

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА**

**Економічний факультет
Кафедра економічної кібернетики**

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

**«Виявлення шахрайських дій на ринку нерухомості за допомогою моделей
машинного навчання»**

студентки 4 курсу
спеціальності 051 «Економіка»
ОПП «Економічна кібернетика»
денної форми навчання
Дікої Данії Василівни

Науковий керівник:
доктор економічних наук,
професор
Чорноус Галина Олександрівна

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

Студент _____
(підпис)

Роботу допущено до захисту перед ЕК
рішенням кафедри економічної кібернетики
від 12 червня 2023 р., протокол № 17

Завідувач кафедри:
доктор економічних наук, професор
Ляшенко Олена Ігорівна

(підпис)

РЕФЕРАТ

Кваліфікаційна робота бакалавра містить: 70 ст.(без додатків), 18 рис., 7 табл., 21 формул, 52 джерела

Ключові слова: ринок нерухомості, шахрайські дії, моделі машинного навчання, нейромережі, виявлення шахрайства, інформаційно-аналітичні системи, кібербезпека, прогнозування шахрайства, автоматизовані системи управління.

Об'єкт дослідження: процеси управління на ринку нерухомості.

Мета дослідження: розробка методики виявлення шахрайських дій на ринку нерухомості за допомогою методів машинного навчання.

Методи дослідження: логічний, діалектичний, історичний та порівняльний методи для обґрунтування наукових засад та вдосконалення понятійного апарату дослідження; методи спостереження, узагальнення, абстрагування, формалізації, аналізу та синтезу для характеристики методичних аспектів створення інформаційно-аналітичної системи; програмно-цільовий метод з метою обґрунтування механізмів розробки та реалізації програм інформатизації аналітичних процесів.

Практична цінність: рекомендації щодо розробки нейромережевої системи, яка дозволить спростити та значно покращити процес прогнозування шахрайських дій на ринку нерухомості.

ABSTRACT

The Bachelor's qualification work includes:: 137 p., 61 fig., 2 tabl., 24 sources.

Keywords: real estate market, fraudulent activities, machine learning models, neural networks, fraud detection, information-analytical systems, cybersecurity, fraud prediction, automated control systems.

Object of the research: management processes in the real estate market.

Purpose of the research: to develop a methodology for detecting fraudulent activities in the real estate market using machine learning methods.

Research methods: logical, dialectical, historical, and comparative methods for substantiating the scientific principles and improving the conceptual framework of the research; methods of observation, generalization, abstraction, formalization, analysis, and synthesis to characterize the methodological aspects of creating an information-analytical system; program-target method for substantiating the mechanisms of developing and implementing informatization programs of analytical processes.

Practical value: recommendations for developing a neural network system that will simplify and significantly improve the process of predicting fraudulent activities in the real estate market.

ЗМІСТ

ВСТУП.....	6
РОЗДІЛ 1 ТЕОРЕТИЧНІ ЗАСАДИ ДОСЛІДЖЕННЯ ШАХРАЙСЬКИХ ДІЙ НА РИНКУ НЕРУХОМОСТІ.....	8
1.1 Аналіз стану ринку нерухомості.....	8
1.2 Поточний стан дослідження проблеми виявлення шахрайських дій на ринку нерухомості.....	17
1.3 Типологія підозрілих дій на ринку нерухомості: основні ознаки та критерії виявлення.....	23
РОЗДІЛ 2 МЕТОДОЛОГІЧНЕ ЗАБЕЗПЕЧЕННЯ ТА МАТЕМАТИЧНИЙ АППАРАТ ВИЯВЛЕННЯ ШАХРАЙСЬКИХ ДІЙ НА РИНКУ НЕРУХОМОСТІ.....	28
2.1 Огляд існуючих методик виявлення шахрайських дій на ринку нерухомості.....	28
2.2 Використання моделей машинного навчання для виявлення шахрайських дій на ринку нерухомості.....	32
2.2.1 Gradient Boosting Machines.....	35
2.2.2 Random Forest.....	38
2.2.3 Model Stacking.....	40
2.3 Розробка моделі машинного навчання для аналізу ринку нерухомості.....	43
2.3.1 Показники ROC, AUC.....	48
2.3.2 Покращення результатів роботи моделей.....	49
РОЗДІЛ 3 ПРАКТИЧНЕ ЗАСТОСУВАННЯ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ ДЛЯ ВИЯВЛЕННЯ ШАХРАЙСЬКИХ ДІЙ НА РИНКУ НЕРУХОМОСТІ.....	51
3.1 Процес розробки моделі.....	51

	5
3.2 Оцінка ефективності моделей машинного навчання	57
ВИСНОВКИ	65
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	68

ВСТУП

Актуальність теми. Ринок нерухомості є одним з великих ринків, вартість «товару» на якому є досить високою, тому і шахрайські дії на такому ринку є більш ймовірними і мають більшу розповсюдженість.

Індустрія попередження шахрайства застосовує широкий спектр методів машинного навчання, постійно вдосконалюючи їх. Вибір найефективніших моделей для виявлення шахрайських операцій є ключовим фактором у створенні систем, що забезпечують безпечні операції на ринку нерухомості. Отже, ця робота присвячена пошуку, аналізу та адаптації методів машинного навчання для виявлення шахрайських транзакцій.

Мета дослідження – розробити методику виявлення шахрайських дій на ринку нерухомості за допомогою методів машинного навчання.

Завдання дослідження:

- провести аналіз поточного стану ринку нерухомості;
- здійснити огляд літератури з поточного стану дослідження проблеми виявлення шахрайських дій на ринку нерухомості;
- визначити типологію підозрілих дій на ринку нерухомості: основні ознаки та критерії виявлення;
- провести огляд існуючих методик виявлення шахрайських дій на ринку нерухомості;
- дослідити використання моделей машинного навчання для шахрайських дій на ринку нерухомості;
- оглянути підходи до розробки моделі машинного навчання для аналізу ринку нерухомості;
- здійснити аналіз функціонування розробленої моделі;
- надати оцінку ефективності моделей машинного навчання.

Предметом дослідження є системи виявлення шахрайських дій на ринку нерухомості на основі нейромереж.

Об'єктом дослідження є процеси управління на ринку нерухомості.

Практичне значення одержаних результатів – рекомендації щодо розробки нейромережевої системи, яка дозволить спростити та значно покращити процес прогнозування шахрайських дій на ринку нерухомості.

Методи дослідження. У виконанні даної роботи застосовано такі загальнонаукові методи дослідження: логічний, діалектичний, історичний та порівняльний методи для обґрунтування наукових принципів і вдосконалення понятійного апарату дослідження. Методичні аспекти створення інформаційно-аналітичної системи охарактеризовано за допомогою методів спостереження, узагальнення, абстрагування, формалізації, аналізу та синтезу. Програмно-цільовий метод використано для обґрунтування механізмів розробки та реалізації програм інформатизації аналітичних процесів.

Інформаційна база дослідження включає наукові праці, публікації вітчизняних науковців, офіційні документи, а також нормативно-правові акти у сфері інформатизації та кібербезпеки. Значна увага приділена науковим розробкам у галузі загальної теорії управління, моделювання складних систем, автоматизованих систем управління, інформаційної аналітики, теорії прийняття рішень та менеджменту. Структура роботи складається з вступу, трьох основних розділів, висновків та списку використаних джерел.

Структура роботи складається зі вступу, трьох основних розділів, висновків, переліку використаних джерел.

