МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Західноукраїнський національний університет
Факультет комп'ютерних інформаційних технологій
Кафедра інформаційно-обчислювальних систем і управління

ВАСИЛЕНКО Остап Андрійович

Інтелектуальна система виявлення шахрайства в електронній комерції / Intelligent Fraud Detection System in E-Commerce

Спеціальність 122 – Комп'ютерні науки
Освітньо-професійна програма – Комп'ютерні науки

Кваліфікаційна робота

Виконав студент групи КН-43
О.А. Василенко

Науковий керівник:
к.е.н., доцент, О. Ю. Ніпіаліді

Кваліфікаційну роботу допущено до
захисту
«___»_____ 2025 р.

В.о. завідувача кафедри
_____ Н.М. Васильків

Тернопіль – 2025

ЗМІСТ

Вступ 3
1 Аналіз предметної області виявлення шахрайства в електронній комерції 6
1.1 Опис предметної області 6
1.2 Аналіз відомих рішень 8
1.3 Постановка задачі дослідження 13
2 Проєктування системи виявлення шахрайства в електронній комерції 16
2.1 Концепція побудови системи 16
2.2 Архітектура системи виявлення шахрайства 17
2.3 Методи та алгоритми машинного навчання для виявлення шахрайства в електронній комерції 19
3 Тестування системи виявлення шахрайства в електронній комерції 31
3.1 Постановка експериментальних досліджень 31
3.2 Результати експериментальних досліджень 34
3.3 Напрями подальших досліджень 38
Висновки 41
Список використаних джерел 43
Додаток А Копія публікації 46

ВСТУП

Фінансовим шахрайством називають дії, спрямовані на отримання матеріальної вигоди за допомогою незаконних і обманних способів [1]. Такі порушення можуть відбуватися у різних сферах фінансової діяльності, зокрема у корпоративному секторі, банківській справі, страхуванні та оподаткуванні. В останні роки у світі спостерігається стрімке зростання кількості випадків фінансового шахрайства, яке проявляється через такі форми, як відмивання грошей, фальсифікація фінансових операцій та інші види злочинної діяльності у сфері фінансів [2].

Попри численні спроби зменшити рівень шахрайства у фінансовому секторі, ця проблема залишається актуальною та продовжує щодня негативно впливати як на економіку, так і на суспільство в цілому. Щороку через шахрайські дії втрачаються значні суми коштів, що підриває довіру до електронної комерції та фінансових послуг [3].

Перші спроби виявлення шахрайських дій у фінансовій сфері були зроблені багато років тому. Більшість ранніх підходів передбачали ручну обробку інформації, яка є трудомісткою, затратною за часом, неточною та вимагає значних фінансових ресурсів. Сучасні дослідження спрямовані на зменшення втрат, спричинених фінансовим шахрайством, однак їх результати не завжди можна застосувати на практиці [3].

Із розвитком штучного інтелекту з’явилися нові можливості для виявлення шахрайських дій за допомогою машинного навчання та інтелектуального аналізу даних [4]. Ці підходи дозволяють автоматично аналізувати великі обсяги транзакцій та виявляти підозрілі шаблони поведінки, які можуть свідчити про порушення. Поєднання методів із учителем і без учителя дозволяє будувати моделі для прогнозування шахрайської поведінки на основі попередніх даних.

Найпоширенішим підходом у виявленні підозрілих фінансових транзакцій є використання алгоритмів класифікації. У такому випадку модель машинного навчання тренується на основі набору даних, що містить приклади типових

транзакцій із позначеними класами (наприклад, «нормальна» або «шахрайська» транзакція) [5]. Після завершення навчання модель застосовується до нових прикладів, щоб класифікувати їх на основі отриманих знань.

Метою кваліфікаційної роботи є побудова ефективної системи виявлення шахрайства в електронній комерції на основі методів машинного навчання, з урахуванням особливостей транзакційних даних, поведінкових патернів користувачів, а також характерного дисбалансу між шахрайськими та легітимними транзакціями.

Виходячи з поставленої мети, у межах дослідження сформульовано такі основні задачі:

– провести аналіз сучасних методів виявлення шахрайства в електронній комерції, охарактеризувати їх переваги та недоліки, виявити наукові та практичні прогалини;

– зібрати, дослідити та підготувати реальні транзакційні дані з маркетплейсів для побудови моделей машинного навчання;

– розробити концепцію інтелектуальної системи виявлення шахрайства, яка включає модулі попередньої обробки даних, генерації ознак, навчання моделей і оцінки результатів;

– реалізувати кілька моделей класифікації шахрайських транзакцій (зокрема, дерева рішень, випадкові ліси, градієнтний бустинг, ансамблеві методи), а також налаштувати гіперпараметри для досягнення максимальної продуктивності;

– визначити та реалізувати стратегії боротьби з дисбалансом класів, у тому числі методи надсемплювання та undersampling, і дослідити їх вплив на точність моделі;

– провести порівняльний аналіз ефективності різних підходів за допомогою обраних метрик якості (точність, повнота, F1-міра, AUC-ROC тощо).

Об’єкт дослідження – процеси виявлення шахрайських транзакцій в електронній комерції, що здійснюються у цифрових торговельних середовищах, зокрема в онлайн-маркетплейсах.

Предмет дослідження – методи і алгоритми машинного навчання, а також підходи до обробки та балансування даних, які використовуються для побудови інтелектуальної системи виявлення шахрайства в електронній комерції.

Результати кваліфікаційної роботи апробовані та опубліковані у матеріалах студентської науково-практичної конференції “Інтелектуальні інформаційні технології в прикладних дослідженнях” (IITAR – 2025), м. Тернопіль, Україна, 27-29 травня 2025 р. (додаток А).

Кваліфікаційна робота складається із вступу, трьох розділів, висновків, списку використаних джерел та додатків.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ВИЯВЛЕННЯ ШАХРАЙСТВА В ЕЛЕКТРОННІЙ КОМЕРЦІЇ

1.1 Опис предметної області

У сучасному світі електронна комерція стрімко розвивається та стає одним із провідних напрямів цифрової економіки. Онлайн-платформи забезпечують зручний механізм купівлі та продажу товарів і послуг, що дозволяє учасникам ринку взаємодіяти без географічних обмежень. До ключових представників галузі електронної комерції належать такі платформи, як Amazon, eBay, Etsy, AliExpress, OLX, Rozetka, Prom.ua, а також спеціалізовані торговельні майданчики в соціальних мережах, зокрема Facebook Marketplace.

Однак з розширенням функціональності електронної торгівлі та зростанням кількості транзакцій значно зростає ризик шахрайських дій. Ці дії можуть проявлятися у різних формах: використання вкрадених платіжних карток, фальшиві акаунти продавців або покупців, маніпуляції з цінами, обман з доставкою, продаж неіснуючих товарів, накручування відгуків, шахрайські схеми повернення товару тощо. Такі дії не лише створюють фінансові ризики для компаній та користувачів, а й підривають довіру до онлайн-сервісів та репутацію бренду [1–3].

Особливої уваги заслуговує явище організованого роздрібного шахрайства (Organized Retail Crime, ORC). Йдеться про систематичні дії шахраїв або груп, спрямовані на зловживання торговельними платформами з метою отримання незаконного прибутку. До прикладу, одна й та сама особа може діяти через декілька фальшивих облікових записів, використовуючи викрадені або підроблені особисті дані, підключаючи автоматизовані інструменти для обходу систем безпеки.

Класичні методи боротьби з шахрайством в електронній комерції зазвичай ґрунтуються на використанні фіксованих правил або ручній перевірці транзакцій. Проте такі підходи мають обмеження: вони не є гнучкими, не враховують змін у поведінці користувачів і шахраїв, мають високу вартість

супроводу та низьку масштабованість. У зв’язку з цим виникає потреба у використанні сучасних інтелектуальних систем, які здатні автоматично аналізувати великі обсяги даних, виявляти аномальні транзакції, адаптуватися до нових сценаріїв шахрайства та зменшувати кількість хибнопозитивних рішень.

У центрі таких рішень лежать алгоритми машинного навчання, які на основі історичних даних можуть будувати моделі для передбачення шахрайства. Перевага цих моделей полягає у здатності виявляти складні, приховані взаємозв’язки між різними параметрами транзакцій, які важко виявити людиною або класичними методами. Застосування таких підходів передбачає не лише технічну реалізацію моделей, а й комплексну підготовку даних, що включає: очищення, нормалізацію, генерацію ознак, кодування категоріальних змінних, боротьбу з дисбалансом класів тощо [4, 5].

Предметна область дослідження також охоплює аспекти оцінки ефективності систем виявлення шахрайства, де важливо не лише досягати високої точності класифікації, але й враховувати специфіку бізнесу: фінансові втрати від пропущеного шахрайства, витрати на перевірку хибнопозитивних випадків, динаміку поведінки користувачів.

Крім того, сучасні підходи виявлення шахрайства все частіше орієнтовані на мультимодальність даних, що включає не лише числові та категоріальні ознаки (наприклад, сума, час, спосіб оплати), але й текстову інформацію (опис товару, коментарі, повідомлення), а також зображення товарів. У майбутньому саме поєднання таких джерел інформації забезпечить найбільшу точність і адаптивність систем безпеки.

Таким чином, предметна область даного дослідження охоплює:

– аналіз електронної комерції як середовища з високим рівнем ризику шахрайства;

– класифікацію типів шахрайських дій у цій галузі;

– проблеми й обмеження традиційних методів виявлення шахрайства;

– сучасні технології штучного інтелекту, зокрема машинного навчання, для розв’язання задач класифікації транзакцій;

– організаційні, технічні та бізнес-аспекти впровадження таких систем.

Дослідження у цій предметній області має не лише наукову, але й прикладну цінність, оскільки дає змогу підвищити рівень кібербезпеки, захистити споживачів, покращити бізнес-метрики та зменшити втрати від шахрайських дій в електронній комерції.

