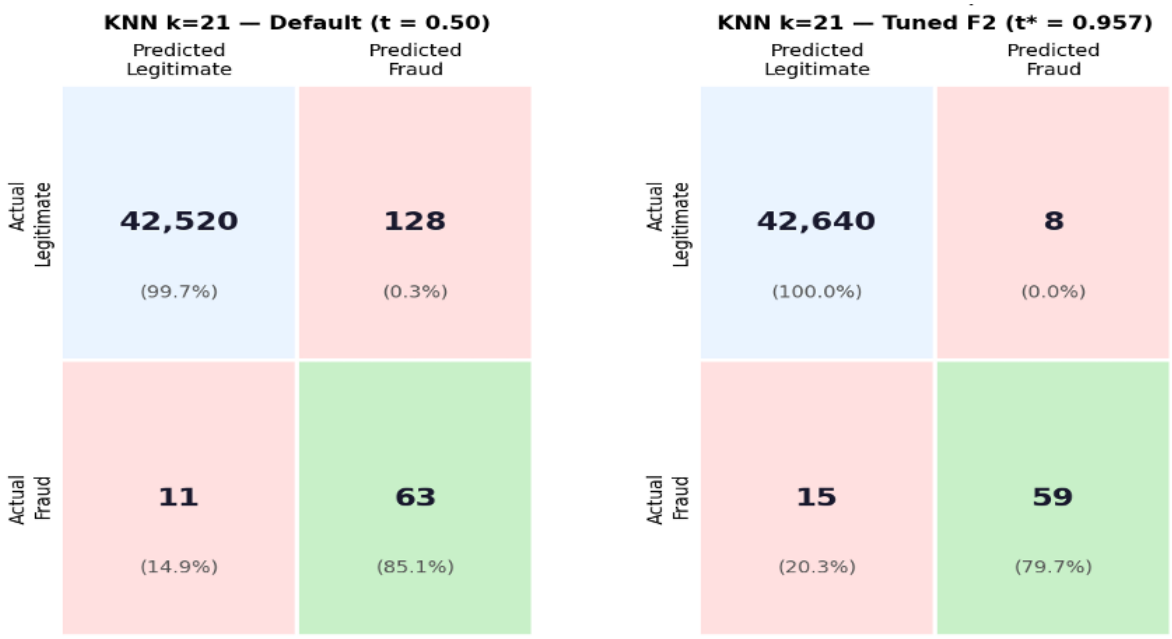

Рисунок 3.15 – Матриці помилок KNN k=21 на тестовій вибірці

Обидва виявляють рівно 59 із 74 фрод-випадків. Різниця повністю в точності: k=21 виробляє на 3 хибні тривоги менше та має точність вищу на 0,038. Пряме порівняння k=15 та k=21 на тестовій вибірці наведено на рисунку 3.17.

Скориговані результати KNN на тестовій вибірці (точність і повнота при одному F2-оптимальному порозі) наведено у таблиці 3.16.

KNN (k=21) досягає конкурентоспроможного F2=0,813 лише з 8 хибними тривогами — збігаючись із випадковим лісом за кількістю хибних

тривог. Основним обмеженням є відсутність важливості ознак, модель не може пояснити, які ознаки визначили рішення, що є регуляторною проблемою.