РОЗДІЛ 1 ТЕОРЕТИЧНІ ЗАСАДИ ДОСЛІДЖЕННЯ ШАХРАЙСЬКИХ ДІЙ НА РИНКУ НЕРУХОМОСТІ

1.1 Аналіз стану ринку нерухомості

Сектор нерухомості є ключовим компонентом сучасної економіки, відіграючи значну роль як в економічному, так і в соціальному прогресі суспільства та його уряду. З моменту здобуття незалежності цей сектор зазнав суттєвого зростання і зараз становить значну частину національного багатства [4, с. 586].

Наведемо типові тенденції на ринку нерухомості, які можуть впливати на ситуацію в Україні.

1. Цінова динаміка: Ціни на нерухомість можуть коливатися в залежності від регіону та типу власності. У більших містах, таких як Київ, Львів чи Одеса, ціни можуть бути вищими через попит на житло та комерційну нерухомість.

2. Вплив політики та економічної ситуації: Політична стабільність та економічна ситуація в країні можуть впливати на довіру інвесторів та споживачів і, відповідно, на активність на ринку нерухомості.

3. Зовнішні фактори: Світові тенденції, такі як зміни відносин міжнародної торгівлі або геополітичні події, також можуть впливати на ринок нерухомості в Україні.

4. Інфраструктурні проекти: Розвиток інфраструктури, такий як будівництво доріг, метро, аеропортів та інші проекти, може змінювати цінову динаміку та попит на нерухомість у відповідних регіонах.

5. Зміни в структурі попиту: Зміни в демографічній ситуації, такі як зростання кількості молодих сімей або підвищення міграційних потоків, можуть впливати на типи нерухомості, які є найбільш популярними.

Для отримання точної інформації про поточний стан ринку нерухомості в Україні можна звернутися до офіційних звітів від органів державної статистики, а також до аналітичних звітів від місцевих агентств нерухомості та міжнародних консалтингових компаній, що вивчають ринок нерухомості.

Згідно з аналітичними звітами та експертними аналізами, на початку 2022 року ринок нерухомості в Україні демонстрував стійке зростання попиту на житло в містах. Проте подорожчання нерухомості на 20-30% порівняно з попереднім роком призвело до спаду ринкової активності [9].

У столичних районах, таких як Київ, Львів та Одеса, спостерігається значне зростання попиту на нові житлові комплекси, особливо на квартири середнього класу. Дані порталу OLX свідчать, що в січні 2022 року кількість оголошень щодо продажу квартир у Києві зросла на 16%, а ціни – на 27% порівняно з груднем 2021 року. Натомість у менших містах і сільській місцевості спостерігається зниження інтересу до житла, що призвело до зниження вартості нерухомості в цих регіонах [9].

Отже, станом на початок 2022 року ринок нерухомості в Україні вважався стабільним, що характеризується зростанням попиту та вартості нерухомості у великих міських центрах, а також зниженням ринкової активності, пов'язаним із зростанням вартості житла.

Стан ринку нерухомості в Україні в перші місяці 2022 року є кульмінацією різних факторів, зокрема підвищеного попиту на житло, збільшення запитів на кредитування та подорожчання будівельних матеріалів і послуг. Особливо помітними є коливання цін на будівельні матеріали, що призводить до збільшення витрат на новобудови та, як наслідок, зростання вартості житла.

Поточний стан ринку нерухомості в Україні з початку 2022 року і до весни 2024 року є кульмінацією різних факторів, зокрема підвищеного попиту на житло, збільшення запитів на кредитування та подорожчання будівельних матеріалів і послуг. Особливо помітними є коливання цін на будівельні матеріали, що призводить до збільшення витрат на новобудови та, як наслідок, зростання вартості житла.

Крім того, політична та економічна ситуація в Україні відіграє ключову роль у формуванні ринку нерухомості, оскільки зміни в податковому законодавстві, курсах обміну валют, нормативно-правовій базі для будівництва та інші зовнішні події безпосередньо впливають на ціни нерухомості та ринкову динаміку.

Тим не менш, незважаючи на деякі перешкоди, з якими стикається сектор нерухомості в Україні, аналітики прогнозують зростання попиту на житло в найближчі роки. Це можна пояснити зростанням населення, демографічними змінами, підвищенням рівня життя та економічним прогресом у країні.

Особливе занепокоєння викликає можливість утворення бульбашки нерухомості через ескалацію цін на нерухомість і підвищений попит на кредити, що потенційно може призвести до спаду ринку та кризового сценарію.

Тому вкрай важливо враховувати ці різноманітні фактори та зосередитися на вдосконаленні системи нерухомості в Україні для сприяння стабільності та росту ринку. Акцент на прозорість, ефективне регулювання та впровадження сучасних технологій може зменшити ризики та підвищити рівень інвестицій у нерухомість. Диверсифікація ринку, підвищення прозорості та вища якість пропозицій нерухомості є важливими для створення середовища, сприятливого для інвестицій, і сприяння стабільному розвитку сектора нерухомості в Україні. Співпраця між державними установами та бізнес-сектором має вирішальне значення для вирішення проблем і забезпечення ефективної роботи ринку [22].

В 2023 році ринок житла потроху почав зростати разом з цінами, оскільки попит на житло був спричинений великою кількістю переселенців із східних областей України. Зростання курсу валют також спричинило збільшення цін за квадратний метр житла протягом 2023 – початку 2024 року.

Динаміка ринку нерухомості за 2020-2023 рр. наведена в Табл. 1.1.

Таблиця 1.1

Динаміка ринку нерухомості за 2020-2023 рр., угод.

	2020	2021	2022	2023
Купівля-продаж квартир та будинків	278 675	325 174	100 710	172 638
Купівля-продаж земельних ділянок	217 810	306 203	133 447	231 881
Кількість іпотечних угод	13 597	22 294	5 410	10 775

Джерело: побудовано автором за даними [9]

За даними дослідження ріелторського порталу Dim.ria [8], у січні 2024 року столиця показала найвищу середню вартість квадратного метра в новобудові. Проте найбільше зростання середньої ціни квадратного метра за рік відбулося в

Житомирській області. На вторинному ринку найбільший сплеск вартості квартир спостерігався в Івано-Франківській та Житомирській областях.

Станом на січень 2024 року у столиці зафіксована найвища середня ціна квадратного метра – 51 377 грн, урахуваючи всі типи квартир за кількістю кімнат.

За місяць квадратний метр знизився на Полтавщині, а на Кіровоградщині та Одещині зріс на 5%.

У січні 2024 року в Україні працювало 75,4% відділів продажу новозбудованого житла.

За місяць по всій країні було зведено 2 новобудови з 5 секцій. Загалом із липня 2023 року до кінця січня 2024 року визнано готовими до заселення 74 будинки на 139 ділянках.

У Київській області за цей період зафіксовано найбільшу кількість зданих новобудов.

Станом на січень було здано близько 30% від загальної кількості новобудов.