1.2 Аналіз відомих рішень

Упродовж останніх років було проведено багато оглядових робіт, присвячених аналізу методів виявлення фінансових правопорушень [6]. Одним із прикладів є дослідження шахрайства з банківськими картками, де розглядаються методи виявлення підробок та підозрілих операцій [7]. У подібному дослідженні також розглядаються можливості застосування машинного навчання для виявлення шахрайських дій на фондовому ринку та у фінансових транзакціях загалом [8].

Попри те, що в науковій літературі вже існує значна кількість оглядових досліджень, присвячених проблематиці фінансового шахрайства, більшість із них зосереджуються на окремих напрямах. Наприклад, у багатьох роботах розглядаються лише окремі випадки шахрайства, такі як махінації з кредитними картками, шахрайство в системах онлайн-банкінгу, зловживання під час видачі банківських кредитів або фальсифікація операцій із платіжними картками. Через таке вузьке спрямування оглядів зберігається потреба у дослідженні, яке б охоплювало всі поширені види фінансового шахрайства та дало б комплексне бачення проблеми.

Нещодавно було опубліковане дослідження, у якому розглядалися різні способи виявлення підозрілих дій у банківських операціях [10]. У ньому автори об’єднали результати попередніх досліджень, що стосуються фальсифікації фінансової звітності, які проводилися у різних галузях. Проте між цією працею та нашим оглядом є низка відмінностей. Зокрема, у згаданому дослідженні

основна увага приділялася залученню результатів із різних наукових напрямів, таких як бухгалтерський облік, аналітика та інформаційні системи.

У науковій роботі [9] було запропоновано систему під назвою FraudBuster, яка призначена для виявлення фінансового шахрайства, пов’язаного із крадіжками невеликих сум грошей протягом певного часу. Основна ідея підходу полягає у моделюванні поведінки користувача щодо витрат з часом. Система вважає транзакції підозрілими, якщо вони відхиляються від сформованої поведінкової моделі та змінюють характер витрат користувача. Таким чином, транзакція вважається шахрайською, якщо вона не відповідає типовому фінансовому профілю користувача.

На відміну від інших досліджень, які зосереджуються переважно на аналізі послідовності фінансових транзакцій [10], автори пропонованої моделі застосовують підхід, який враховує послідовність дій користувача перед здійсненням транзакції. Це дозволяє точніше оцінити, чи є конкретна транзакція легітимною. Такий підхід дав змогу сформулювати задачу виявлення шахрайства як задачу класифікації послідовностей, що, у свою чергу, потребує мінімальних зусиль щодо обробки ознак. Застосована методологія забезпечила високу ефективність розпізнавання шахрайських дій.

Результати дослідження показали, що запропонований підхід до обробки ознак не лише забезпечує конкурентоспроможність із сучасними методами, але й демонструє кращі результати у поєднанні з простими засобами виявлення аномалій. Оскільки використані ознаки не містять ідентифікаційної інформації, вдалося створити перший публічно доступний набір даних для задач виявлення шахрайства в інтернет-транзакціях.

В іншому дослідженні [11] для побудови моделі клієнтської поведінки було використано концепцію Recency, Frequency, Monetary (RFM) — нещодавність, частота та грошова вартість покупок. Такий підхід дозволив отримати інформативні ознаки для аналізу витрат користувачів. Це дослідження продемонструвало переваги грамотної інженерії ознак, навіть у поєднанні з базовими алгоритмами класифікації, такими як логістична регресія. Результати

експериментів підтвердили суттєве зростання точності класифікації завдяки вдало сформованим ознакам.

Крім того, автори зазначеного дослідження також дослідили вплив використання ознак, отриманих за допомогою методів виявлення аномалій. Виявилось, що додавання таких ознак сприяє покращенню результатів класифікації. У нашому підході ми також використовуємо ознаки аномалій, однак при оцінюванні їх ефективності за аналогічних умов дисбалансу даних наша методика демонструє ще кращі результати порівняно з наведеними підходами.

У роботі [12] запропоновано підхід, згідно з яким історія останніх транзакцій кожного окремого користувача розглядається як послідовність, що використовується для визначення достовірності кожної окремої операції. Для цього автори використали рекурентні нейронні мережі з механізмом уваги. Застосування механізму уваги дало змогу підвищити точність моделі та краще інтерпретувати результати класифікації. Для кодування метаданих кожної транзакції було використано декілька змінних, серед яких день тижня, година доби, сума платежу, а також ідентифікатор пристрою. У подальших кроках ці змінні перетворювались у векторні представлення — так звані ембеддинги. Оскільки транзакції оброблялись пакетно у вигляді послідовностей, модель змогла ефективно використовувати історичні дані різної довжини для кожного клієнта.

У дослідженні [13] було представлено модель виявлення шахрайства з використанням глибоких нейронних мереж у поєднанні з імовірнісними графічними моделями. Ця модель також базувалася на підході класифікації послідовностей. Для перевірки ефективності автори провели експеримент, використовуючи реальні транзакційні дані, та порівняли запропоновану модель з базовими методами. Результати показали, що врахування послідовних залежностей між транзакціями та передбаченими класами дозволяє досягти вищої точності. Додатково було запропоновано новий алгоритм undersampling,

який перевершив традиційні методи над- та недосемплювання при роботі з незбалансованими наборами даних [14].

1.2.1 Стан виявлення шахрайства

У постійно змінному середовищі цифрової злочинності важливо мати повне уявлення про основні типи шахрайства, а також про стратегії, які використовуються зловмисниками. Знання характерних ознак, які передують шахрайським діям або супроводжують їх, є важливою передумовою для побудови ефективних систем захисту.

Одним із поширених видів шахрайства є викрадення особистих даних. У цьому випадку зловмисники отримують доступ до особистої інформації користувача, щоб видавати себе за нього під час проведення фінансових операцій. Такі дії важко виявити за допомогою традиційних алгоритмів, проте методи машинного навчання, які аналізують поведінкові шаблони користувача, можуть виявити нетипові дії, що свідчать про можливе викрадення особистості.

Ще однією поширеною загрозою є шахрайство з кредитними картками, яке полягає у несанкціонованому використанні платіжної інформації для здійснення покупок або зняття коштів. Сучасні підходи до виявлення таких дій часто базуються на правилах, сформованих експертами, однак ці методи мають обмежену гнучкість і не завжди здатні вчасно реагувати на нові схеми обману. На відміну від них, алгоритми машинного навчання можуть виявляти приховані закономірності у фінансових транзакціях, що дозволяє вчасно виявити аномальні дії та запобігти шахрайству.

Таким чином, розуміння основних видів шахрайства та методів їх виявлення є критично важливим для побудови ефективних систем захисту. Надзвичайно важливо не лише фіксувати факт правопорушення, але й розпізнавати тонкі поведінкові патерни, що можуть свідчити про підготовку або реалізацію шахрайських дій.

1.2.2 Обмеження сучасних методів виявлення шахрайства

Традиційні підходи до виявлення шахрайства відіграють важливу роль у системах фінансової безпеки, однак вони мають низку істотних обмежень, які ускладнюють ефективну протидію сучасним кіберзагрозам.

Одним із ключових недоліків таких систем є їхня залежність від фіксованих наборів правил. Більшість традиційних рішень базуються на заздалегідь визначених умовах, які сигналізують про потенційно шахрайські дії. Проте зловмисники постійно вдосконалюють свої методи, адаптуються до нових умов та навмисно обходять існуючі правила. У результаті правила, що раніше були ефективними, з часом втрачають актуальність і не здатні виявити нові типи шахрайства. Це створює критичну проблему — втрата релевантності детекційних механізмів у динамічному середовищі загроз.

Ще однією суттєвою проблемою є висока ймовірність хибних спрацьовувань, що проявляється у вигляді помилкових позитивних та негативних результатів. У першому випадку система помилково визначає легітимну транзакцію як шахрайську, що може призвести до блокування рахунків, призупинення обслуговування клієнтів та втрати довіри. У другому — шахрайські операції можуть залишитися непоміченими, що веде до фінансових втрат та порушення безпеки. Високий рівень хибних спрацьовувань значно знижує ефективність системи виявлення шахрайства та ставить під сумнів її практичну цінність.

Крім того, традиційні системи часто не справляються з великими та різнорідними наборами даних. У сучасних умовах обсяги транзакційної інформації стрімко зростають, а дані характеризуються великою кількістю параметрів, різними джерелами походження та високою швидкістю надходження. Базові методи обробки інформації виявляються неспроможними забезпечити обробку транзакцій у реальному часі, що відкриває зловмисникам нові можливості для маніпуляцій. Відсутність здатності до масштабування та гнучкого реагування призводить до того, що організації залишаються вразливими до постійно змінюваних стратегій шахраїв [15].

У ряді досліджень розглядалися альтернативні підходи до проєктування систем виявлення шахрайства. Наприклад, у роботі [16] було розроблено концептуальну стратегію виявлення шахрайства на основі методу кейс-стаді. Об’єктом дослідження стала реальна ситуація в одному з ланцюгів постачання в країнах Азії, де виробник транспортних засобів зіткнувся з фактами шахрайства та неналежної поведінки співробітників. У ході дослідження було використано як якісні, так і кількісні дані для аналізу ситуації, що дозволило отримати глибше розуміння внутрішніх ризиків та схем зловживання.

Метод кейс-стаді є доцільним у тих випадках, коли необхідно дослідити конкретне середовище або обмежену галузь, що дозволяє зменшити витрати часу та ресурсів, а також зосередити увагу на деталях реального сценарію [17–19]. Подібні дослідження демонструють важливість контекстуального аналізу ризиків шахрайства, який часто не враховується при використанні виключно автоматизованих методів. Таким чином, ефективна система виявлення шахрайства повинна поєднувати автоматичні засоби аналізу з глибоким розумінням контексту, в якому відбуваються транзакції.