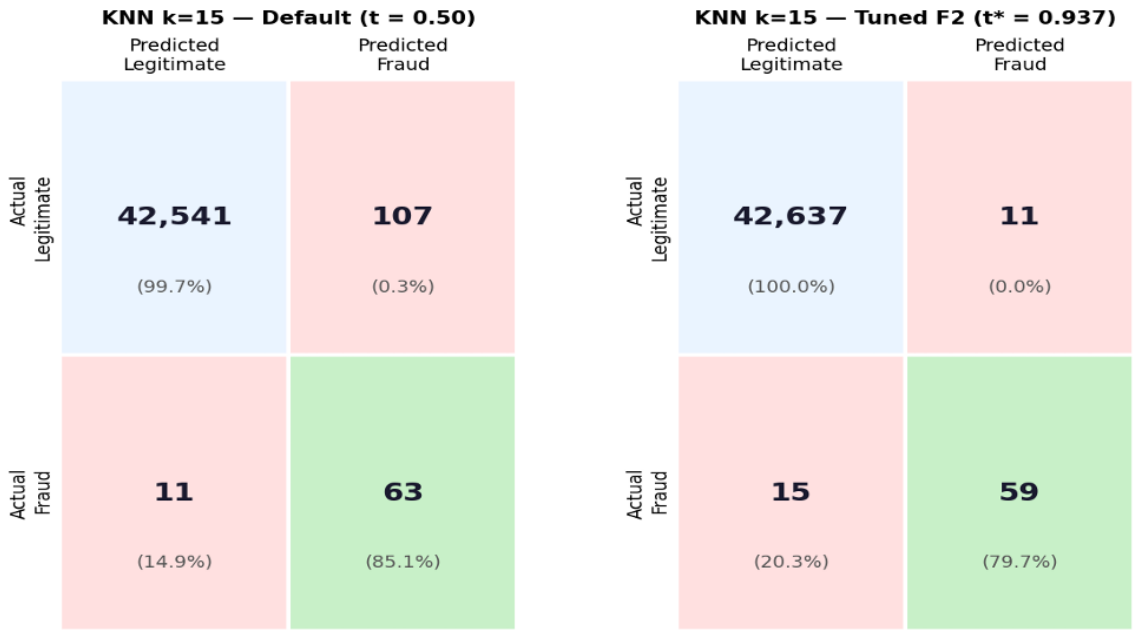

Рисунок 3.16 – Матриці помилок KNN k=15 на тестовій вибірці

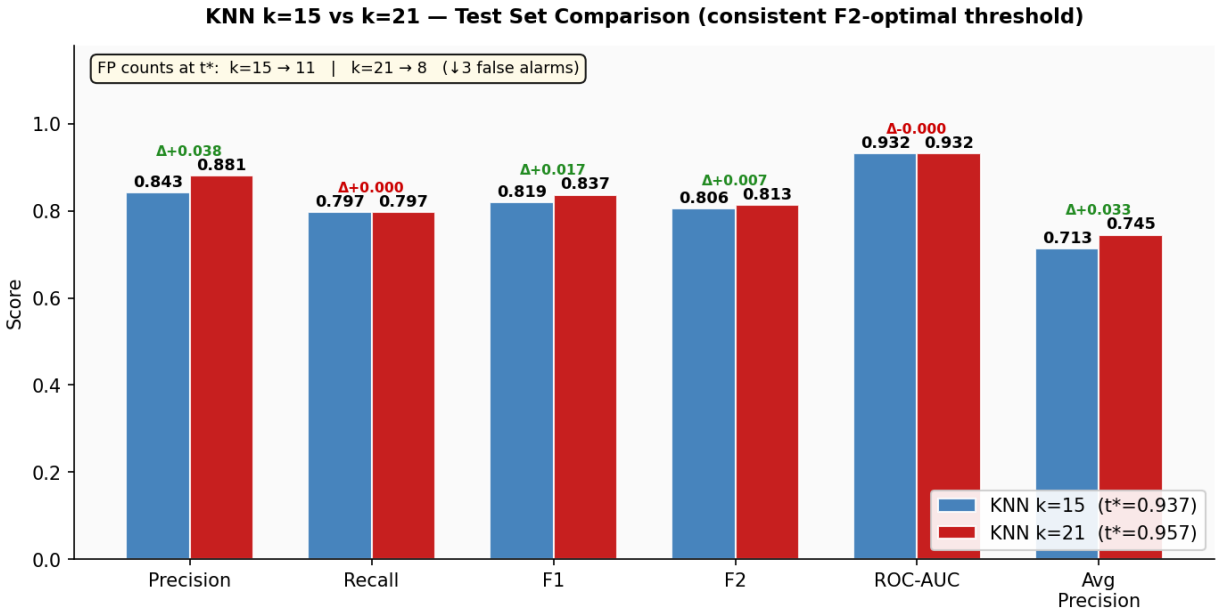

Рисунок 3.17 – Пряме порівняння k=15 та k=21 на тестовій вибірці

Таблиця 3.16 – KNN, скориговані результати на тестовій вибірці

Конфігурація	*t**	*TP*	*FP*	*FN*	*Precision*	*Recall*	*F2*	*ROC-AUC*
KNN k=15	0,937	59	11	15	0,843	0,797	0,806	0,932
KNN k=21	0,957	59	8	15	0,881	0,797	0,813	0,932

3.3.7 Багатошаровий персептрон та TabNet

Процес розробки MLP не обмежувався одноразовим навчанням та оцінюванням; модель ітеративно перепроєктовувалась у трьох поколіннях, кожне з яких усувало недоліки, виявлені в попередній версії. Загальний прогрес за трьома версіями представлено на рисунку 3.18.

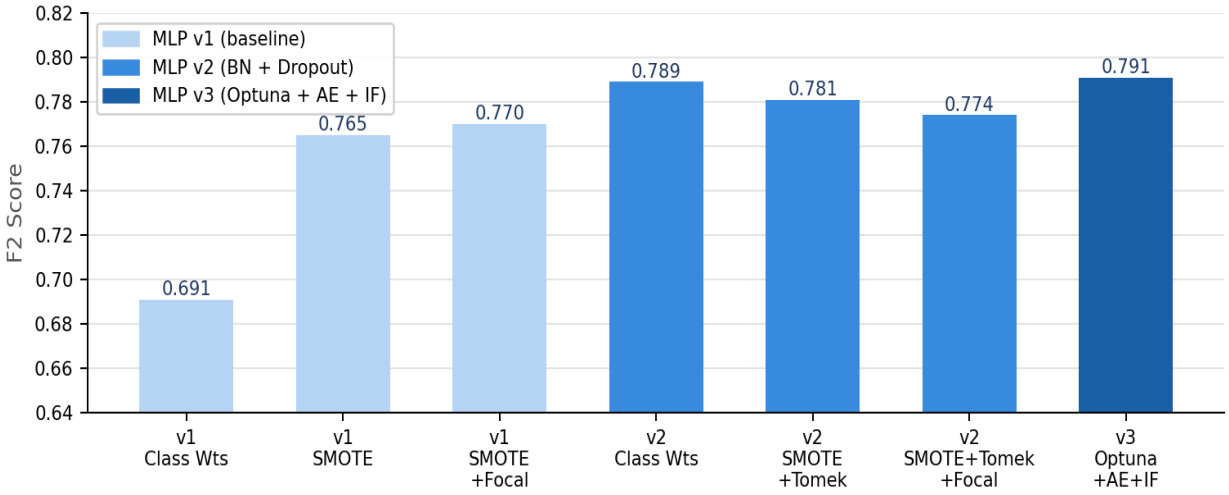

Рисунок 3.18 Прогрес між версіями MLP за F2-оцінкою та іншими метриками

Перша модель *MLP v1* слугувала свідомо простою відправною точкою: тришарова повнозв'язна мережа (30→256→128→64→1), навчена за трьома різними стратегіями боротьби з дисбалансом класів. Жодної нормалізації

пакетів та регуляризації Dropout не застосовувалося, тому будь-яке варіювання результатів можна безпосередньо приписати стратегії дисбалансу, а не архітектурним відмінностям.

Найкраща конфігурація v1: SMOTE + Focal Loss (F2=0,770). Focal Loss знижував ваги для легких негативних прикладів під час навчання, а SMOTE забезпечував синтетичними позитивними зразками. Разом вони дозволили моделі зосередитися на граничних фродових транзакціях. Результати всіх конфігурацій MLP v1 наведено у таблиці 3.17.

Результати v1 підтвердили, що стратегія боротьби з дисбалансом класів суттєво впливає на якість, однак також виявили стелю: навіть найкраща конфігурація не перевищувала F2=0,770. Щоб досягти подальшого прогресу, в архітектуру v2 було введено Batch Normalisation після кожного лінійного шару та Dropout зі скиданням 50% між прихованими шарами. Стратегії управління дисбалансом залишилися незмінними, що дозволило провести пряме порівняння.

Таблиця 3.17 – Конфігурації MLP v1, базова архітектура (3 шари, без BN/Dropout)

Конфігурація	*t**	*TP*	*FP*	*FN*	*F2*	*ROC-AUC*
Зважування класів	0,887	60	78	14	0,691	0,957
SMOTE	0,980	58	25	16	0,765	0,963
SMOTE + Focal Loss	0,709	59	28	15	0,770	0,975

Найкраща конфігурація v2: зважування класів (F2=0,789). Проста стратегія дисбалансу — зважування класів — перевершила складніші підходи зі збільшенням вибірки. Регуляризація BN+Dropout, вочевидь, зменшила перенавчання, яке маскувало переваги зважування класів у v1. Результати всіх конфігурацій MLP v2 наведено у таблиці 3.18.

Таблиця 3.18 – Конфігурації MLP v2, покращена архітектура (BN + Dropout)

Конфігурація	*t**	*TP*	*FP*	*FN*	*F2*	*ROC-AUC*
Зважування класів	0,990	62	35	12	0,789	0,972
SMOTE+Tomek	0,887	60	28	14	0,781	0,962
SMOTE+Tomek + Focal Loss	0,635	61	37	13	0,774	0,981

Версія 3 переслідувала два одночасних покращення. По-перше, ручний підбір гіперпараметрів був замінений деревоподібним руйнівником сімей Парзена (TPE) через Optuna — з 100 пробними запусками на валідаційній вибірці. По-друге, до 30 ознак PCA було додано два скалярні ознаки на основі неконтрольованого навчання: помилку реконструкції автоенкодера (AE) та оцінку аномалії Isolation Forest (IF), навчених виключно на негативних прикладах.