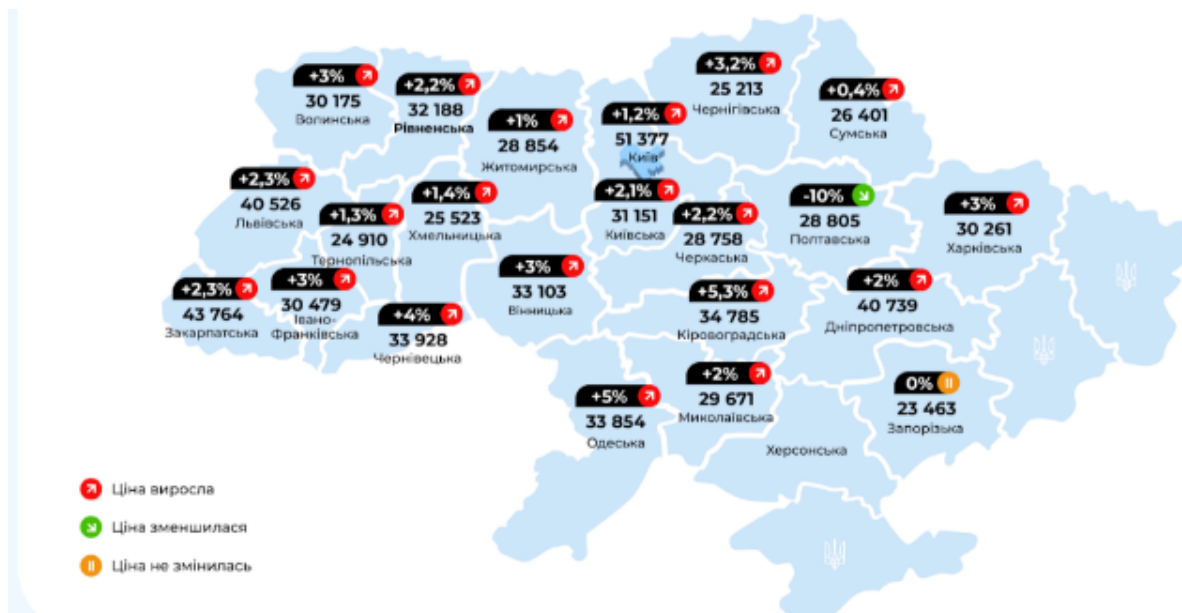


Рис. 1.1 Розмір та динаміка цін за 1 м² в новобудові (фактичне значення і зміна відносно 2023 року)

Джерело: [8]

За даними аналітиків Dim.ria, станом на березень 2024 року в Україні працює 76% відділів продажу первинної нерухомості. У звіті вказується, що за один місяць по країні було зведено сім новобудов із дев'яти секцій. За рік в Україні завершено будівництво 31 новобудови на 53 секції, у тому числі 5 будинків на 7 секцій у Києві. Київська область має найвищий відсоток завершених новобудов – 45%, випереджаючи всі інші регіони.

У березні 2024 року середня ціна квадратного метра в столиці зафіксована на рівні 53 737 грн з урахуванням усіх типів квартир за кількістю кімнат. Порівняння цін за рік показує найбільше зростання цін у Житомирській області – 32%. Крім того, помітно зросли ціни в Івано-Франківській області на 26%, у Рівненській та Черкаській областях на 24% та 21% відповідно.

Дослідження первинного ринку нерухомості Києва показало, що найвища середня ціна квадратного метра в Печерському районі – 99 664 грн, а найнижча – у Дарницькому районі – 37 286 грн.

У березні 2024 року спостерігався помітний сплеск наявності вторинного житла в різних регіонах, значний приріст спостерігався в Івано-Франківській, Волинській та Житомирській областях. І навпаки, у Полтавщині кількість пропозицій зменшилася на 21%, а в Чернівецькій області – менше – на 12% порівняно з аналогічним періодом березня 2023 року.

За даними дослідження, протягом року на вторинному ринку в Україні відбулося помітне зростання вартості житла. У березні 2024 року середня ціна на однокімнатну квартиру у Волинській області порівняно з лютим 2024 року зросла на 19%, навпаки, ціна на такі квартири у Дніпропетровській області знизилася на 4%. За даними експертів Dim.ria, у Києві найвища середня вартість однокімнатної квартири – 93 168 доларів США, на другому місці – Львівська область – 69 680 доларів США та Закарпатська область – 53 401 долар США.

Аналіз цін на житло в різних районах Києва показує, що середня вартість однокімнатної квартири в Печерському районі становить 113 011 доларів, а в Шевченківському районі – 86 862 долари. Найнижча середня ціна квартири у Деснянському районі – 47 042 дол.

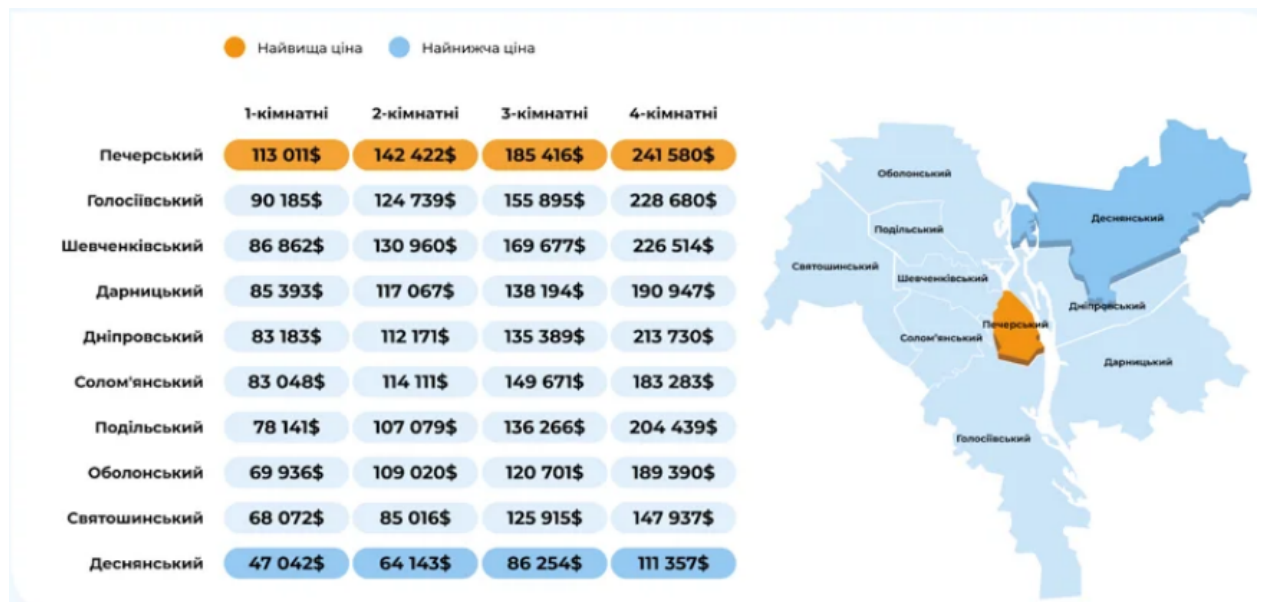


Рис. 1.2. Порівняльна вартість квартир на вторинному ринку м. Києва
Джерело: [8]

У секторі нерухомості в Україні існує багато проблем і перешкод, які вимагають ретельного розгляду та стратегічного планування перед прийняттям будь-яких інвестиційних рішень [15].

Ринок нерухомості в Україні зараз страждає від непрозорості та неадекватних регуляторних заходів. Надзвичайно необхідно посилити існуючі норми та правила, що регулюють цей сектор, наприклад, запровадити надійні механізми моніторингу для дотримання вимог транзакцій, створити ефективні механізми вирішення спорів між учасниками ринку та захист прав інвесторів шляхом ефективного правового захисту.

Крім того, відсутність прозорості на ринку нерухомості може вплинути на довіру інвесторів і стримати інвестиції в сектор нерухомості України. Це може загальмувати зростання економіки країни та перешкодити загальному розвитку ринку. Тому необхідно зосередитися на вдосконаленні нормативно-правової бази та сприянні прозорості, щоб забезпечити стале зростання ринку нерухомості в Україні.