1.3 Постановка задачі дослідження

У зв’язку зі зростанням обсягів електронної комерції та активною цифровізацією фінансових операцій, постає гостра потреба у розробці ефективних методів виявлення шахрайських дій. Сучасні шахраї дедалі частіше використовують складні схеми, які важко виявити за допомогою традиційних методів, що базуються на фіксованих правилах чи ручному аналізі. Наявні системи виявлення шахрайства нерідко демонструють низьку ефективність у реальному середовищі, особливо у випадках з високим рівнем динаміки транзакцій та великою кількістю анонімних або обмежено структурованих даних.

Таким чином, виникає необхідність формалізації задачі розроблення інтелектуальної системи, яка була б здатна виявляти потенційно шахрайські транзакції з високим рівнем точності та адаптивності.

Метою кваліфікаційної роботи є побудова ефективної системи виявлення шахрайства в електронній комерції на основі методів машинного навчання, з урахуванням особливостей транзакційних даних, поведінкових патернів користувачів, а також характерного дисбалансу між шахрайськими та легітимними транзакціями.

Виходячи з поставленої мети, у межах дослідження сформульовано такі основні задачі:

– провести аналіз сучасних методів виявлення шахрайства в електронній комерції, охарактеризувати їх переваги та недоліки, виявити наукові та практичні прогалини;

– зібрати, дослідити та підготувати реальні транзакційні дані з маркетплейсів для побудови моделей машинного навчання;

– розробити концепцію інтелектуальної системи виявлення шахрайства, яка включає модулі попередньої обробки даних, генерації ознак, навчання моделей і оцінки результатів;

– реалізувати кілька моделей класифікації шахрайських транзакцій (зокрема, дерева рішень, випадкові ліси, градієнтний бустинг, ансамблеві методи), а також налаштувати гіперпараметри для досягнення максимальної продуктивності;

– визначити та реалізувати стратегії боротьби з дисбалансом класів, у тому числі методи надсемплювання та undersampling, і дослідити їх вплив на точність моделі;

– провести порівняльний аналіз ефективності різних підходів за допомогою обраних метрик якості (точність, повнота, F1-міра, AUC-ROC тощо).

Таким чином, формалізована задача полягає у побудові системи машинного навчання, яка здатна в умовах обмеженої та нерівномірної вибірки

транзакцій ефективно ідентифікувати шахрайські дії, адаптуючись до змін поведінки користувачів і особливостей ринку.

Предметна область дослідження охоплює аналіз цифрових транзакцій, побудову моделей класифікації, оцінку ефективності систем та врахування бізнес-контексту. Особливу роль відіграє обробка великих і різнорідних даних, вирішення проблеми дисбалансу класів і використання поведінкових ознак. Мультимодальний підхід, що поєднує числові, текстові та візуальні дані, відкриває нові можливості для підвищення точності систем виявлення шахрайства в електронній комерції.

2 ПРОЄКТУВАННЯ СИСТЕМИ ВИЯВЛЕННЯ ШАХРАЙСТВА В ЕЛЕКТРОННІЙ КОМЕРЦІЇ

2.1 Концепція побудови системи

У сучасних умовах стрімкого розвитку електронної комерції одним із найважливіших завдань є забезпечення довіри користувачів до торговельних платформ. Одним із ключових аспектів цього завдання є своєчасне виявлення шахрайських дій, що можуть завдавати значних фінансових збитків як користувачам, так і самим платформам. Для вирішення цієї проблеми необхідна побудова інтелектуальної системи, здатної автоматично розпізнавати ознаки шахрайства на основі транзакційної, поведінкової та профільної інформації користувачів.

Запропонована концепція побудови системи виявлення шахрайства базується на інтеграції класичних підходів до обробки даних із сучасними технологіями машинного навчання. Основна ідея полягає у тому, щоб використовувати попередні знання про поведінку користувачів, шаблони покупок, характеристики товарів і дії продавців для побудови адаптивної моделі, яка зможе виявляти відхилення від нормальної поведінки та класифікувати транзакції як потенційно шахрайські.

Формально, система складається з кількох функціональних рівнів:

1. Рівень збору та обробки даних — відповідає за отримання інформації з різних джерел: історія транзакцій, дії користувачів на платформі, атрибути товарів, відгуки, профілі продавців. На цьому етапі виконується попередня обробка, включаючи очищення, нормалізацію, обробку пропущених значень, кодування категоріальних змінних, масштабування числових ознак та видалення аномалій.

2. Рівень генерації ознак — виконує побудову нових ознак, що мають високу прогностичну здатність. Зокрема, використовуються поведінкові патерни користувача, частотність транзакцій, розбіжність між типовою сумою і

поточною, а також індикатори ризикованої активності (наприклад, багаторазові повернення, повторні спроби оплати тощо).

3. Аналітичний рівень — включає модулі машинного навчання, які здійснюють класифікацію транзакцій на основі навченої моделі. Концепція передбачає використання ансамблевих методів (наприклад, випадковий ліс, градієнтний бустинг, стекування моделей), які поєднують переваги окремих класифікаторів та забезпечують високу точність за умов нерівномірного розподілу класів.

4. Рівень балансування класів — розглядає проблему суттєвого дисбалансу між класами «шахрайство» та «нормальна транзакція». Для її вирішення передбачено використання таких методів, як SMOTE, SMOTENC, ENN та Tomek Links, які дозволяють синтетично збалансувати вибірку та зберегти репрезентативність моделі.

5. Рівень оцінки результатів — використовує розширений набір метрик, включаючи точність, повноту, F1-міру, AUC-ROC, а також спеціалізовані бізнес-метрики, які враховують потенційні втрати від хибнопозитивних та хибнонегативних рішень.

6. Рівень виведення та реакції — відповідає за формування звітів, позначення підозрілих транзакцій та, за потреби, автоматичне блокування або перенаправлення транзакції на ручну перевірку.

Ключова особливість запропонованої концепції полягає у гнучкості та модульності, що дозволяє адаптувати систему до особливостей конкретної торговельної платформи, структури даних і змін у поведінці шахраїв. Усі компоненти системи можуть бути масштабованими, а їхнє функціонування — оптимізованим у режимі реального часу.

2.2 Архітектура системи виявлення шахрайства

Запропонована архітектура інтелектуальної системи виявлення шахрайства включає чотири окремі експерименти (рисунок 2.1), кожен з яких

має на меті підвищення точності виявлення випадків організованого шахрайства у сфері роздрібної торгівлі.

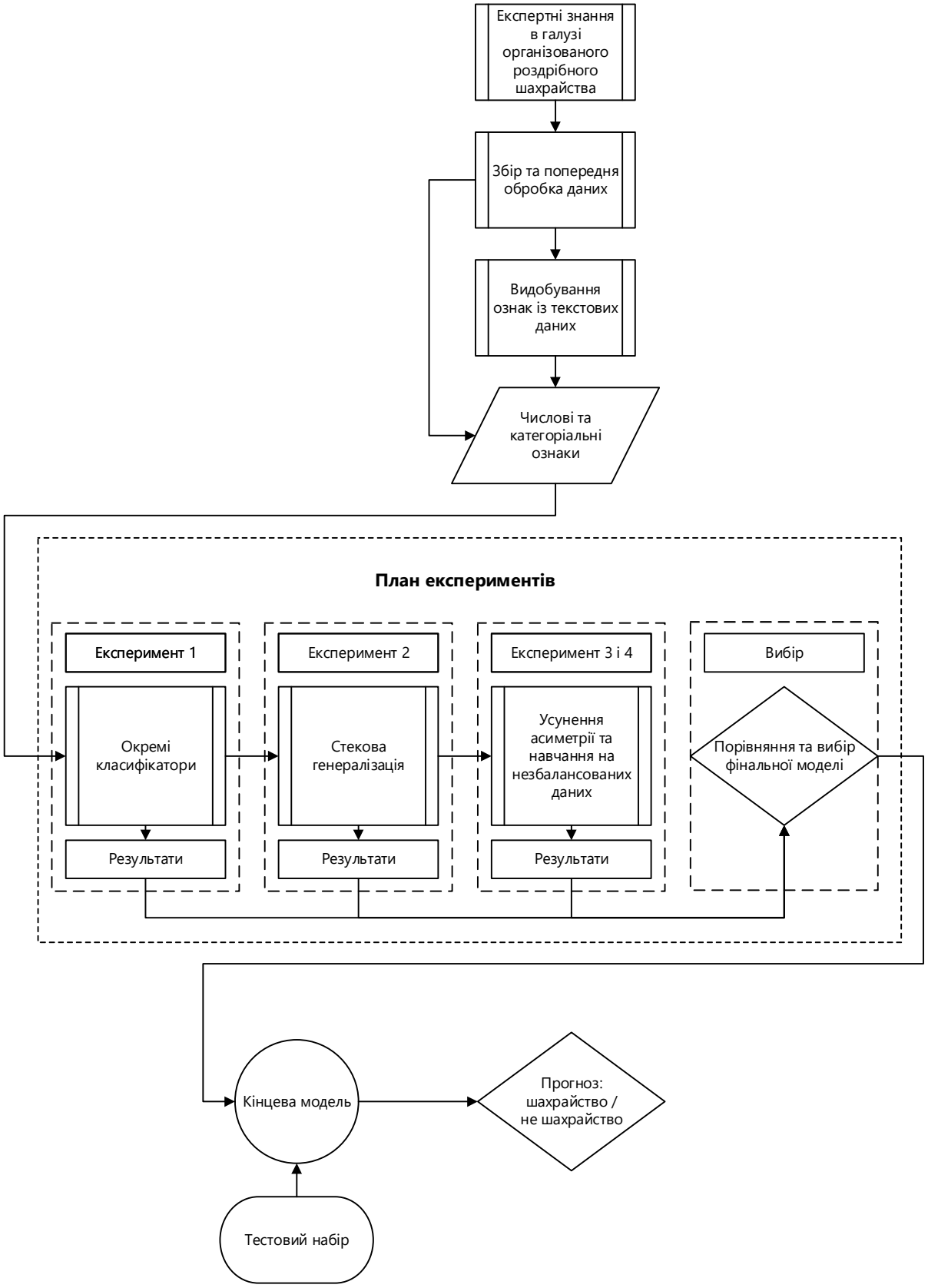

Рисунок 2.1 – Інформаційні потоки та дані, що використовуються у системі виявлення шахрайства для сфери роздрібної торгівлі

Комплексне проведення експериментів дозволяє встановити найбільш ефективну модель, яка забезпечує високий рівень достовірності у виявленні шахрайських транзакцій.