Оптимальна архітектура, знайдена Optuna, помітно вужча у пізніших шарах порівняно з вручну спроєктованою v2. Це свідчить про те, що більш агресивна компресія ознак у пізніх шарах допомагає уникнути перенавчання до основної незначущої більшості. Комбінація ознак AE+IF додала +0,002 F2 поверх чистого Optuna, підтверджуючи, що неконтрольовані детектори аномалій несуть доповнювальний сигнал.

3.3.8 Порівняння MLP та TabNet

TabNet застосовує принципово інший підхід до навчання на табличних даних. Замість застосування однакової трансформації до всіх ознак одночасно, TabNet використовує механізм послідовної уваги, який на кожному кроці вибирає підмножину ознак для обробки. Це покликане відтворити поведінку дерева рішень у диференційовній формі.

Найкраща архітектура використовує чотири кроки прийняття рішень із широкими розмірностями представлення ознак та відносно великий розмір пакета (4096) — на порядок більше, ніж у MLP. Великий пакет може відображати потребу TabNet у більш стабільній оцінці уваги на кожному кроці.

На рисунку 3.19 порівнюються дві моделі за п'ятьма метриками. Картина, яка виникає, не є картиною явного переможця в усіх вимірах; натомість обидві моделі демонструють різні профілі сильних та слабких сторін, що відображають їхні архітектурні відмінності.

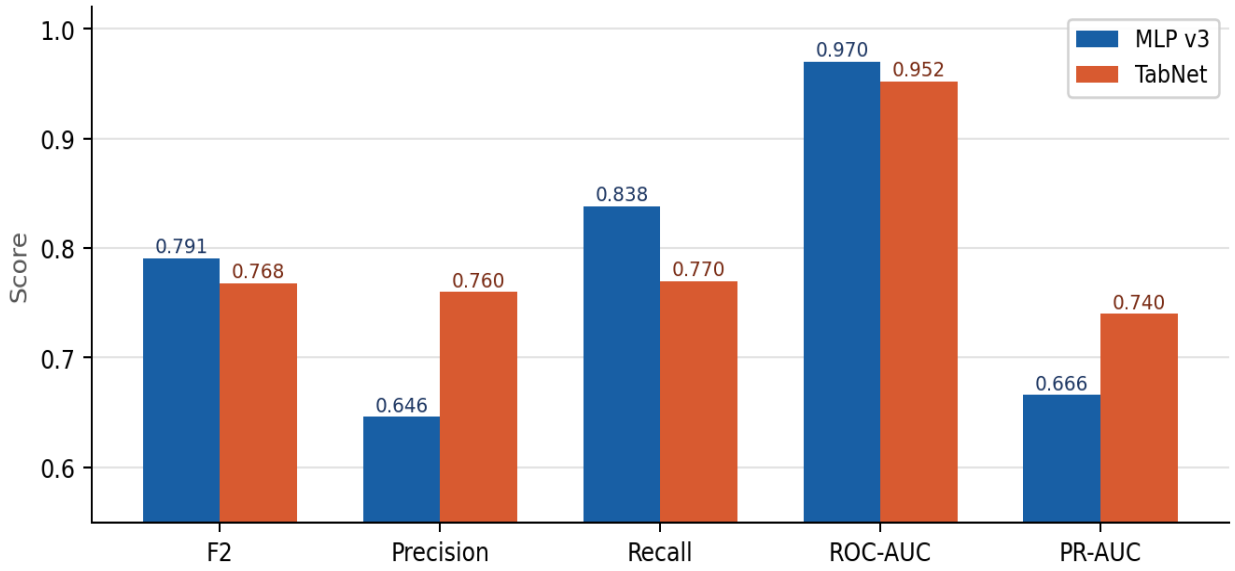

Рисунок 3.19 Порівняння MLP v3 та TabNet за п'ятьма метриками оцінки

MLP v3 лідирує за F2-оцінкою, повнотою та ROC-AUC. Перевага за повнотою є особливо значущою у контексті виявлення фроду: 68 додаткових фродових транзакцій (у розрахунку на 1000 фродових зразків тестової вибірки) виявляється за рахунок генерування 16 додаткових хибнопозитивних результатів. Аналіз помилок представлено у таблиці 3.19.

Таблиця 3.19 – Аналіз помилок: хибнопозитивні та хибнонегативні результати

Тип помилки	*MLP v3*	*TabNet*	*Різниця*	*Наслідок*
Хибнопозитивні (FP)	34	18	+16 MLP	Незручності для клієнта
Хибнонегативні (FN)	12	17	+5 TabNet	Невиявлений фрод

Рекомендації щодо вибору моделі залежно від пріоритетів наведено у таблиці 3.20.

Таблиця 3.20 – Коли надавати перевагу кожній моделі

Надавати перевагу MLP v3, коли...	*Надавати перевагу TabNet, коли...*
Максимізація частки виявленого фроду (повнота) є основною метою.	Мінімізація помилкових відхилень (незручності для клієнта) є пріоритетом.
Необхідний вищий ROC-AUC для ранжування транзакцій за ризиком.	Вища точність потрібна для черг перевірки операторами.
Хибнонегативні мають вищу фінансову або регуляторну вартість, ніж хибнопозитивні.	PR-AUC є цільовою метрикою.

На рисунку 3.20 компроміс проілюстровано конкретно. MLP v3 генерує 34 хибнопозитивних та 12 хибнонегативних результатів; TabNet генерує 18 хибнопозитивних та 17 хибнонегативних результатів. Якщо пропущений фрод (хибнонегативний) оцінюється приблизно як у 5–10 разів дорожче, ніж помилкове блокування транзакції (хибнопозитивний), то MLP v3 є кращим вибором для мінімізації очікуваних витрат.

Фундаментальна відмінність у дизайні архітектурі моделей представлена на рисунку 3.21. MLP v3 передає кожну ознаку через усі нейрони на кожному шарі, тоді як TabNet вибірково активує підмножини ознак на кожному кроці обробки.

У MLP v3 кожна вхідна ознака проходить через всі три приховані шари послідовно. Кожний шар спочатку застосовує лінійне перетворення, потім

Batch Normalisation, потім Dropout (маскує 33–42% нейронів під час навчання, запобігаючи спільній адаптації) і, нарешті, активацію ReLU. Звужувальна послідовність [256→32→16] стискає простір ознак перед кінцевим скалярним виходом. Поріг перетворює неперервну оцінку ймовірності на бінарну мітку; значення t*=0,911 відображає сильно асиметричну вартість, де пропущений фрод є набагато дорожчим, ніж помилкове блокування [19, 20, 25].