Значну загрозу безпеці та комфорту мешканців українського ринку нерухомості становить незаконне будівництво. Порушуючи права власності та сприяючи конфліктам між зацікавленими сторонами, він підриває довіру до

ринку та призводить до фінансових втрат для інвесторів та забудовників. Негативні наслідки незаконного будівництва зумовлюють необхідність активних державних заходів щодо регулювання та запобігання такій практиці. Крім того, екологічні проблеми є серйозною проблемою для українського сектору нерухомості. Нехтування збереженням навколишнього середовища може призвести до забруднення, зниження якості життя та зниження привабливості для інвестицій та бізнесу. Необхідно розробити та впровадити стратегії сталого розвитку, одночасно забезпечуючи дотримання законів і правил охорони навколишнього середовища.

Таким чином, питання, пов'язані з самовільним будівництвом та екологічними проблемами, є значною перешкодою для сектора нерухомості в Україні. Вирішення цих проблем потребує активного втручання з боку урядових органів, створення надійних регуляторних механізмів для нагляду за будівництвом і прийняття стратегій сталого розвитку, спрямованих на сприяння збереженню навколишнього середовища та підвищення добробуту мешканців [16].

Нестабільність політичної та економічної ситуації в країні може мати наслідки для сектора нерухомості, послабити довіру інвесторів і призвести до коливань цін. Невизначеність, спричинена політичною та економічною нестабільністю, може мати значний вплив на ринок нерухомості України. Цей брак стабільності може підірвати довіру інвесторів і зменшити їх бажання інвестувати в нерухомість. Крім того, ринок може стати свідком непостійних і значних коливань цін, що ще більше ускладнить процес переговорів.

У разі економічної кризи знецінення вартості нерухомості може призвести до фінансових труднощів для фізичних осіб і підприємств, тоді як політична нестабільність може вимагати коригування законів і ринкових правил, створюючи додаткові ризики для інвесторів. Крім того, нестабільність як політичних, так і економічних умов може перешкоджати мешканцям брати участь у операціях з нерухомістю, оскільки високий рівень безробіття та

економічна нестабільність можуть обмежити купівельну спроможність людей і згодом вплинути на ринковий попит.

Недостатній розвиток транспортної, енергетичної та іншої необхідної інфраструктури може мати негативний вплив на бажаність і вартість нерухомості. Ця проблема є поширеною на українському ринку нерухомості, де недостатня транспортна інфраструктура може обмежити доступ до основних зручностей, таких як можливості працевлаштування, навчальні заклади та торгові центри. Крім того, відсутність розвитку інфраструктури може перешкоджати будівництву нових об'єктів нерухомості, роблячи це складним і потенційно нездійсненним заходом.

Недостатній прогрес у розвитку енергетичної інфраструктури може призвести до проблем, пов'язаних з енергопостачанням нерухомості та збільшенням витрат на опалення, що в кінцевому підсумку вплине на бажаність і вартість нерухомості. Крім того, обмежений розвиток інших форм інфраструктури, таких як водопостачання та водовідведення, може створити проблеми як для власників нерухомості, так і для мешканців [13].

Для ефективного вирішення цієї проблеми вкрай важливо створити інфраструктуру, яка відповідає потребам суспільства та ринку нерухомості, що розвивається. Цього можна досягти шляхом залучення інвестицій для будівництва та розвитку інфраструктури разом із впровадженням відповідних регуляторних заходів та отриманням державної підтримки.

На українському ринку нерухомості існує помітний недолік у дотриманні стандартів якості будівництва, що створює потенційні ризики для безпеки та благополуччя мешканців. Це питання викликає особливе занепокоєння в Україні через те, що воно може спричинити загрозу безпеці та здоров'ю мешканців. Випадки невідповідної будівельної практики, такі як неправильне встановлення інженерних мереж і неякісні будівельні матеріали, можуть спровокувати нещасні випадки, пожежі та інші небезпечні обставини. Крім того, неякісна будівельна практика також може вплинути на комфорт мешканців через такі проблеми, як шум, неефективність тепла та інші пов'язані проблеми.

Йдеться про неналежний нагляд і регулювання будівельної галузі в Україні. Хоча в країні існують закони та нормативні акти, що регулюють будівництво, їх виконання та нагляд відсутні. Вкрай важливо регулярно оновлювати та вдосконалювати нормативну базу для приведення у відповідність із сучасними стандартами та забезпечення якості та безпеки будівельних проектів. Дослідження показують, що відсутність суворих будівельних стандартів може значно поставити під загрозу добробут і безпеку мешканців. Наприклад, неякісні матеріали та некваліфікована робоча сила можуть призвести до структурних дефектів і потенційного руйнування будівель, що загрожує життю мешканців.

Крім того, відсутність суворих правил якості може призвести до проблем, пов'язаних з енергоефективністю будівель і забрудненням навколишнього середовища. Нерегульований скид небезпечних матеріалів і неадекватна ізоляція можуть сприяти підвищенню споживання енергії та завдати шкоди навколишньому середовищу.

Враховуючи ці перешкоди, надзвичайно важливо запровадити та підтримувати відповідні стандарти якості будівництва, щоб гарантувати благополуччя та задоволення мешканців та пом'якшити несприятливий вплив на навколишнє середовище. Це вимагає проведення ретельних перевірок якості будівельних матеріалів і практики, а також забезпечення кваліфікації будівельного персоналу.

Для того, щоб сприяти стабільному зростанню сектора нерухомості в Україні, вирішення таких проблем, як самовільне будівництво, неналежний розвиток інфраструктури, політична та економічна нестабільність, а також відсутність стандартів якості будівництва є обов'язковим.

Щоб подолати незаконне будівництво, необхідно посилити нормативно-правову базу та застосувати суворі заходи контролю. Крім того, надзвичайно важливою є боротьба з корупцією в будівельній галузі. Що стосується розвитку інфраструктури, важливо визначити пріоритети потреб мешканців та інвесторів. Підвищення якості транспортних мереж, енергетичних систем, водопостачання та інших об'єктів інфраструктури має першочергове значення.

Для посилення стабільності в політичному та економічному ландшафті України вкрай необхідно запровадити реформи та боротися з корупцією. Ці дії не тільки викличуть довіру інвесторів, а й сприятимуть стабільному зростанню сектора нерухомості. Крім того, встановлення та забезпечення дотримання стандартів якості є важливими для забезпечення безпеки та комфорту мешканців нерухомості. Прозорість і підзвітність у практиці будівництва та дотримання цих стандартів мають вирішальне значення для забезпечення якості будівельних проектів. Також особливо важливим є питання запобігання шахрайству на ринку нерухомості, чому ми і приділимо більш детальну увагу.

1.2 Поточний стан дослідження проблеми виявлення шахрайських дій на ринку нерухомості

Огляд літератури з поточного стану дослідження проблеми виявлення шахрайських дій на ринку нерухомості включає аналіз різноманітних досліджень, які розглядають цю проблему з різних точок зору, використовуючи різні методи та підходи. Деякі з основних напрямків досліджень включають:

1. Аналіз тенденцій шахрайства на ринку нерухомості: Цей напрямок досліджень вивчає різні види шахрайства на ринку нерухомості, такі як фальшиві списки, шахрайство з використанням підроблених документів або недобросовісні практики в агентствах нерухомості [21].

2. Технологічні рішення для виявлення шахрайства: Цей напрямок досліджень розглядає використання різних технологій, таких як машинне навчання, аналіз даних та штучний інтелект, для виявлення шахрайських дій на ринку нерухомості. Це може включати в себе розробку алгоритмів для виявлення підозрілих транзакцій або виявлення недостовірних списків нерухомості [17].