На рисунку 2.1 представлено загальну структуру потоків даних, що реалізується у рамках запропонованої системи. Ця схема відображає основні етапи обробки інформації, починаючи з надходження сирих даних та закінчуючи прийняттям рішення щодо наявності шахрайських ознак. Система адаптована до специфіки роздрібної торгівлі та може використовуватись для побудови стандартизованих рішень у сфері боротьби з шахрайством.

2.3 Методи та алгоритми машинного навчання для виявлення шахрайства в електронній комерції

2.3.1 Логістична регресія

У процесі побудови системи виявлення шахрайства в електронній комерції логістична регресія використовується як один із базових методів бінарної класифікації. Цей підхід дозволяє оцінити ймовірність того, що конкретна транзакція є шахрайською, на основі сукупності вхідних характеристик. Завдяки своїй простоті та інтерпретованості логістична регресія є зручною не лише для початкового моделювання, але й для подальшого аналізу впливу ознак на прийняття рішень.

Математична модель логістичної регресії базується на застосуванні сигмоїдальної (логістичної) функції до лінійної комбінації ознак:

$$P(y = 1 \mid X) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n)}}, \quad (2.1)$$

де $P(y = 1 \mid X)$ – це ймовірність того, що транзакція належить до класу шахрайських ($y = 1$);

$x_1, x_2, \dots, x_n$ – вхідні ознаки транзакції (наприклад, сума, частота, тип пристрою, час операції);

β_0 – вільний член (зсув);

$\beta_1, \beta_2, \beta_n$ – вагові коефіцієнти, які навчаються під час тренування моделі.

Функція перетворює необмежене значення лінійної комбінації ознак у значення в межах від 0 до 1, що інтерпретується як ймовірність шахрайства. Якщо ця ймовірність перевищує певне встановлене порогове значення (наприклад, 0.5), то транзакція вважається шахрайською.

Модель навчається на позначених (розмічених) даних, де для кожної транзакції відомо, чи була вона шахрайською. Навчання полягає у підборі коефіцієнтів β_i, які мінімізують логістичну втрату (log loss) – функцію, що штрафує за неправильні передбачення:

$$L(\beta) = -\frac{1}{m}\sum_{i=1}^{m}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)], \quad (2.2)$$

де m – кількість навчальних прикладів;

y_i – істинна мітка класу;

$\hat{y}_i$ – передбачена ймовірність, отримана з моделі.

Завдяки математичній прозорості та наявності коефіцієнтів, логістична регресія дозволяє аналітикам оцінити, які саме характеристики транзакцій найбільше впливають на ризик шахрайства. Наприклад, високий позитивний коефіцієнт при ознаці може вказувати на те, що її наявність значно підвищує ймовірність шахрайства.

Таким чином, логістична регресія є потужним і водночас простим інструментом для виявлення шахрайських транзакцій. Вона може використовуватись як самостійна модель у системах, що потребують пояснюваності, або як базовий компонент у складі складніших ансамблевих рішень. Її застосування дозволяє створити перший, інтерпретований прототип системи, на основі якого можуть будуватися більш складні моделі.

2.3.2 Дерева рішень

Застосування дерева рішень у задачі виявлення шахрайства в електронній комерції є ефективним підходом, що поєднує інтерпретованість, швидкість і

здатність працювати з різними типами даних. Дерево рішень використовується для побудови моделі, яка на основі послідовного розгалуження логічних умов визначає, чи є транзакція шахрайською чи ні. Цей метод особливо корисний у тих випадках, коли важливо мати чітке пояснення прийнятого рішення, що є актуальним у контексті фінансової безпеки та відповідальності перед користувачами платформи.

Побудова дерева рішень починається з вибору найінформативнішої ознаки, яка найкраще розділяє вхідні транзакції на «шахрайські» та «нормальні». Наприклад, першою умовою може бути перевірка, чи перевищує сума транзакції певний поріг. У разі позитивної відповіді дерево переходить до наступного вузла, де розглядається нова умова, наприклад: чи є розташування користувача незвичним або чи використовувався новий пристрій. Такий підхід дозволяє будувати ієрархію перевірок, що веде до кінцевого листка дерева, в якому міститься прогноз (шахрайство або ні).

З математичної точки зору, модель дерева рішень побудована на основі жадібного алгоритму, який на кожному кроці вибирає ознаку, що максимально зменшує невизначеність (ентропію) або мінімізує показник нечистоти (наприклад, критерій Джині або ентропійний приріст):

$$Information\ Gain = H(D) - \sum_{i=1}^{k} \frac{|D_i|}{|D|} \cdot H(D_i), \qquad (2.3)$$

де $H(D)$ – ентропія вихідної множини;

D_i – підмножина даних після розбиття за певною ознакою;

$|D|, |D_i|$ – кількість елементів у відповідних множинах.

Цей критерій дозволяє вибирати ті ознаки, які дають найбільший «виграш інформації» при кожному розгалуженні дерева.

Перевагою дерева рішень є його природна здатність працювати з категоріальними та числовими змінними, без необхідності масштабування чи перетворення. Крім того, дерево рішень легко візуалізується, а кожен шлях у

дереві можна інтерпретувати як набір умов, що призводить до класифікації транзакції як шахрайської.

У контексті електронної комерції дерева рішень можуть бути використані як базові класифікатори або як частина ансамблевих методів (наприклад, Random Forest або Gradient Boosting), які покращують загальну точність, зменшують ризик переобучення і краще працюють у складних умовах з нерівномірними класами.

Таким чином, дерева рішень є зручною та потужною моделлю, що дозволяє реалізувати систему виявлення шахрайства, орієнтовану на точність, інтерпретованість і адаптацію до змінних даних платформи. Вони ефективні в умовах, де потрібно приймати рішення у реальному часі та пояснювати їх як користувачам, так і адміністраторам системи.

2.3.3 Наївний баєсівський класифікатор

Застосування наївного баєсівського класифікатора у задачі виявлення шахрайства в електронній комерції є доцільним, коли йдеться про створення простої, швидкої та інтерпретованої моделі, здатної працювати з великими обсягами даних у реальному часі. Цей метод ґрунтується на теоремі Баєса, яка дозволяє обчислювати ймовірність належності транзакції до певного класу – шахрайського або нормального – за наявними ознаками, під припущенням незалежності між ознаками.

Математично модель виглядає наступним чином:

$$P(y \mid x_1, x_2, \dots, x_n) = \frac{P(y) \cdot \prod_{i=1}^{n} P(x_i|y)}{P(x_1, x_2, \dots, x_n)}, \quad (2.4)$$

де $P(y \mid x_1, x_2, \dots, x_n)$ – ймовірність того, що транзакція належить до класу y (наприклад, шахрайство) за умов спостережуваних ознак $x_1, \dots, x_n$;

$P(y)$ – апріорна ймовірність класу;

$P(x_i|y)$ – ймовірність спостереження ознаки x_i за умови, що транзакція належить до класу y;

$P(x_1, x_2, \dots, x_n)$ – є спільною ймовірністю, однаковою для всіх класів і не впливає на класифікацію.

Особливістю наївного баєсівського підходу є припущення про незалежність ознак, тобто модель вважає, що всі характеристики транзакції (сума, час, тип пристрою, місце розташування тощо) не впливають одна на одну. Хоча це припущення рідко виконується в реальності, на практиці класифікатор часто демонструє хорошу продуктивність завдяки простоті обчислень і здатності швидко реагувати на нові дані.

У задачі виявлення шахрайства на торгових платформах наївний баєсівський класифікатор можна використовувати для:

– первинного фільтрування підозрілих транзакцій;

– оцінки ймовірності шахрайства у режимі реального часу;

– розгортання легкої моделі на пристроях з обмеженими ресурсами (наприклад, у мобільних застосунках).

З технічного боку, для числових ознак часто використовується гаусівський наївний Баєс, де припускається, що розподіл кожної ознаки в межах класу є нормальним. Для категоріальних ознак застосовуються інші варіації: мультинормальний або бернуллівський баєсівський класифікатор, залежно від типу даних.

Недоліком є те, що порушення припущення незалежності між ознаками може знижувати точність моделі, особливо у складних задачах, де ознаки мають сильні кореляції. Проте навіть у таких умовах наївний Баєс часто слугує надійним базовим класифікатором для початкового аналізу або як частина ансамблю в поєднанні з більш складними моделями.

Таким чином, наївний баєсівський класифікатор є корисним інструментом для швидкого виявлення підозрілих транзакцій, побудови інтерпретованих моделей і формування первинної лінії захисту в системах безпеки електронної комерції.

2.3.4 Метод k-ближчих сусідів

Застосування методу k-ближчих сусідів (k-Nearest Neighbors, KNN) у задачі виявлення шахрайства в електронній комерції є доцільним у випадках, коли важливо порівнювати нову транзакцію з відомими прикладами з минулого. Метод належить до алгоритмів навчання без побудови моделі, оскільки він не створює узагальненої моделі під час навчання, а натомість приймає рішення на основі схожості між об’єктами.