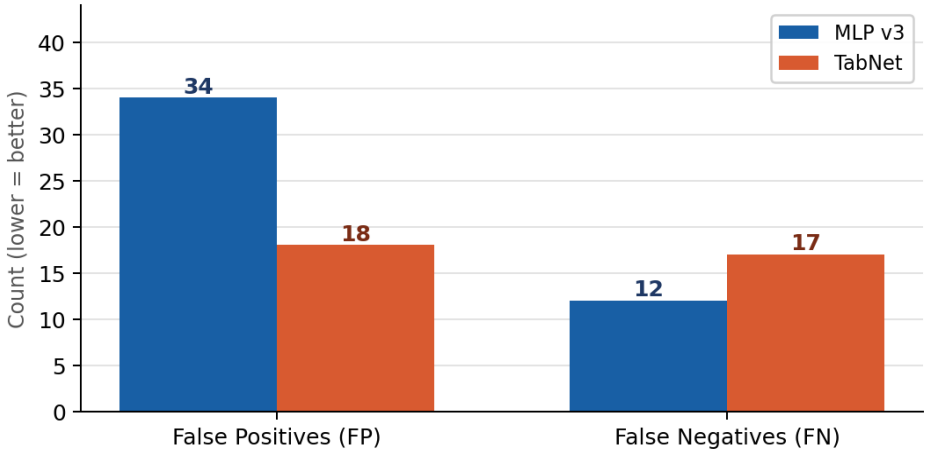

Рисунок 3.20 Матриці помилок: MLP v3 та TabNet (справа)

У TabNet ті самі 30-вимірні вхідні дані обробляються чотири рази послідовно — по одному на кожному кроці прийняття рішень. На кожному кроці механізм уваги обчислює розподіл важливості ознак, що зосереджується на підмножині ознак і дозволяє решті "пройти крізь". Потім трансформатор ознак застосовує нелінійне перетворення лише до вибраних ознак. Накопичені представлення всіх чотирьох кроків підсумовуються та передаються кінцевому вихідному шару. Регуляризатор розрідженості заохочує кожен крок зосереджуватися на невеликих підмножинах ознак, а не розподіляти увагу рівномірно [23].

Архітектурна відмінність безпосередньо відображається у спостережуваному розходженні метрик. Щільна зв'язність MLP v3 означає, що кожен нейрон може потенційно реагувати на будь-яку комбінацію вхідних

ознак — включаючи тонкі взаємодії високого порядку між компонентами PCA, які можуть асоціюватися з фродовою поведінкою [24]. Це сприяє вищій повноті, модель з більшою ймовірністю вловлює нетипові фродові транзакції, що не вписуються в жоден ізольований ознаковий шаблон. Натомість вибіркова увага TabNet створює більш інтерпретовані, але вужчі правила рішень, які краще узгоджуються з точністю та PR-AUC за рахунок деякого зниження повноти.

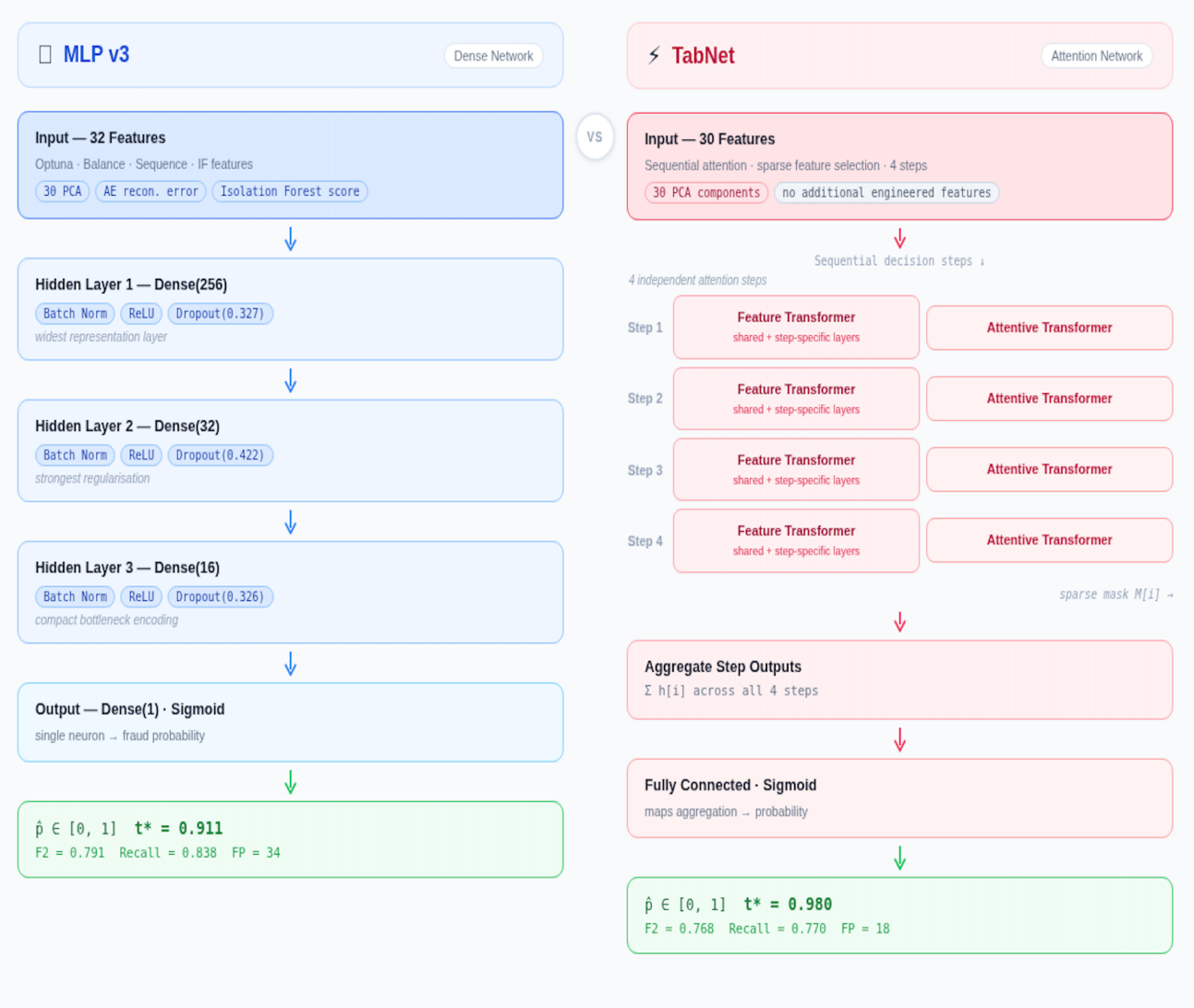

Рисунок 3.21 Порівняльна діаграма архітектур: MLP v3 та TabNet

3.4 Порівняння моделей

Усі моделі було оцінено на відкладеній тестовій вибірці (42 722 транзакції, 74 фродові) при F2-максимізуючому валідаційному порозі. Метрики KNN (k=15 та k=21) отримано з повної оцінки тестової вибірки; точність і повнота обидві повідомляються при одному F2-оптимальному порозі t* (з виправленням артефакту змішаного порогу у скані чутливості до k). Продуктивність усіх восьми моделей на тестовій вибірці наведено у таблиці 3.21.