3. Роль правового середовища в протидії шахрайству на ринку нерухомості: Цей напрямок досліджень аналізує роль правового середовища у протидії шахрайству на ринку нерухомості. Він розглядає ефективність правових механізмів, таких як реєстрація власності, захист прав споживачів та регулювання діяльності агентів нерухомості [18].

4. Соціально-економічні аспекти шахрайства на ринку нерухомості: Цей напрямок досліджує соціальні та економічні чинники, що призводять до зростання шахрайства на ринку нерухомості, такі як економічна нестабільність, недостатнє регулювання та низький рівень освіти споживачів [20].

5. Міжнародний досвід та найкращі практики: Цей напрямок досліджень розглядає досвід інших країн у боротьбі з шахрайством на ринку нерухомості та вивчає найкращі практики, які можуть бути застосовані для покращення ситуації в конкретній країні або регіоні [19].

Огляд літератури в цій області може включати як емпіричні дослідження, так і теоретичні розвідки, що сприяють розумінню проблеми шахрайства на ринку нерухомості та розробці стратегій для її подолання.

Для формування уявлень про шахрайські дії на ринку нерухомості звернемось до правової кваліфікації даних дій.

Проблема шахрайства в сфері нерухомості є багатогранною, що впливає з різноманітних організаційних та правових факторів. Таким чином, численним дисциплінам у сфері правових досліджень наполегливо пропонується надати свій досвід і керівництво у боротьбі з цією проблемою за допомогою проактивних заходів, таких як запобігання, виявлення та ретельне розслідування злочинної діяльності [2, с. 16].

Точка зору О.В. Дикого є обґрунтованою, оскільки насторожує наявність організованої злочинної діяльності на ринку нерухомості. Ці злочинні елементи є частиною більшої проблеми, пов'язаної з організованою злочинністю. Виконавцями злочинів у сфері нерухомості, як правило, є добре поінформовані злочинні угруповання, які переслідують осіб, які перебувають у різних життєвих обставинах, що вимагають від них зміни умов життя. Злочини проти власності викликають особливе занепокоєння, оскільки вони часто скоюються дуже стабільними злочинними групами з організованими структурами. Крім того, особи з, здавалося б, заможних верств суспільства також займаються злочинною діяльністю на ринку нерухомості, що свідчить про те, що злочинність нормалізувалася в цьому секторі. Також, існує реальна загроза незафіксованої

віктимізації серед населення, оскільки багато жертв можуть не звертатися за допомогою до правоохоронних органів через відсутність довіри до їхньої здатності надати ефективну підтримку [4, с. 590].

Поява ринкової економіки призвела до чітких характеристик криміналізації на ринку нерухомості, включаючи постійну еволюцію злочинних методів і технік, особливість злочинних тактик, труднощі в класифікації злочинної діяльності та значну загрозу громадській безпеці.

Поширеність злочинної діяльності часто пов'язана з розбіжностями та невідповідностями в житлових і правових нормах, а також з браком правових знань серед населення. Наш аналіз злочинної поведінки в економічному секторі показує тривожний рівень насильства, коли шахраї використовують такі тактики, як вимагання, викрадення та зловживання владою для незаконного придбання власності. Як правило, від наслідків таких злочинів страждають законотрухняні власники майна, в результаті чого жертви часто залишаються бездомними та злиденними [3; с. 110].

У світлі викликів сучасності виникає необхідність розрізняти шахрайські дії на ринку нерухомості, які часто приховуються під виглядом цивільно-правових правочинів зловмисниками, які прагнуть шляхом обману незаконно заволодіти майновими або житловими правами [3; с. 111].

Згідно з висновками К. Чередника та К. Чаплінського [2], умови, що панують на ринку нерухомості, свідчать про шахрайські дії, які відбуваються на ньому:

- 1) Трапляються законодавчі прогалини та правові колізії, які виникають при вирішенні питань щодо набуття та переходу прав на нерухоме майно, а також придбання об'єктів нерухомого майна в новозбудованих будинках;
- 2) Відсутність належного контролю з боку контролюючих органів за діяльністю суб'єктів та організацій, які здійснюють операції з нерухомим майном;
- 3) Недотримання суб'єктами операцій з нерухомим майном норм права при здійсненні операцій з нерухомим майном призводить до недостатнього

захисту прав і законних інтересів громадян, які беруть участь у цивільних правовідносинах, при укладенні різних видів угод з нерухомим майном;

4) Невиконання посадових обов'язків особами, до яких доручено видавати документи та дозвільні документи, необхідні для вчинення правочинів, що призвело до порушення прав, а в окремих випадках і до втрати від права власності на нерухоме майно;

5) Відсутність належного правового представництва потребуючих осіб та невиправлення неясностей у цивільно-правових угодах під час їх укладення;

6) Державні та правоохоронні органи активно підтримують злочинні організації, які займаються незаконною діяльністю у сфері нерухомості;

7) Поширена участь у складі ЗУ осіб та представників органів, які відповідають за супроводження правочинів з нерухомістю;

8) Вчинення шахрайських діяч у сфері нерухомості, а також інші кримінальні злочини, такі як вбивство та вимагання;

9) Неадекватне та несвоєчасне реагування правоохоронних органів на факти шахрайства на ринку нерухомості [2; с. 28].

Методи вчинення шахрайства в галузі нерухомості залежать від різних факторів, включаючи нормативне середовище в певному штаті чи регіоні, рівень правового нагляду, що регулює операції з нерухомістю, характеристики жертви та те, чи відбувається транзакція в первинний або вторинний ринок нерухомості. Крім того, вид та специфіка цивільно-правових угод відіграють значну роль у визначенні способів обману осіб при набутті права власності на майно.

У дослідженні шахрайських дій у сфері інвестування коштів у житлове будівництво І. М. Попова [5] наголошує на важливості класифікації протиправних дій за різними стадіями процесу. Вона стверджує, що класифікація шахрайських дій необхідна для ефективного розмежування взаємопов'язаних ланцюгів злочинної діяльності та встановлення в них ієрархічної структури:

1) Оманливі практики, пов'язані з вимаганням коштів від інвесторів без наміру виконання обіцянок щодо будівництва житла;

2) Шахрайські дії, що здійснюються під час виконання обов'язків, пов'язаних із житловим будівництвом;

3) Шахрайство на етапі здачі об'єкта будівництва в експлуатацію [5, с. 85].

С.С. Кузьменко [6] також наголошує, що злочинна діяльність, пов'язана з розвитком проектів у сфері нерухомості, має широкий розмах. Ці незаконні дії можуть варіюватися від придбання землі (в різні способи) до завершення будівельних об'єктів та реєстрації прав власності. Автор рекомендує при визначенні методів враховувати механізми інвестування та суб'єктів, залучених до цих дій. Зазначається, що в будівельній сфері застосовуються як прямо визнані законодавством способи, так і ті, що спеціально не визначені в нормативних документах, що регулюють діяльність з інвестування в нерухомість. Останні більш поширені. Слідчі та судові дані свідчать про те, що лише 9% випадків шахрайства включають такі механізми, як FON, FFB та цільові облігації, при цьому схема ISI не використовується взагалі. Більшість (91%) випадків шахрайства пов'язані з використанням альтернативних механізмів інвестування в розвиток нерухомості [6].