У контексті виявлення шахрайства, KNN працює так: коли надходить нова транзакція, алгоритм обчислює відстані між цією транзакцією та всіма іншими транзакціями в навчальній вибірці, після чого вибирає k найближчих сусідів (тобто найбільш схожих транзакцій). На основі міток класів цих сусідів (наприклад, «шахрайство» або «нормальна транзакція») здійснюється голосування, і визначається клас нової транзакції. Якщо більшість із сусідів позначені як шахрайські, то система також класифікує нову транзакцію як шахрайську.

Найчастіше для обчислення схожості використовується евклідова відстань:

$$d(x, x') = \sqrt{\sum_{i=1}^{n}(x_i - x_i')^2}, \tag{2.5}$$

де x – новий об'єкт;

x' – об'єкт з навчальної вибірки;

n – кількість ознак.

Метод KNN має кілька переваг:

– він простий у реалізації та інтуїтивно зрозумілий;

– він не вимагає складної процедури навчання;

– він добре працює з локальними аномаліями, які часто характерні для шахрайських дій.

Проте в задачі виявлення шахрайства є і певні виклики:

– KNN є обчислювально затратним при великих обсягах даних, оскільки для кожного нового прикладу потрібно обчислювати відстань до всіх інших прикладів у базі;

– алгоритм є чутливим до масштабування ознак, тому перед його застосуванням необхідно здійснити нормалізацію або стандартизацію числових даних;

– при високій розмірності даних (багато ознак) ефективність KNN може знижуватись — це явище називається прокляттям розмірності.

У рамках системи виявлення шахрайства в електронній комерції, метод k-ближчих сусідів може бути використаний як:

– базовий класифікатор для порівняння з іншими моделями;

– підсистема локального аналізу при підозрілих транзакціях;

– частина ансамблевої моделі, зокрема в стекованій архітектурі, де KNN надає одну з оцінок, яка враховується метамоделлю.

Таким чином, метод k-ближчих сусідів дозволяє виявляти потенційно шахрайські дії, орієнтуючись на подібність транзакцій, і є корисним інструментом у системах, де важливу роль відіграє аналіз поведінкових патернів і структурна близькість даних.

2.3.5 Метод опорних векторів

Застосування методу опорних векторів (Support Vector Machine, SVM) у задачі виявлення шахрайства в електронній комерції є ефективним, особливо коли потрібно відокремити шахрайські транзакції від нормальних за допомогою чіткої межі. Цей метод ґрунтується на побудові гіперплощини, яка максимально розділяє дві категорії даних у багатовимірному просторі ознак.

Основна ідея методу полягає в тому, щоб знайти таку межу, яка максимально віддалена від найближчих прикладів кожного класу (їх називають опорними векторами). Ця межа дозволяє SVM не лише точно класифікувати відомі транзакції, а й з високою ймовірністю правильно розпізнавати нові.

Математична модель методу опорних векторів у найпростішому лінійному випадку формулюється як задача оптимізації:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \quad \text{за умови:} \quad y_i(w \cdot x_i + b) \geq 1, \tag{2.6}$$

де w – вектор ваг (нормаль до гіперплощини);

b – зсув гіперплощини;

x_i – вхідний вектор ознак для транзакції;

$y_i \in \{-1, +1\}$ – клас транзакції (шахрайство або ні).

У разі, коли дані не є лінійно роздільними, застосовують ядрові функції (наприклад, RBF – радіально-базисну функцію), які дозволяють перенести дані у вищий простір ознак, де вже можна побудувати розділюючу гіперплощину.

SVM добре працює в задачах виявлення шахрайства завдяки таким властивостям:

– висока точність навіть при складних, нелінійних залежностях між ознаками;

– стійкість до переобучення, особливо при правильному налаштуванні параметра регуляризації;

– підтримка невеликої вибірки за рахунок використання лише опорних векторів для побудови моделі.

Проте метод має і певні обмеження:

– погано масштабується на великі набори даних (SVM може бути повільним при десятках тисяч транзакцій);

– чутливий до дисбалансу класів, тому в задачі виявлення шахрайства рекомендовано поєднувати його з методами вирівнювання класів, наприклад SMOTE або Tomek Links;

– важко інтерпретується у порівнянні з деревами рішень або логістичною регресією.

У контексті системи виявлення шахрайства в електронній комерції SVM може використовуватись:

– як основна класифікаційна модель, якщо дані мають складну структуру;

– у поєднанні з іншими алгоритмами в рамках ансамблевих або стекованих моделей;

– для виявлення аномалій, якщо використовувати одно-класовий варіант SVM (One-Class SVM).

Таким чином, метод опорних векторів є потужним інструментом для побудови високоточних моделей виявлення шахрайства, які здатні працювати з як лінійними, так і з нелінійними залежностями, забезпечуючи надійність і точність систем безпеки електронної комерції.

2.3.6 Випадковий ліс

Застосування методу «випадковий ліс» (Random Forest) у задачі виявлення шахрайства в електронній комерції є одним із найбільш ефективних підходів, що поєднує високу точність, стійкість до переобучення та здатність працювати з різнорідними даними. Цей метод належить до ансамблевих алгоритмів машинного навчання і базується на об'єднанні великої кількості дерев рішень, кожне з яких навчається на випадковій частині даних та випадковому підмножинному наборі ознак.

Принцип роботи полягає в тому, що для кожної нової транзакції кожне дерево у «лісі» робить свій прогноз (шахрайство чи ні), після чого застосовується голосування більшості, і на основі його результатів видається фінальне рішення. Такий підхід дозволяє уникнути надмірної чутливості до окремих ознак або шумових даних, що є поширеними у транзакційній інформації.

З математичної точки зору, метод працює за таким алгоритмом:

1. Із вихідного навчального набору D випадково створюються підмножини даних $D_1, D_2, \ldots, D_t$ за методом bootstrap sampling.

2. Для кожної підмножини будується дерево рішень, де на кожному вузлі випадковим чином вибирається підмножина ознак, і з них обирається найкраща для поділу.

3. Після побудови всіх дерев прогнозування нової транзакції відбувається за принципом множинного голосування.

Метод має низку вагомих переваг у контексті виявлення шахрайства:

– висока точність класифікації навіть у складних умовах із великим числом ознак;

– ефективна робота з незбалансованими даними, що часто зустрічаються у шахрайських вибірках;

– можливість автоматично оцінювати важливість ознак, що дозволяє зрозуміти, які саме параметри транзакцій найбільше впливають на ризик шахрайства;

– висока стійкість до перенавчання завдяки усередненню великої кількості дерев.

Система виявлення шахрайства в електронній комерції може використовувати випадковий ліс:

– як основний класифікатор, що забезпечує баланс між точністю і швидкістю роботи;

– у складі ансамблевої або стекованої архітектури, де Random Forest виступає як одна з базових моделей;

– для аналізу важливості ознак, з метою виявлення нових потенційних індикаторів шахрайства;

– як бенчмарк для порівняння з більш складними або нейронними моделями.

Під час реалізації системи важливо налаштовувати такі параметри як кількість дерев (наприклад, 100–500), максимальна глибина дерева, мінімальна кількість зразків на листку, щоб досягти оптимального співвідношення між продуктивністю та точністю.

Таким чином, метод випадкового лісу є надійним і потужним інструментом для побудови високоточних систем виявлення шахрайських транзакцій, здатних працювати в умовах великого обсягу даних, складної структури ознак та змінної поведінки користувачів.

2.3.7 Градієнтний бустинг

Застосування методу градієнтного бустингу (Gradient Boosting) у задачі виявлення шахрайства в електронній комерції є одним із найефективніших підходів до побудови високоточних класифікаційних моделей. Цей метод належить до класу ансамблевих методів машинного навчання, які поєднують багато слабких моделей (як правило, дерев рішень) у потужний класифікатор шляхом послідовного навчання з урахуванням помилок попередніх моделей.

Основна ідея полягає в тому, що кожне нове дерево навчається не на вихідних мітках класу, а на залишкових помилках попередніх дерев, тобто на тому, що попередні моделі не змогли правильно передбачити. Таким чином, бустинг поступово "виправляє" помилки, зменшуючи функцію втрат і підвищуючи точність моделі на кожному кроці.

З математичної точки зору, градієнтний бустинг ітеративно оптимізує довільну функцію втрат:

$$L(y, F(x)), \tag{2.7}$$

де y – справжнє значення мітки (наприклад, 0 для легітимної транзакції, 1 – для шахрайської);

$F(x)$ – сума передбачень усіх попередніх моделей/

На кожному кроці додається нове дерево $h_m(x)$, яке апроксимує градієнт похідної функції втрат:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \tag{2.8}$$

де γ_m – коефіцієнт навчання, що регулює внесок нового дерева в модель.

У контексті виявлення шахрайства в електронній комерції, градієнтний бустинг має такі переваги:

– висока точність: метод ефективно навчається навіть на складних вибірках із великою кількістю ознак і взаємозв’язків;

– стійкість до нерівномірного розподілу класів: за умови правильного налаштування функції втрат і ваг класів;

– гнучкість: дозволяє налаштовувати тип функції втрат, кількість дерев, глибину, швидкість навчання, що робить його придатним для адаптації до будь-якої структури транзакційних даних;

– оцінка важливості ознак: градієнтний бустинг дозволяє визначити, які саме характеристики (сума, час, країна, спосіб оплати, історія користувача) найсильніше впливають на класифікацію транзакції.

Найпопулярнішими реалізаціями цього методу є XGBoost, LightGBM та CatBoost, які широко застосовуються в реальних системах безпеки онлайн-платформ завдяки своїй швидкодії, підтримці пропущених значень і роботі з категоріальними змінними.

Градієнтний бустинг ефективно використовується в системі виявлення шахрайства:

– як основний класифікатор завдяки найвищій точності серед більшості моделей;

– у стекованих ансамблях, де він виступає як метамодель або один із потужних базових класифікаторів;

– для аналізу важливості ознак і побудови довірених прогнозів, що можуть пояснюватися за допомогою інтерпретаторів, таких як SHAP.