Таблиця 3.21 – Продуктивність усіх восьми моделей на тестовій вибірці при F2-оптимальних порогах

Модель	*Поріг*	*TP*	*FP*	*FN*	*Precision*	*Recall*	*F2*	*ROC-AUC*	*Сер. точн.*
KNN k=15	0,937	59	11	15	0,843	0,797	0,806	0,932	0,713
KNN k=21	0,957	59	8	15	0,881	0,797	0,813	0,932	0,745
Random Forest	0,17	62	8	12	0,886	0,838	0,847	0,937	0,937
XGBoost (базовий)	0,29	61	11	13	0,847	0,824	0,836	0,829	0,966
XGBoost (SMOTE+AE+IF)	0,39	—	—	—	0,768	0,851	0,808	0,833	0,984
LightGBM (Optuna)	0,457	62	19	12	0,765	0,838	0,800	0,822	0,974
Логістична регресія	0,99	60	29	14	0,674	0,811	0,736	0,779	0,968
Дерево рішень (гл=5)	0,99	56	36	18	0,609	0,757	0,675	0,722	0,891
MLP v3	0.911	62	34	12	0.646	0.838	0.791	0.97	0.666
TabNet	0.98	57	18	17	0.76	0.77	0.768	0.952	0.74

Порівняння всіх моделей та відповідних метрик представлено на рисунку 3.22. Криві Точність–Повнота для всіх моделей на тестовій вибірці представлено на рисунку 3.23.

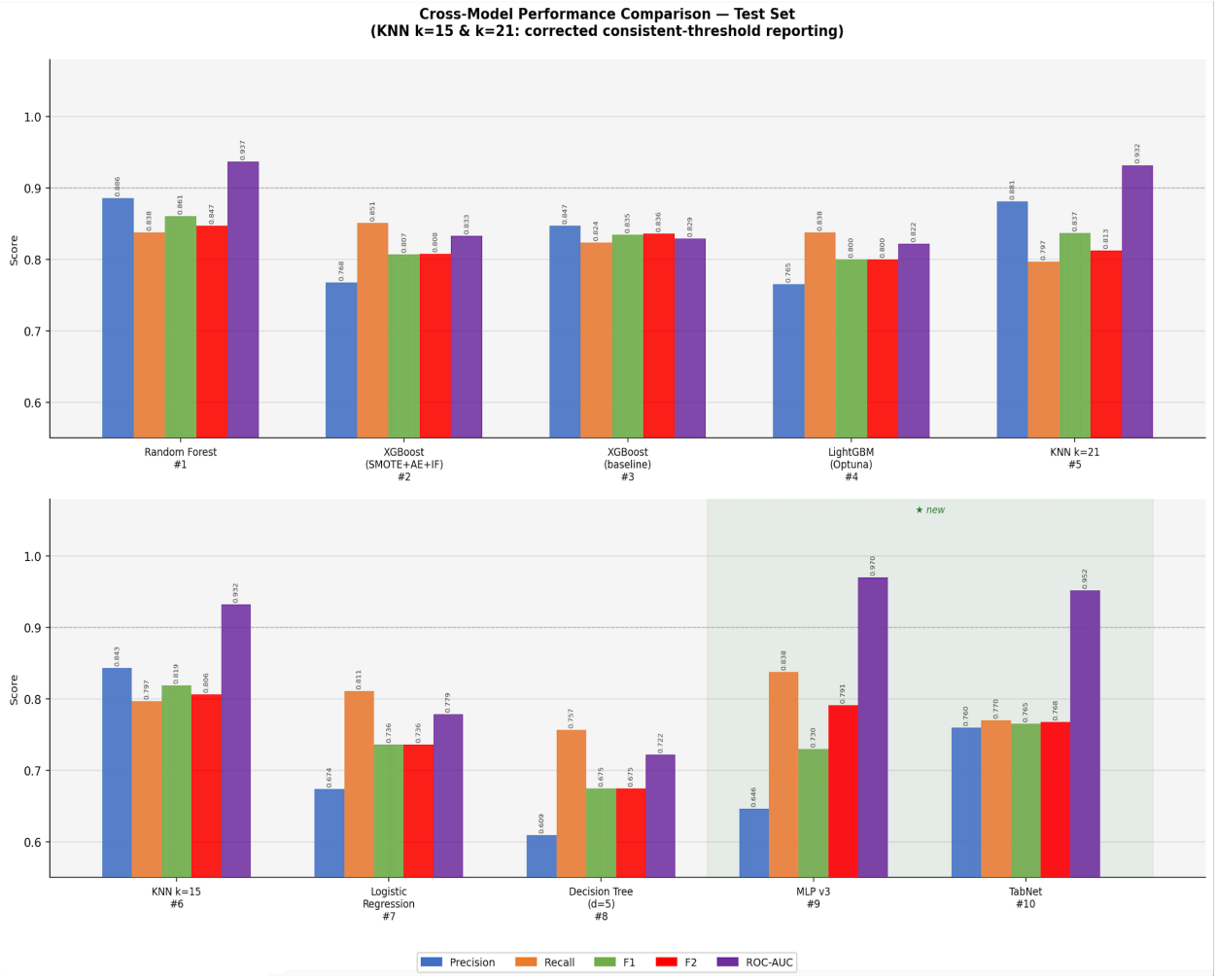

Рисунок 3.22 – Порівняння моделей

На тестовій вибірці KNN k=21 досяг найвищого F2 серед конфігурацій KNN (F2=0,813, Точність=0,881, Повнота=0,797, FP=8), незначно перевершивши KNN k=15 (F2=0,806, FP=11). Обидва результати KNN використовують скориговане однорідне звітування порогів: точність і повнота вимірюються при одному t* (0,937 та 0,957 відповідно). Серед усіх моделей Random Forest залишається лідером за F2 (0,847), за яким ідуть конфігурації XGBoost.