У свою чергу, У своєму дослідженні шахрайства організованими злочинними групами в будівельній галузі Опанасенко виділив кілька основних методів. Йдеться про укладення попередніх договорів бронювання квартир, які не гарантують інвестору право на отримання майна після завершення будівництва, участь у незаконних договорах інвестування житла без необхідних дозволів, надання неправдивих відомостей у договорах інвестування будівництва, внесення змін до основних договорів через додаткових угод, приховування фінансової неспроможності юридичних осіб, а також здійснення подвійного продажу квартир у об'єктах житлового будівництва [7, с. 9].

Щодо глобального досвіду протидії шахрайству на ринку нерухомості, можна зазначити, що такий вид обману менш поширений на міжнародній арені, ніж в Україні. Наприклад, у країнах Європи та Сполучених Штатах застосовуються три основні рівні захисту інвесторів у будівельній сфері: юридичний захист, професійний консалтинг та страхування.

Наприклад, згідно з даними Федерального бюро розслідувань (ФБР), у США злочини у сфері нерухомості найчастіше виявляються у формі:

- корпоративного шахрайства;
- іпотечного шахрайства;
- інвестиційного шахрайства.

Основною формою фінансової злочинності у США є іпотечне шахрайство, особливо шахрайство субкредиторів у сфері іпотечного кредитування. Це відбувається під час субкредитування позичальників, які не можуть отримати кредит від основного кредитора через об'єктивні чи суб'єктивні причини. Найпоширеніші протиправні дії під час субкредитування включають:

- Надання позичальником іпотечному брокеру або кредитору хибної інформації;
- Фальсифікацію фінансових звітів керівництвом компаній, які отримують кошти, з метою створення уявлення про успішну компанію. Це часто включає коригування бухгалтерських записів та штучне завищення активів і доходів. Деякі компанії маніпулюють ризиками кредитного портфеля та використовують різні облікові схеми для "роздування" фінансових звітів.

У першому випадку шахрайство є досить поширеним і в сучасній Україні, тоді як другий випадок є більш специфічним. Керівництво компаній, що залучають кошти під іпотечні програми, часто фальсифікує звітні відомості під час деструктивних процесів на ринку іпотечного кредитування та вторинному ринку іпотечних цінних паперів. Це робиться для того, щоб створити уявлення про успішну компанію та залучити більше коштів, що врешті-решт призводить до значних збитків для субкредиторів, які не встигають вилучити свої інвестиції.

З початком фінансової кризи кількість випадків іпотечного шахрайства зросла, особливо в регіонах з високим попитом на іпотечні кредити. Хоча через високу латентність точний рівень іпотечного шахрайства у США залишається невідомим, зареєстрований рівень зростає щороку. Аналітики ФБР прогнозують, що під час "буму" на ринку нерухомості обсяги іпотечного кредитування збільшаться, що підвищить ризик шахрайства у цій сфері [9].

До зростання ризику шахрайства також сприяє використання Інтернету та пов'язаних з ним технологій для отримання та оформлення кредитних заяв. Велика кількість "знеособлених" операцій (коли ідентифікаційні документи клієнта скорочені або відсутні) створює сприятливі умови для шахрайства. Отримання іпотечного кредиту через Інтернет або телефон часто призводить до того, що позичальник і кредитор ніколи не зустрічаються, навіть під час закриття кредиту. Іноді кредитори надсилають документи позичальникам кур'єрською службою і отримують їх назад таким же чином [10].

Шахрайства на ринку нерухомості як окремий вид або в контексті дослідження фінансових шахрайств в цілому досліджували багато українських та зарубіжних науковців, зокрема, автоматизованому виявленню шахрайських дій присвячено працю А. Мінц [12], а проблематиці фінансового моніторингу – дослідження Н. Внукової [11] та інших фахівців.

Отже, шахрайським діям на ринку нерухомості присвячено ряд досліджень, проте в них не надто досліджено можливість запобігання таким діям і прогнозування їх ймовірності для окремої операції.

1.3 Типологія підозрілих дій на ринку нерухомості: основні ознаки та критерії виявлення

На ринку нерухомості існують різні види підозрілих дій, які можуть вказувати на недобросовісність, шахрайство або порушення законодавства. Ось деякі основні ознаки та критерії виявлення таких дій (Табл. 1.2).

Таблиця 1.2

Ознаки підозрілих дій на ринку нерухомості

№ з/п	Дія	Характеристика
1	Непрозорість в угодах	включає в себе недостатню документацію, незрозумілі умови угоди, а також відмову надавати додаткову інформацію
2	Недостовірна інформація	Якщо надана інформація про об'єкт нерухомості (наприклад, про стан будівлі, правовий статус власності тощо) суперечить дійсності або є неперевіреною, це може бути підозрілим

№ з/п	Дія	Характеристика
3	Надмірний тиск або поспішність	Шахраї можуть намагатися змусити покупця чи продавця прийняти рішення швидко, не надаючи достатньо часу на аналіз угоди або перевірку умов.
4	Відмова від прозорості у фінансових операціях	Недостатня чіткість у фінансових документах, відмова від надання документів, що підтверджують походження коштів або фінансових операцій.
5	Незвичайні методи оплати	Використання незвичайних або неперевірених методів оплати може бути ознакою нелегальних дій
6	Зміни у власності або ціні без належних пояснень	Різкі зміни у власності об'єкту нерухомості або в ціні, які не мають обґрунтування, можуть бути підозрілими
7	Штучне підвищення цін	Маніпуляції з ринковими цінами нерухомості для отримання непропорційних прибутків
8	Порушення прав інших осіб	випадки фальшивого представлення прав на нерухомість або випадки, коли одна сторона намагається придбати нерухомість, яка насправді належить іншій особі
9	Надмірні вимоги до власників нерухомості	вимоги високих комісій чи інших платежів без належного обґрунтування.
10	Відмова від договірних зобов'язань	Невиконання умов укладених договорів або відмова від їх виконання без належних пояснень

Джерело: побудовано автором на основі [1]

Виявлення таких підозрілих дій може вимагати уважності, ретельного аналізу угоди і співпраці з професіоналами, такими як адвокати чи фахівці з нерухомості.

Дослідження випадків шахрайства на ринку нерухомості виявило безліч незаконних схем. Цю шахрайську діяльність зазвичай можна розділити на три основні групи:

1) Перша категорія охоплює незаконну діяльність, яка явно спрямована на привласнення майна інших і часто передбачає додаткові порушення закону, наприклад, загрозу громадській безпеці. Це групування складається з брехливої та агресивної поведінки, вимагання, залякування тощо. Ця модель злочинної поведінки зазвичай спрямована проти ізольованих, літніх або економічно знедолених осіб. Це також може включати незаконне захоплення нерухомого майна шляхом маніпулювання частками власності та виключення інших співвласників. Діяльність цієї категорії є особливо небезпечною для суспільства,

оскільки вона спрямована не лише на заволодіння чужими активами, а й становить значний ризик для життя та добробуту людей;

2) Другу категорію можна позначити як «замасковані правопорушення», оскільки дії явно порушують закон, але маскуються в правовому полі. Ці шахрайські дії в сфері нерухомості включають:

- підробку необхідних документів для оформлення угоди, таких як довіреність або рішення суду, що підтверджує право власності на майно;
- продаж майна за недійсними або анульованими довіреностями;
- укладання фіктивних угод;
- обман осіб, які не мають правових знань, під час продажу майна під іпотеку;
- «подвійний продаж» квартир, у тому числі в девелоперських проектах.