Таким чином, метод градієнтного бустингу є надзвичайно потужним і гнучким інструментом для виявлення шахрайства в електронній комерції, здатним до точного моделювання складних закономірностей у великих наборах даних і забезпечення високого рівня захисту користувачів та бізнесу.

3 ТЕСТУВАННЯ СИСТЕМИ ВИЯВЛЕННЯ ШАХРАЙСТВА В ЕЛЕКТРОННІЙ КОМЕРЦІЇ

3.1 Постановка експериментальних досліджень

3.1.1 План проведення експериментів

Перший експеримент передбачає індивідуальне тестування низки базових моделей класифікації. На початковому етапі виконується попередня обробка даних та виділення ознак, у тому числі за допомогою числових метрик. Ці етапи мають вирішальне значення для ефективності роботи класифікаторів. Детальну структуру обробки можна переглянути на рисунку 2.1.

У рамках цього експерименту було навчено сім різних класифікаторів, відібраних на основі аналізу літературних джерел. При цьому не застосовувалися жодні методи усунення асиметрії класів (наприклад, oversampling або undersampling). Для налаштування параметрів кожної моделі використовувався метод графічного перебору у поєднанні з багаторазовою стратифікованою k-кратною перехресною перевіркою. Стратифікація дозволяє зберегти збалансоване співвідношення класів у кожній підвибірці, що особливо важливо при наявності дисбалансу між кількістю нормальних та шахрайських транзакцій.

У другому експерименті реалізовано більш складну модель, яка об'єднує всі сім класифікаторів із першого експерименту в єдину ансамблеву архітектуру, побудовану за принципом стекування. Цей підхід дозволяє отримати більш точні прогнози завдяки комбінації сильних сторін кожного окремого класифікатора (архітектуру представлено на рисунку 3.1).

Суть стекування полягає в тому, що кожен базовий класифікатор генерує прогноз, а спеціальна мета-модель використовує всі ці прогнози як вхідні ознаки для формування остаточного рішення. У процесі навчання мета-моделі використовується відкладена вибірка, яка дозволяє уникнути переобучення та зберегти незалежність між навчальними даними базових моделей і даними, що надходять на вхід мета-рівню.

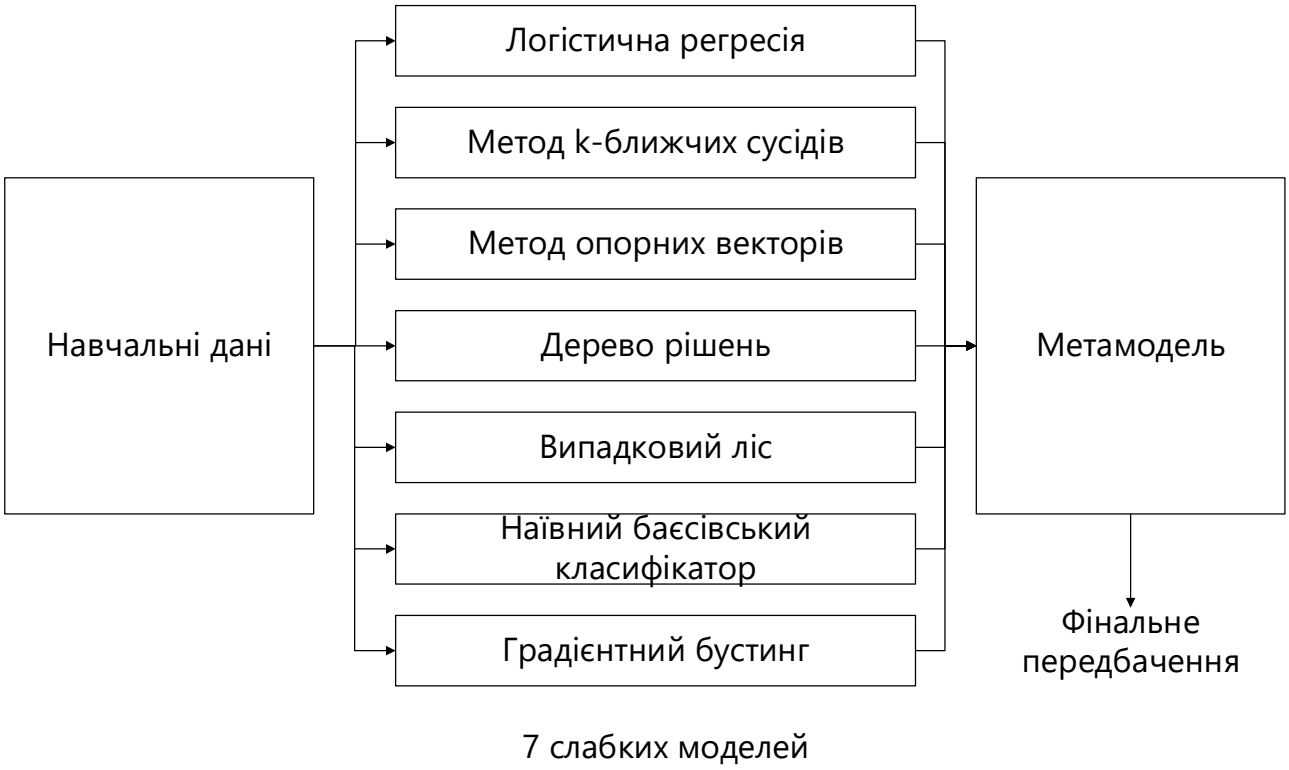

Рисунок 3.1 – Застосування методу стекування із використанням семи класифікаторів, кожен з яких виконує роль слабкого навчального алгоритму

Для формування таких вибірок застосовувався k-кратний перехресний аналіз з використанням out-of-fold прогнозів, тобто прогнозів, зроблених на тих даних, які не входили у навчальну частину кожної підмоделі. Це дає змогу отримати якісні дані для мета-моделі та забезпечити високу узагальнювальну здатність у реальних умовах.

Таким чином, запропонований підхід дозволяє поєднати класичні методи класифікації із сучасними ансамблевими техніками, такими як bagging (об’єднання моделей через повторну вибірку) та boosting (послідовне посилення слабких моделей), що сприяє значному підвищенню точності виявлення шахрайських дій.

Оскільки в даних, пов’язаних із виявленням шахрайства, спостерігається природжений дисбаланс між кількістю шахрайських і не шахрайських транзакцій, третій і четвертий експерименти були спрямовані на усунення цієї проблеми. Значна перевага одного класу над іншим часто призводить до того, що

класифікатор навчається розпізнавати лише домінуючий клас, ігноруючи менш представлений, який є найбільш критичним у контексті виявлення шахрайства.

У рамках цих експериментів було повторно реалізовано процедури з експериментів 1 та 2 (див. рисунок 2.1), але з урахуванням методів балансування класів, які попередньо були відібрані на основі аналізу характеристик набору даних. Основна мета полягала в тому, щоб визначити найкращу комбінацію методу балансування та класифікатора, яка забезпечить максимальну ефективність виявлення шахрайських операцій у конкретному контексті.

Застосовувані підходи до вирішення проблеми дисбалансу включали як методи недосемплювання (undersampling), так і надсемплювання (oversampling). Наприклад, використовувалися такі техніки, як SMOTE (Synthetic Minority Oversampling Technique), Tomek Links та Random Undersampling. Результати цих експериментів дозволили виявити, які саме комбінації моделей і методів ребалансування забезпечують найкращі показники точності, повноти та F1-міри у задачі виявлення організованого шахрайства.

3.1.2 Експериментальні дані

У цьому параграфі описуються методики та параметри, які були використані у серії проведених експериментів. Вибрані класифікатори включають як базові моделі (логістична регресія, дерево рішень, наївний баєсівський класифікатор), так і просунуті алгоритми машинного навчання (градієнтний бустинг, випадковий ліс, XGBoost, нейронні мережі). Кожна з моделей була налаштована із використанням підбору гіперпараметрів за допомогою Grid Search у комбінації з стратифікованою *k*-кратною перехресною перевіркою.

Для оцінки якості моделей використовувались метрики, чутливі до дисбалансу класів, зокрема точність (accuracy), повнота (recall), специфічність, точність позитивного прогнозу (precision), F1-міра та площа під ROC-кривою (AUC-ROC). Завдяки цьому забезпечено комплексну оцінку ефективності

моделей при роботі з даними, де шахрайські випадки трапляються значно рідше, ніж нормальні транзакції.

Для реалізації та тестування системи виявлення шахрайства було використано реальні транзакційні та поведінкові дані одного з провідних міжнародних інтернет-маркетплейсів. Метою аналізу було виявлення організованих шахрайських схем у роздрібній торгівлі (Organized Retail Crime, ORC).

Зважаючи на обмеження у маркуванні даних, до вибірки було включено 3606 виробників, зареєстрованих у США, з якими проводилося пілотне тестування. Основну частину інформації становили атрибути продавців (тип облікового запису, історія реєстрації, обсяг торгівлі) та деталі товарів (категорія, ціна, рейтинг, популярність).

Щоб зменшити вплив нерепрезентативних сплесків активності, фокус дослідження було обмежено на продавцях із високою активністю, яких було ідентифіковано як топ-продавців за останні 90 днів. Такий підхід дозволив сформувати стабільну базу даних, у якій відображаються закономірності, характерні для великих обсягів торгівлі, та забезпечити надійність побудованої моделі в реальних умовах.

3.2 Результати експериментальних досліджень

Для швидкого обчислювального експерименту було здійснено стратифіковану випадкову вибірку, у рамках якої випадковим чином відібрано 50 000 рядків із набору даних маркетплейсу. Стратифікація забезпечила справедливе представлення усіх підгруп у вибірці, що дозволило уникнути статистичної упередженості при аналізі результатів.