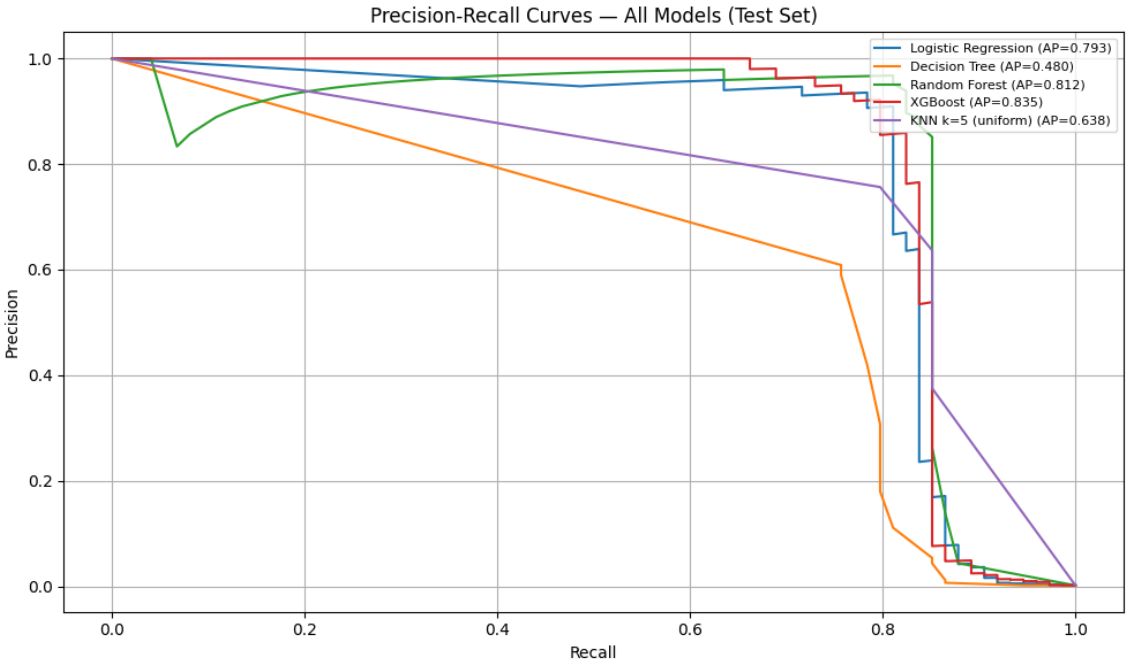

Рисунок 3.23 – Криві Точність–Повнота для дерев та регресійних моделей на тестовій вибірці

Вдалось показати, що на структурованих табличних даних ансамблеві методи на основі дерев перевершують архітектури нейронних мереж. Нейронні мережі розроблені для виявлення ієрархічних представлень із сирих, багатовимірних, неструктурованих вхідних даних, таких як зображення чи текст, і їхня перевага суттєво зменшується, коли ознаки вже є значущими числовими змінними, а не сирими сигналами, що потребують глибокого перетворення. Моделі на основі дерев, навпаки, розділяють простір ознак за допомогою послідовних бінарних розбиття, які природним чином обробляють нерегулярні, негладкі межі рішень, ігнорують неінформативні ознаки без явної регуляризації та ефективно узагальнюють з невеликих навчальних наборів. За умов дисбалансу класів ансамблі з градієнтним підвищенням концентрують потужність моделі саме на класі меншин за допомогою своєї ітеративної процедури підгонки залишків, тоді як нейронні мережі оптимізують глобальні втрати за всіма параметрами одночасно, що ускладнює досягнення фокусу на класі меншин навіть за допомогою явного зважування класів. Крім того,

моделі на основі дерев вимагають менше гіперпараметрів, менш чутливі до масштабування ознак і створюють краще калібровані вихідні дані ймовірності — всі ці властивості особливо цінні в умовах виявлення шахрайства, де навчальний набір невеликий, ознаки попередньо оброблені, а для оптимізації порогу потрібні надійні оцінки ймовірності [26].

Висновки до розділу 3

Було оцінено вісім підходів машинного навчання до виявлення фрод-транзакцій за кредитними картками на наборі даних ULB, просуваючись від лінійної базової моделі через ансамблеві методи до повністю розробленого конвеєра LightGBM з несупервізованими аномалійними ознаками та байєсівською оптимізацією гіперпараметрів та двух нейронних мереж з різними архітектурами.

По всіх моделях найбільші покращення продуктивності досягнуто не через алгоритмічну складність, а через правильне калібрування порогу. Результати, розглянуті у розділі 3.3 підтверджують, що оптимізація порогу на вибірці є необхідною передумовою для будь-якого практичного класифікатора фроду, незалежно від базового алгоритму.

Випадковий ліс та XGBoost стабільно перевершували одиночні моделі. Повна оцінка тестової вибірки підтверджує KNN як конкурентоспроможний, але обчислювально витратний варіант.

Доповнення 30 PCA-ознак помилкою реконструкції автоенкодера та аномалійною оцінкою Isolation Forest забезпечило вимірне покращення для моделей.

Щодо нейронних моделей, у завданні виявлення фродових транзакцій вартість пропущеного фроду (хибнонегативний) зазвичай значно перевищує вартість помилкового блокування (хибнопозитивний). MLP v3 максимізує F2-

оцінку та досягає вищої повноти ніж TabNet — це безпосередньо відповідає операційному пріоритету мінімізації пропущеного фроду. Обидві нейронні моделі відстають від базового рівня Random Forest.

Залишковий розрив у 0,056 балів порівняно з Random Forest відповідає добре відомій перевазі ансамблевих деревоподібних методів на однорідних табличних наборах даних. Ансамблі дерев мають вбудовані механізми відбору ознак, не потребують нормалізації та природно вловлюють взаємодії ознак без явної архітектурної індуктивної переваги. Нейронні мережі можуть відновити цей розрив лише за рахунок більш агресивного проектування ознак або суттєво більших наборів даних. Також нейронні мережі показують кращі показники навчання на датасетах з неперетвореними даними, що є ускладненим у банківській сфері, де дані є секюрними, та в розглянутому датасеті дані

Вибір моделі для розгортання у виробничому середовищі залежить від відносної вартості хибнонегативних результатів проти хибнопозитивних. У проблемі боротьби з чарджбеками варто зробити ухил у збалансованість з підвищеною важливістю Recall ознаки.

ВИСНОВКИ

У дипломній роботі було досліджено проблему виявлення фродових транзакцій у цифрових платіжних системах як один із ключових напрямів мінімізації ризиків чарджбеків для бізнесу. Актуальність теми зумовлена стрімким зростанням кількості онлайн-платежів, розвитком електронної комерції та збільшенням обсягів транзакцій без фізичної присутності картки. У таких умовах мерчанти стикаються не лише з прямими фінансовими втратами від фроду операцій, а й з додатковими витратами на обробку спорів, комісіями, штрафами платіжних систем, репутаційними ризиками та можливістю обмеження платіжного процесингу.