3) До третьої категорії належать дії, які не передбачають безпосереднього набуття права власності на чуже майно, але можуть призвести в кінцевому підсумку до визнання угоди недійсною. До цієї категорії належать такі дії, як приховування інформації щодо осіб, які мають право користування житловими приміщеннями, вимога отримання згоди третьої особи на здійснення операції, а також, осіб, які звільнили своє житло під виправні чи соціальні установи [3; с. 113].

Найпоширеніші шахрайські практики під час купівлі та продажу нерухомості в новобудовах зазвичай передбачають:

- Угода передбачає продаж права на розробку, а не самого фізичного майна на етапі будівництва, таким чином наражаючи сторони на потенційний ризик незавершення проекту або його введення в експлуатацію.
- Шахрайський акт багаторазового продажу одного і того ж майна різним особам.
- У контексті продажу нерухомості поширеною є практика заміни виставленого об'єкта об'єктом іншого призначення, наприклад, презентація відремонтованої квартири, а в кінцевому підсумку продаж об'єкта з мінімальною внутрішньою обробкою.

Також варто більш детально описати наступні схеми:

Фальсифікація документів. На ринку нерухомості злочинці використовують підроблені документи про право власності та особисті документи для полегшення продажу нерухомості. Незважаючи на те, що покупці часто залежать від нотаріального засвідчення як гарантії, залишається ймовірність як ненавмисної людської помилки, так і навмисної співучасті в шахрайських діях.

Довіреність. Використання довіреності для купівлі нерухомості є юридично дозволеною дією, яка може бути виправдана за певних обставин. Однак ця практика також може бути вразливою для використання особами-шахраями. Таким чином, доцільно бути обережними та уникати операцій з нерухомістю, придбаною через довірену особу. Випадки шахрайського отримання довіреності, а також подальше оскарження дійсності угоди не є рідкістю в таких сценаріях.

«Прописані» мешканці. Власники житла можуть вирішити продати свою нерухомість, навіть якщо інші особи зареєстровані як мешканці. Деякі домовласники можуть усно погоджуватися на те, щоб мешканці змінили свою реєстраційну адресу, або стверджувати, що це вже зроблено, але не надають жодних документів на підтвердження цього. Згідно зі статтею 7 Закону України «Про свободу пересування та вільний вибір місця проживання в Україні», зареєстрованого мешканця можна виселити лише за його згодою або за рішенням суду, що вимагає звернення нового власника до суду з позовом про виселення колишнього резидент. Ситуація ускладнюється, коли йдеться про неповнолітніх, оскільки їх неможливо виселити без додаткового дозволу органу опіки та піклування. Хоча законодавство певною мірою врегульовує це питання, на практиці зареєстровані неповнолітні мешканці можуть створювати проблеми з використанням майна.

Борг за комунальні послуги. Власники нерухомості пропонують свою нерухомість на продаж разом із будь-якими неоплаченими комунальними платежами. Ці рахунки можуть коливатися від невеликих сум, які можуть призвести до припинення надання послуг постачальником, до більших сум, які можуть бути передані новому власнику у разі несплати.

Занижена оціночна вартість. Договір купівлі-продажу укладається за ціною, нижчою від ринкової вартості речі. Згодом продавець вирішує розірвати угоду, в результаті чого покупець отримує початково узгоджену суму, яка може значно відрізнятись від фактично здійсненої оплати.

Спільна власність. Мається на увазі можливість продажу майна без дозволу співвласника у випадку спільної власності, а також продаж майна без згоди подружжя у випадках спільної власності.

Нерухомість під арештом або в заставі. Ці типи нерухомості зазвичай продаються за зниженою ціною, що змушує покупців, які поспішно купують, не звертати уваги на необхідні перевірки. Крім того, це майно може бути предметом юридичних ускладнень, таких як перебування під арештом, під заставу або участь у різних судових спорах.

Спадщина. Особи, які успадковують нерухомість, часто вирішують продати нерухомість незабаром після отримання права власності. Тим не менш, існує потенційний ризик появи непередбачених спадкоємців, які оскаржуватимуть продаж майна та претендуватимуть на частину квадратних метрів, що призведе до потенційних судових проблем.

Завдаток. Шахраї використовують різноманітні схеми, що включають авансові платежі, із загальною тактикою, щоб отримати депозити від кількох покупців, перш ніж втекти з коштами під виглядом законних власників.

Отже, існує ряд підозрілих дій, які можуть бути в майбутньому класифіковані як шахрайські, тому запобігання їм є нагальною потребою, а отже розробка механізму оцінки ризиків та виявлення шахрайських дій є актуальним і потребує детальної розробки.

РОЗДІЛ 2 МЕТОДОЛОГІЧНЕ ЗАБЕЗПЕЧЕННЯ ТА МАТЕМАТИЧНИЙ АППАРАТ ВИЯВЛЕННЯ ШАХРАЙСЬКИХ ДІЙ НА РИНКУ НЕРУХОМОСТІ

2.1 Огляд існуючих методик виявлення шахрайських дій на ринку нерухомості

Задачею системи моніторингу дій на ринку нерухомості для боротьби з шахрайством в системах є безперервний контроль за основним його компонентом – вартістю та відповідним платежем покупця продавцеві. Моніторинг шахрайських дій у реальному часі може бути:

- ручний;
- напівручний;
- автоматичний.

Метод ручного моніторингу передбачає покрокову перевірку платежів працівниками відділу безпеки в реальному часі. Цей підхід до моніторингу вважається застарілим і неефективним через великий обсяг платежів, які обробляються сучасними платіжними системами, що ускладнює його обробку одному співробітнику служби безпеки. Насправді цей метод потенційно може принести більше проблем, ніж користі, оскільки неефективний моніторинг може призвести до значних фінансових втрат. Як результат, цей тип моніторингу шахрайства рідко використовується в поточній практиці та може бути реалізований лише на ранніх стадіях розвитку системи, а не в довгостроковій перспективі.

Підхід до практичного моніторингу передбачає перехід виявлення шахрайських дій до автоматизованої системи, а остаточні рішення та дії делегуються відділу безпеки. Цей метод є більш ефективним, ніж ручний моніторинг, оскільки він покладається на автоматизовану систему, яка дотримується заздалегідь визначеного алгоритму для виявлення шахрайства. Однак необхідний постійний нагляд з боку персоналу відділу безпеки, оскільки автоматизована система обмежена виявленням шахрайських дій і не має можливості впроваджувати профілактичні заходи.

Автоматичний режим моніторингу пропонує найбільш значні переваги порівняно з попередніми типами моніторингу. Цей метод не тільки автоматизує процес виявлення шахрайських дій, але й автоматизує вибір профілактичних заходів. Системи моніторингу протидії шахрайству такого роду дуже затребувані через потребу в ефективному та постійному контролі в режимі реального часу, що неможливо досягти за допомогою інших систем моніторингу. Це пов'язано з великим об'ємом операцій, які можуть виконуватися протягом декількох секунд і вимагають миттєвої обробки.

Після вивчення основних форм моніторингу в режимі реального часу виникає питання про те, які методології є найбільш ефективними для виявлення шахрайських дій. Враховуючи динамічний характер ринку нерухомості, поява нових видів шахрайства вимагає постійного вдосконалення та коригування. Однак усі форми шахрайства в першу чергу зосереджуються на параметрах платежу та транзакціях, здійснених платіжною системою.

Для ефективного вибору та впровадження заходів протидії шахрайству важливо спочатку визначити конкретну сферу уваги для постійного моніторингу. Ця сфера уваги, відома як об'єкт контролю, може охоплювати різні аспекти, такі як вартість квадратного метра нерухомості, рахунки клієнтів, рахунки постачальників тощо. По суті, об'єкт управління виступає як основна функціональна одиниця в системі. Згодом слід застосувати відповідні методи виявлення шахрайства відповідно до обраного об'єкта контролю.