Основна увага була зосереджена на побудові моделі виявлення шахрайства на основі числових та категоріальних ознак. Спершу ці ознаки були ідентифіковані у структурі записів оголошень, після чого оцінювалась їх

відповідність демографічним характеристикам, поведінці та звичкам користувачів торгової платформи.

Для обробки даних була створена окрема конвеєрна система (pipeline), яка відповідала за:

– кодування категоріальних змінних;
– масштабування числових ознак;
– обробку пропущених значень;
– видалення аномалій і дублікатів;
– запуск експериментів 1 та 2.

У рамках експериментів 3 та 4 до конвеєру було додано модуль вирівнювання класів, який застосовував методи надсемплювання (oversampling) та недосемплювання (undersampling) з метою усунення дисбалансу між шахрайськими та нормальними транзакціями.

Серед застосованих методів балансування даних варто відзначити:

– SMOTENC — модифікований SMOTE, адаптований для категоріальних змінних;
– SMOTENC + ENN — поєднання синтетичного надсемплювання з видаленням шумових точок;
– SMOTENC + Tomek Links — комбінація надсемплювання із видаленням перетинаючихся пар сусідніх точок різних класів.

На навчальній вибірці (in-sample) всі три підходи досягли показника точності 90%, однак на тестовій вибірці (out-of-sample) результати впали приблизно до 55%, що свідчить про потенційне перенавчання або чутливість до нерепрезентативних даних.

Після порівняння різних комбінацій класифікаторів і методів балансування даних за всіма метриками, випадковий ліс (Random Forest) продемонстрував найкращі результати. На другому місці за загальною ефективністю опинився класифікатор стекування (Stacked Generalization). Інші моделі не показали значущого покращення незалежно від обраних методів балансування.

На рисунку 3.2 наведено детальну візуалізацію покращення продуктивності кожного класифікатора в умовах застосування різних стратегій балансування класів.

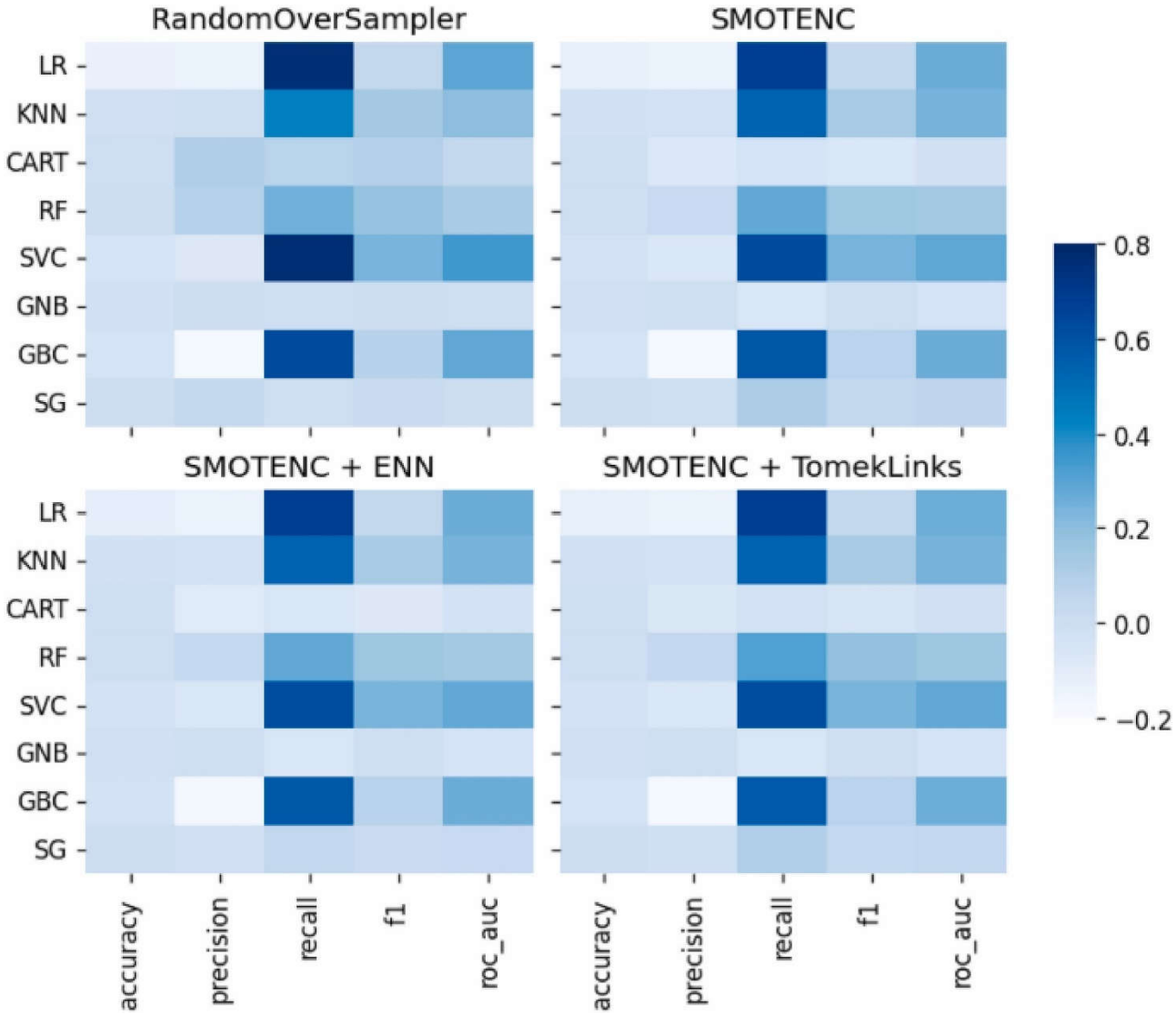

Рисунок 3.2 – Покращення продуктивності класифікаторів на тестовій вибірці при застосуванні методів балансування класів на рівні даних

Окрім оцінки точності моделей, було проведено аналіз попиту за днями тижня, результати якого наведено у рисунку 3.3 у вигляді теплової карти. Згідно з результатами, найбільший попит на продукцію інтернет-магазину спостерігався у період з грудня по травень, причому пікові значення були зафіксовані на початку квітня. Такі результати дозволяють компаніям

оптимізувати маркетингові кампанії та стимули для покупців, щоб вирівняти навантаження на торговельні системи впродовж року.

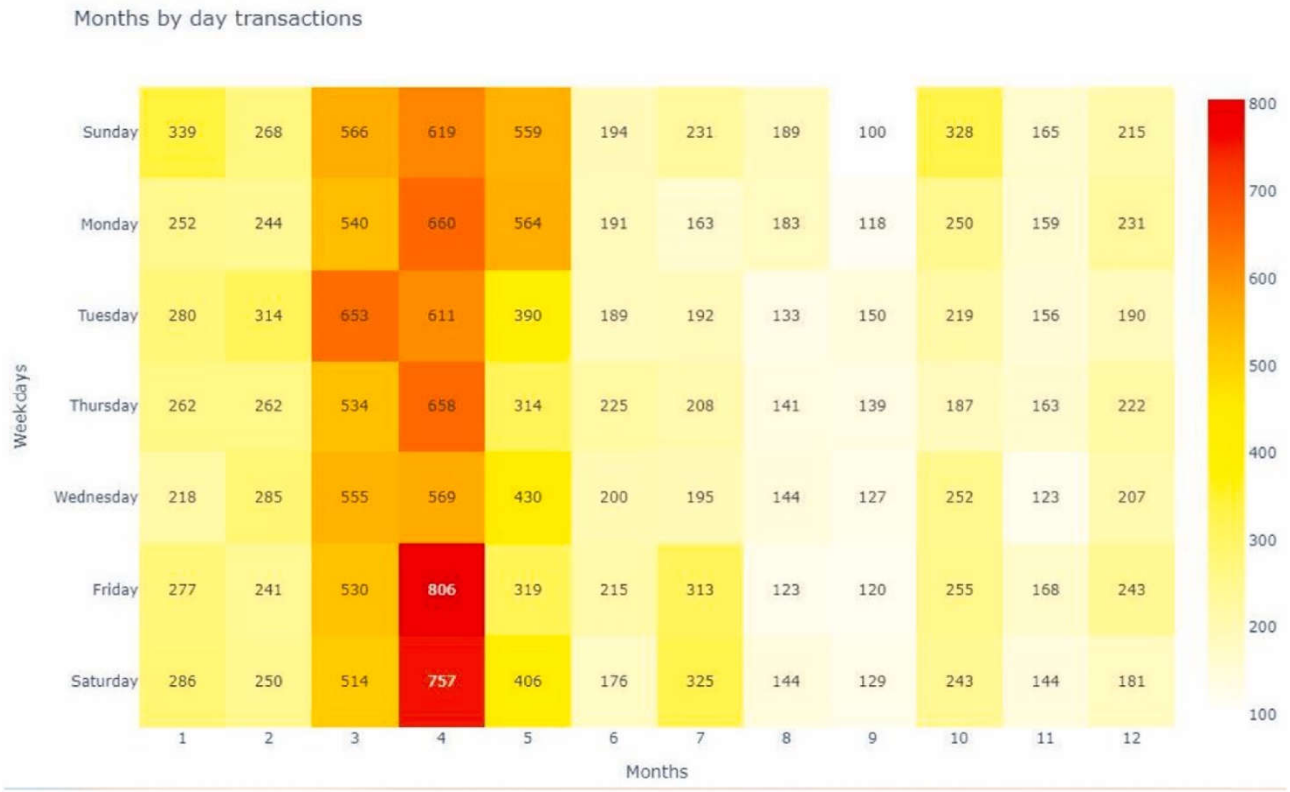

Рисунок 3.3 – Теплова карта транзакцій для одного з інтернет-магазинів за днями тижня протягом місяця

Також було визначено топ-10 найважливіших ознак, які найбільше впливають на здатність моделі передбачати шахрайські дії. Для оцінки важливості ознак було використано SHAP-аналіз (SHapley Additive exPlanations), який дозволяє інтерпретувати внесок кожної змінної у рішення моделі. Високі додатні значення SHAP вказують на ознаки, що підвищують ймовірність класифікації транзакції як шахрайської, у той час як від'ємні значення SHAP свідчать про ознаки, які притаманні легітимним транзакціям. Відповідну візуалізацію представлено на рисунку 3.4.