У першому розділі було розглянуто сутність чарджбеку, історію його виникнення, основних учасників процесу та механізм його роботи. Встановлено, що чарджбек має подвійну природу. З одного боку, він є важливим інструментом захисту прав споживачів, оскільки дає власникам карток можливість оскаржувати несанкціоновані, помилкові або спірні транзакції. З іншого боку, для бізнесу чарджбеки створюють значні фінансові, операційні та репутаційні ризики. Кожен чарджбек може призводити до втрати доходу, додаткових комісій, витрат на підготовку доказової бази та участь у процедурі репрезентменту. У разі систематичного перевищення допустимих рівнів чарджбеків мерчант може потрапити до моніторингових програм платіжних систем, що підвищує ризик штрафів, погіршення умов еквайрингу або втрати можливості приймати карткові платежі.

Також було проаналізовано основні типи чарджбеків і визначено роль фроду як одного з ключових факторів їх виникнення. У роботі розмежовано дружній фрод і справжній фрод, оскільки ці явища мають різну природу та потребують різних підходів до управління. Дружній фрод пов’язаний із діями самого власника картки, який може оскаржити легітимну транзакцію через непорозуміння, невпізнаваний платіжний дескриптор, незадоволення

продуктом, складну процедуру повернення коштів або навмисне зловживання механізмом чарджбеку. Було встановлено, що частину таких випадків бізнес може зменшити шляхом покращення клієнтського досвіду, прозорої комунікації, зрозумілих умов оплати, ефективної підтримки та якісної політики повернення коштів.

На відміну від дружнього фроду, справжній фрод виникає внаслідок несанкціонованого використання платіжних даних третьою особою. Було показано, що після завершення такої транзакції можливості мерчанта суттєво обмежені, оскільки товар або послуга вже можуть бути надані, а власник картки справді не авторизував операцію. Тому постфактум-реагування не дозволяє повністю уникнути фінансових втрат. Це підтверджує необхідність превентивного підходу, за якого ризикові транзакції мають виявлятися до моменту їх завершення або до надання товару чи послуги.

У другому розділі було сформульовано задачу виявлення фродових транзакцій як задачу бінарної класифікації з учителем в умовах сильного дисбалансу класів. Було розглянуто математичні основи машинного навчання для цієї задачі, зокрема формалізацію навчальної вибірки, функцію прийняття рішення, функції втрат, матрицю помилок та основні метрики оцінювання моделей. Окрему увагу приділено тому, що стандартна метрика accuracy є недостатньо інформативною для задачі фрод-детекшну, оскільки за наявності сильного дисбалансу модель може демонструвати високу загальну точність, фактично не виявляючи шахрайських операцій.

У зв’язку з цим у роботі було обґрунтовано використання метрик Precision, Recall, F1, F2, ROC-AUC та Average Precision. Особливе значення для задачі мінімізації чарджбеків має Recall, оскільки кожна пропущена фродова транзакція може призвести до прямої фінансової втрати та подальшого чарджбеку. Водночас надмірна кількість хибнопозитивних спрацювань також є небажаною, оскільки вона може блокувати легітимних клієнтів і створювати додаткове операційне навантаження. Саме тому F2-міра

була використана як важлива метрика, що надає більшої ваги повноті, але водночас враховує точність моделі.

У роботі було проаналізовано кілька моделей машинного навчання, які можуть застосовуватися для виявлення фродових транзакцій: логістичну регресію, дерево рішень, Random Forest, XGBoost, LightGBM, KNN, MLP та TabNet. Було встановлено, що прості моделі, такі як логістична регресія та дерево рішень, мають перевагу в інтерпретованості, однак обмежені у здатності виявляти складні нелінійні патерни шахрайства. Ансамблеві методи, зокрема Random Forest, XGBoost і LightGBM, краще підходять для задачі фрод-детекшну, оскільки здатні моделювати складні залежності між ознаками, працювати з незбалансованими даними та забезпечувати вищу якість класифікації.

У третьому розділі було виконано практичну реалізацію та експериментальне порівняння моделей на наборі даних ULB Credit Card Fraud Detection. Набір містить 284 807 транзакцій, з яких лише 492 є фродовими, що відповідає частці фроду 0,173%. Такий розподіл відображає реалістичну ситуацію у фінансових даних, де шахрайські операції є рідкісними, але мають високу вартість помилки. Було проведено попередню обробку даних, стандартизацію ознак, стратифікований поділ на тренувальну, валідаційну та тестову вибірки, а також застосовано різні підходи до роботи з дисбалансом класів.

Для покращення здатності моделей виявляти фродові транзакції було використано ваги класів, метод SMOTE, комбінацію SMOTE+Tomek Links, а також додаткові аномалійні ознаки, отримані за допомогою автоенкодера та Isolation Forest. Це дозволило посилити сигнал міноритарного класу та покращити здатність моделей розрізняти легітимні й фродові транзакції. Окремо було виконано оптимізацію порогу класифікації для кожної моделі на валідаційній вибірці з метою максимізації F2-міри. Було встановлено, що саме налаштування порогу є одним із найважливіших етапів побудови практичної

системи фрод-детекшну, оскільки стандартний поріг 0,5 не є оптимальним для сильно незбалансованих даних.

За результатами експериментів найкращу якість за F2-мірою продемонструвала модель Random Forest. Вона досягла F2 = 0,847, Precision = 0,886 та Recall = 0,838 при лише 8 хибнопозитивних спрацюваннях на тестовій вибірці. Це свідчить про високий баланс між здатністю виявляти фродові транзакції та низьким рівнем помилкового блокування легітимних операцій. XGBoost також показав сильні результати та підтвердив свою ефективність для задачі фрод-детекшну. Розширена конфігурація XGBoost із використанням аномалійних ознак продемонструвала найвищий показник ROC-AUC, що свідчить про високу здатність моделі ранжувати транзакції за рівнем ризику.

LightGBM показав добру масштабованість і придатність до застосування на великих наборах даних, а також переваги завдяки використанню оптимізованих гіперпараметрів. KNN продемонстрував конкурентоспроможні результати, однак його практичне використання обмежується високими обчислювальними витратами під час прогнозування та відсутністю механізмів пояснення важливості ознак. Логістична регресія та дерево рішень, попри нижчі результати порівняно з ансамблевими моделями, залишаються корисними як базові та більш інтерпретовані підходи.

Нейронні мережі, як правило, перевершують інших, коли неструктуровані дані потребують глибокого вилучення ознак (зображення, текст, послідовності). На структурованих табличних даних зі спроектованими або PCA ознаками, невеликим класом меншості та скромним розміром набору даних, ансамблі градієнтного підсилення постійно перемагають, оскільки вони краще відповідають геометрії задачі.