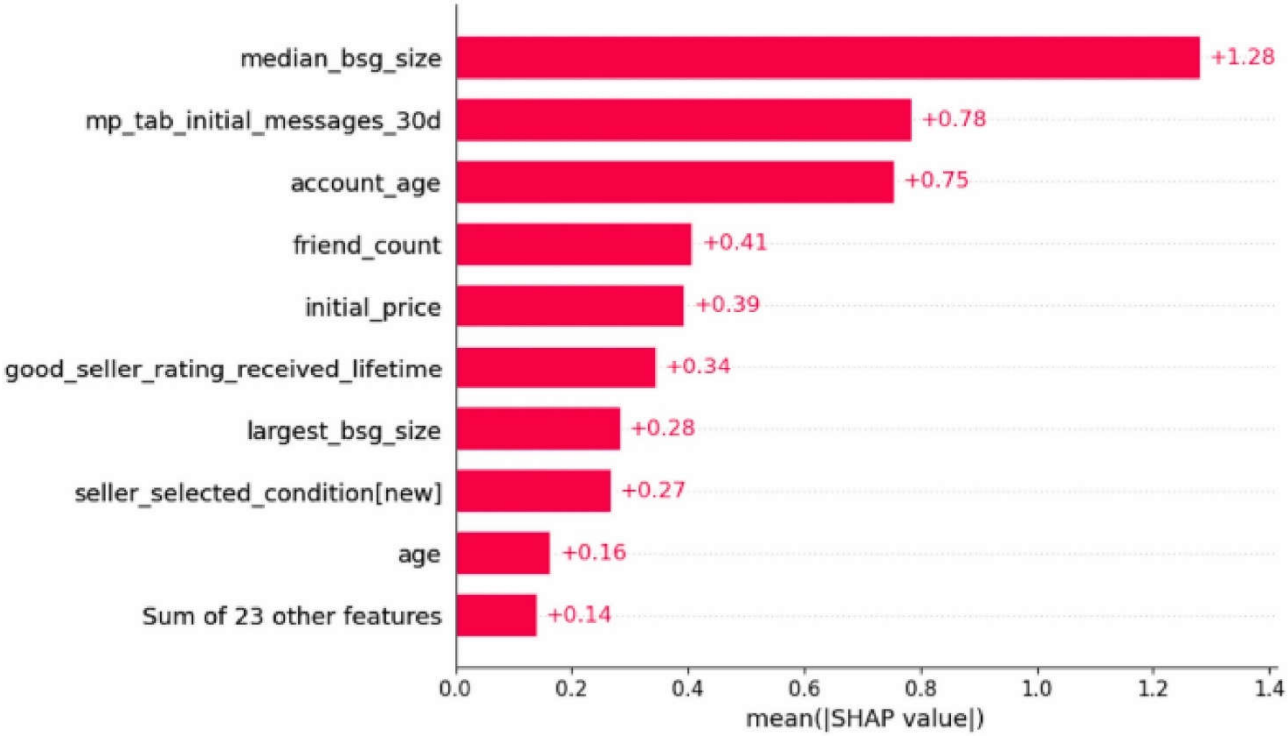

Рисунок 3.4 – Вплив ключових ознак на рішення моделі щодо класифікації транзакції як шахрайської або легітимної

3.3 Напрями подальших досліджень

Багато сучасних інтернет-маркетплейсів, зокрема Marketplace компанії Meta та платформа eBay, протягом тривалого часу стикаються із серйозними загрозами кібербезпеки, пов’язаними з організованими шахрайськими схемами у сфері роздрібної торгівлі. Із розширенням обсягів даних про поведінку користувачів, їхні звички, покупки та особисті характеристики, традиційні методи виявлення шахрайства — зокрема фільтрація за ключовими словами, ручний пошук чи застосування фіксованих правил – стають усе менш ефективними та складними в реалізації.

У зв’язку з цим у даній роботі було запропоновано автоматизований підхід до виявлення шахрайських дій в електронній комерції, що базується на методах машинного навчання з учителем. Проведені експерименти показали, що

запропонована система демонструє вищу точність і продуктивність у порівнянні з попередніми підходами, які використовували лише правила або методи без учителя. Слід зазначити, що подібна модель раніше не застосовувалась у контексті організованого роздрібного шахрайства (ORC), а в тих поодиноких випадках, де вона згадувалась, зазвичай обмежувалась лише вирішенням проблеми дисбалансу або застосуванням одноетапної обробки даних.

Наш підхід базувався на поєднанні інформативного відбору ознак, що враховує експертну думку, з індивідуально налаштованими стратегіями обробки даних, використанням методів балансування класів, оптимізацією вибору моделей і гіперпараметрів, а також орієнтацією на бізнес-орієнтовані метрики оцінювання ефективності. Усі ці компоненти дозволили досягти рівня точності, що відповідає сучасним вимогам у галузі виявлення шахрайства.

Основу експериментального набору ознак становили числові та категоріальні змінні, які були отримані з транзакційної та профільної інформації користувачів. Проте для подальшого підвищення якості моделі можливим напрямом удосконалення є впровадження мультимодального підходу, який включатиме числові, текстові та візуальні дані. Наприклад, додатковий аналіз описів товарів, зображень продуктів та текстів відгуків може суттєво покращити розпізнавання шахрайських шаблонів без необхідності глибокої участі галузевих експертів.

Підсумовуючи, запропонована система має потенціал до широкого впровадження в електронній комерції для виявлення організованого шахрайства. Подальші дослідження можуть бути спрямовані на:

– інтеграцію нових типів даних для навчання моделей;

– розвиток методів самонавчання або напівавтоматичного оновлення моделей;

– розширення застосування системи на інші домени (логістика, послуги, цифрові підписки тощо);

– дослідження етичних і правових аспектів впровадження подібних інтелектуальних рішень у комерційних середовищах.

Запропонована концепція створює передумови для формування гнучких, адаптивних та інтерпретованих систем виявлення шахрайства, здатних ефективно реагувати на нові загрози в умовах постійно змінюваного середовища цифрової торгівлі.

ВИСНОВКИ

1. Електронна комерція є одним із ключових напрямів цифрової економіки, але водночас виступає вразливим середовищем для реалізації шахрайських схем. З розвитком онлайн-торгівлі зростає кількість та складність шахрайських дій, які негативно впливають на безпеку, довіру споживачів та фінансову стабільність платформи. Особливу загрозу становить організоване роздрібне шахрайство, що потребує системного та автоматизованого підходу до виявлення.

2. Існуючі системи виявлення шахрайства здебільшого базуються на фіксованих правилах або ручному аналізі, що не забезпечує достатньої гнучкості та адаптивності. Методи штучного інтелекту, зокрема машинного навчання, демонструють вищу здатність до виявлення прихованих закономірностей у транзакційних даних. Вони дозволяють створювати адаптивні моделі, здатні виявляти як типові, так і нові форми шахрайства в умовах постійних змін.

3. Запропонована концепція побудови системи виявлення шахрайства орієнтована на модульну, масштабовану та гнучку архітектуру, яка дозволяє адаптуватися до особливостей конкретної платформи електронної комерції. Система охоплює всі ключові етапи — від збору та обробки даних до навчання моделей та автоматичної реакції на підозрілі дії. Завдяки використанню поведінкових ознак, адаптивних алгоритмів і бізнес-метрик забезпечується висока ефективність і релевантність прийнятих рішень.

4. Архітектура системи реалізує послідовне проведення експериментів для визначення оптимальних підходів до класифікації шахрайських транзакцій, зокрема із використанням методів балансування класів та ансамблевих моделей. Такий підхід дає змогу досягти точного виявлення шахрайства навіть за умов суттєвого дисбалансу даних. Система розроблена з урахуванням потреб роздрібного сектору, що забезпечує її прикладну цінність та можливість впровадження в реальні торговельні середовища.

5. Проаналізовано ключові алгоритми машинного навчання, які застосовуються для виявлення шахрайських транзакцій в електронній комерції.

Кожен із розглянутих методів – логістична регресія, дерева рішень, наївний баєсівський класифікатор, метод *k*-ближчих сусідів, метод опорних векторів, випадковий ліс і градієнтний бустинг – має свої переваги та обмеження, які необхідно враховувати під час розробки системи. Зокрема, моделі відрізняються рівнем інтерпретованості, чутливістю до дисбалансу класів, швидкістю обробки даних та адаптивністю до складної структури ознак.

6. У ході експериментальних досліджень було реалізовано поетапну архітектуру тестування, що включала базові моделі, стековані ансамблі та комбінації з методами балансування класів. Отримані результати підтвердили ефективність стекування як стратегії підвищення точності виявлення шахрайства, особливо в умовах сильної асиметрії даних. Випадковий ліс і стекований класифікатор виявилися найкращими за всіма основними метриками, демонструючи потенціал для практичного впровадження.

7. Для досягнення високої продуктивності системи було використано реальні дані з онлайн-маркетплейсу, а також сучасні стратегії передобробки, включно з SMOTENC, Tomek Links та іншими техніками балансування. Завдяки застосуванню SHAP-аналізу вдалося інтерпретувати вагомість ключових ознак, що підвищує довіру до моделі та її адаптованість до бізнес-середовища. Запропонована система є гнучкою до змін у структурі даних і може бути вдосконалена шляхом інтеграції мультимодальних джерел – текстових, візуальних і поведінкових.