У результаті роботи було підтверджено, що ефективне виявлення фродових транзакцій потребує комплексного підходу, який поєднує якісну попередню обробку даних, коректну роботу з дисбалансом класів, вибір відповідної моделі машинного навчання та оптимізацію порогу прийняття

рішень. Було встановлено, що складність моделі сама по собі не гарантує найкращого результату: значний вплив на якість має саме калібрування порогу відповідно до бізнес-мети, тобто досягнення балансу між мінімізацією пропущеного фроду та обмеженням кількості хибних спрацювань.

Практичне значення отриманих результатів полягає в тому, що запропонований підхід може бути використаний як основа для побудови системи раннього виявлення фродових транзакцій. Така система здатна допомогти бізнесу зменшити кількість шахрайських операцій, скоротити рівень чарджбеків, знизити фінансові втрати, зменшити операційне навантаження на команди ризик-менеджменту та підтримки, а також уникнути санкцій і моніторингових програм платіжних систем. Особливо важливим є те, що виявлення фроду до завершення транзакції дозволяє не лише реагувати на вже здійснені порушення, а й запобігати їхнім наслідкам.

Таким чином, мету дипломної роботи було досягнуто: розроблено, проаналізовано та порівняно моделі машинного навчання для виявлення фродових транзакцій з метою мінімізації ризиків чарджбеків. У роботі було показано, що ансамблеві моделі, зокрема Random Forest та XGBoost, є найбільш ефективними для досліджуваної задачі, а правильне налаштування порогу класифікації є критично важливим для практичного застосування. Отримані результати підтверджують доцільність використання методів машинного навчання у сфері платіжного ризик-менеджменту та демонструють їхню важливість для забезпечення фінансової стійкості бізнесу в умовах зростання цифрових платежів і фродових загроз.

ПЕРЕЛІК ПОСИЛАНЬ

1. Chargeback.io (no date) History of chargebacks [Online]. Available at: https://www.chargeback.io/blog/history-of-chargebacks [Accessed: 1 May 2026].

2. Adyen (no date) Understanding chargebacks [Online]. Available at: https://www.adyen.com/knowledge-hub/understanding-chargebacks [Accessed: 1 May 2026].

3. Mastercard (2025) What's the true cost of a chargeback in 2025? [Online]. Available at: https://www.mastercard.com/us/en/news-and-trends/Insights/2025/what-s-the-true-cost-of-a-chargeback-in-2025.html [Accessed: 1 May 2026].

4. Chargeback Gurus (no date) The chargeback process [Online]. Available at: https://www.chargebackgurus.com/blog/chargeback-process [Accessed: 1 May 2026].

5. Solidgate (no date) VAMP [Online]. Available at: https://solidgate.com/glossary/vamp/ [Accessed: 1 May 2026].

6. Mastercard (2025) 2025 chargeback state of the market report. Purchase, NY: Mastercard.

7. Checkout.com (no date) What is friendly fraud and how can you deal with it? [Online]. Available at: https://www.checkout.com/blog/what-is-friendly-fraud-and-how-can-you-deal-with-it [Accessed: 1 May 2026].

8. Ethoca (no date) First-party fraud, friendly fraud, third-party fraud: what do these terms mean? [Online]. Available at: https://www.ethoca.com/blog/first-party-fraud-friendly-fraud-third-party-fraud-what-do-these-terms-mean [Accessed: 1 May 2026].

9. Hastie, T., Tibshirani, R. and Friedman, J. (2009) The elements of statistical learning: data mining, inference, and prediction. 2nd edn. New York: Springer.

10. Breiman, L., Friedman, J., Olshen, R. and Stone, C. (1984) Classification and regression trees. Belmont, CA: Wadsworth.

11. Quinlan, J.R. (1986) 'Induction of decision trees', Machine Learning, 1(1), pp. 81–106.

12. Breiman, L. (2001) 'Random forests', Machine Learning, 45(1), pp. 5–32.

13. Friedman, J.H. (2001) 'Greedy function approximation: a gradient boosting machine', Annals of Statistics, 29(5), pp. 1189–1232.

14. Chen, T. and Guestrin, C. (2016) 'XGBoost: a scalable tree boosting system', Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, 13–17 August, pp. 785–794.

15. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.-Y. (2017) 'LightGBM: a highly efficient gradient boosting decision tree', Advances in Neural Information Processing Systems, 30, pp. 3146–3154.

16. Cover, T.M. and Hart, P.E. (1967) 'Nearest neighbor pattern classification', IEEE Transactions on Information Theory, 13(1), pp. 21–27.

17. Nizan, O. (2024) 'k-NNN: nearest neighbors of neighbors for anomaly detection', Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW), Waikoloa, HI, January.

18. Goodfellow, I., Bengio, Y. and Courville, A. (2016) Deep learning. Cambridge, MA: MIT Press.

19. Ioffe, S. and Szegedy, C. (2015) 'Batch normalization: accelerating deep network training by reducing internal covariate shift', *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, Lille, France, 6–11 July, pp. 448–456.

20. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014) 'Dropout: a simple way to prevent neural networks from overfitting', *Journal of Machine Learning Research*, 15(1), pp. 1929–1958.

21. Liu, F.T., Ting, K.M. and Zhou, Z.-H. (2008) 'Isolation forest', *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM)*, Pisa, Italy, 15–19 December, pp. 413–422.

22. Bergstra, J., Bardenet, R., Bengio, Y. and Kégl, B. (2011) 'Algorithms for hyper-parameter optimization', *Advances in Neural Information Processing Systems*, 24, pp. 2546–2554.

23. Arik, S. Ö., & Pfister, T. (2021). *TabNet: Attentive Interpretable Tabular Learning*. Proceedings of the AAAI Conference on Artificial Intelligence, 35(8), 6679–6687.

24. Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986). "Learning representations by back-propagating errors." *Nature*, 323(6088), pp. 533–536.

25. Akiba, T., Sano, S., Yanase, T., Ohta, T. and Koyama, M. (2019). "Optuna: A next-generation hyperparameter optimization framework." *Proceedings of the 25th ACM SIGKDD*, pp. 2623–2631.

26. Grinsztajn, L., Oyallon, E. and Varoquaux, G. (2022) 'Why tree-based models still outperform deep learning on tabular data', Advances in Neural Information Processing Systems (NeurIPS), 35, pp. 507–520.

27. Panibratov, R.S. and Bidyuk, P.I. (2025) 'Analysis of actuarial risk with generalized linear models', *System Research & Information Technologies*, (4). DOI: 10.20535/SRIT.2308-8893.2025.4.04.

28. ppdkk (2026) *diploma_notebooks*. Available at: https://github.com/ppdkk/diploma_notebooks (Accessed: 21 May 2026).