Основним методом виявлення шахрайських дій є систематичне дослідження показників об'єкта управління. Моніторинг здійснюється за попередньо визначеними сценаріями, розробленими відділом безпеки, які адаптовані до конкретного об'єкта моніторингу та рівня його ризику. Сценарій слугує орієнтиром, окреслюючи поведінку, яка вважається неприйнятною та потенційно вказує на шахрайство з боку покупця чи продавця. Щоб розробити ефективний сценарій, важливо класифікувати об'єкти контролю на категорії високого ризику, середнього ризику та низького ризику. На основі цієї класифікації розробляється сценарій для оцінки конкретних показників, таких як порогові значення.

Порогове значення представляє максимально допустиму кількість транзакцій для певного об'єкта контролю, наприклад продавця, протягом визначеного періоду часу. Якщо цей поріг перевищено, об'єкт позначається як шахрайський, що запускає запобіжні заходи. Порогове значення визначається рівнем ризику, пов'язаного з об'єктом контролю, причому об'єктам підвищеного ризику призначаються нижчі порогові значення. Крім того, сценарії можуть включати баланси облікових записів користувачів як індикатор потенційного шахрайства, коли система автоматично перевіряє баланси облікових записів на відповідність заздалегідь визначеним пороговим значенням, встановленим персоналом служби безпеки під час платіжних операцій.

Застосування методу покрокового аналізу параметрів дії на ринку нерухомості для конкретного об'єкта контролю передбачає встановлення гранично допустимих значень цих параметрів. Будь-яке порушення цих цінностей призведе до негайного застосування запобіжного заходу. Визначення допустимих показників різняться в залежності від об'єкта контролю, оскільки різні послуги мають різні показники оплати. Одним з поширених показників є вартість квадратного метра. Якщо це значення занижено більш ніж на 20%, це може свідчити про шахрайство. Наприклад, встановлення вартості об'єкта нижче 20% не дозволяє шахраям негайно здійснити операцію купівлі-продажу після отримання коштів. Крім того, частота зняття коштів з рахунку також може сигналізувати про можливі шахрайські дії, спрямовані на швидке зняття коштів. Якщо система моніторингу виявляє шахрайські платежі, то рахунок, як один із параметрів платежу, часто автоматично блокується.

Моніторинг шахрайства можна використовувати для розробки сценарію, який блокує платежі на основі IP-адреси на етапі обробки платежу, ефективно запобігаючи платежам із певних регіонів чи місцевостей без необхідності розглядати інші потенційні сценарії шахрайства. Наприклад, якщо платежі по операції купівлі-продажу здійснюються на карту фізичної особи, хоча сторонами договору є юридичні особи тощо.

Використання методів покрокового аналізу платіжних показників дозволяє детально описувати правила сценарію та оперативно реагувати. Поряд з таким підходом поширеним є використання «білих» і «чорних» списків. До «чорного» списку входять об'єкти контролю, які раніше порушували певні сценарні правила. Ці списки зазвичай створюються окремо для кожного об'єкта керування, зберігаючи порушене правило. Наприклад, ріелтор або номер картки, які беруть участь у шахрайських платежах, можуть бути включені в «чорний» список, щоб заборонити будь-які транзакції в системі до з'ясування законного власника картки. Включення до «чорного» списку призводить до автоматичного блокування оплати об'єкта контролю без проходження перевірки на визначені сценарії відділом безпеки до виключення зі списку. У «чорний» список можуть потрапити не тільки номери гаманців і карток або IP-адреси, але й персональні дані клієнта, такі як прізвище та по батькові клієнта, якщо їх визнають шахрайськими, що призведе до обмеження всіх операцій незалежно від типу.

Білі списки складаються з фізичних або юридичних осіб, які пройшли попередню перевірку та вважаються такими, що заслуговують на довіру, тому не потребують додаткової перевірки для виявлення шахрайства. Це контрастує з чорними списками, які ідентифікують підозрілу або шахрайську діяльність, за якою потрібно ретельно стежити. Включення фізичних або юридичних осіб до білого списку ґрунтується на попередній перевірці та укладенні окремих угод для зменшення потенційних ризиків шахрайства.

Альтернативний підхід до аналізу історії транзакцій продавця нерухомості в автономному режимі передбачає вивчення серії платежів для оцінки моделей депозитів і зняття коштів на основі різних операцій. Цей аналіз зосереджується на частоті, термінах і сумі здійснених платежів. Підозра може виникнути, якщо з однієї картки чи гаманця часто знімаються великі суми, що призводить до таких можливих заходів, як тимчасове блокування до з'ясування будь-яких проблем і подальше розблокування, якщо це буде визнано необхідним.

2.2 Використання моделей машинного навчання для виявлення шахрайських дій на ринку нерухомості

Після проведення ретельного огляду відповідної літератури та визначення дослідницького завдання наступний етап передбачає оцінку основних методологій і математичних основ, які будуть використовуватися в перспективній інформаційній системі, призначеній для виявлення шахрайських операцій.

Одними з найбільш популярних алгоритмів машинного навчання, є алгоритми, побудовані на основі методу дерев рішень. Дерева рішень є графічним методом організації послідовного процесу прийняття рішень. Вони складаються з вузлів рішень, кожен з яких має гілки, що відповідають різним альтернативним рішенням. У дереві також присутні вузли шансів (випадкові змінні), причому корисність кожної гілки обчислюється на листку цієї гілки. Таким чином, очікувана корисність будь-якого рішення може бути визначена шляхом зваженого підсумовування всіх гілок від рішення до відповідних листків.

Особливість цієї математичної моделі полягає в тому, що вона описує процес прийняття рішень таким чином, щоб врахувати кожне можливе рішення, попередні та наступні події або інші рішення, а також наслідки кожного кінцевого рішення.

Оскільки листи являють собою частини даних з однією цільовою ознакою, саме ця ознака буде прогнозом для заданих описових ознак. Кількісна міра, що визначає, наскільки добре ознака розбиває дані, обчислюється за допомогою ентропії Шеннона. Ентропія визначає кількісну міру неоднорідності елементів у множині:

$$H(t, D) = - \sum_{i \in I} [P(t = i) * \log_2(P(t = i))], \quad (2.1)$$

де $P(t = i)$ – ймовірність того, що цільова ознака t належить класу i ;

l – кількість різних класів цільової ознаки у наборі даних D .

Мірою інформативності ознаки, що використовується в деревах рішень, є приріст інформації. Для обчислення приросту інформації використовується

формула ентропії для множини даних D відносно цільової ознаки. Ентропія після розбиття множини даних за конкретною ознакою d визначається формулою (2.2):

$$rem(d, D) = \sum_{l \in \text{рівні}(d)} \frac{|D_{d=l}|}{|D|} \cdot H(d, D_{d=l}), \quad (2.2)$$

Дерева рішень допомагають розв'язувати задачі як класифікації, так і прогнозування, при цьому показують високу точність. Таких алгоритмів є безліч і кожен з них має свої особливості, переваги і недоліки. Розглянемо ті, які будуть застосовуватися при прогнозуванні цін на нерухомість в цій роботі.

Формула (2.1) визначає ентропію, що залишається після поділу множини даних за певною ознакою d . На основі цього поділу можна обчислити приріст інформації. Приріст інформації (Information Gain, IG) визначається як різниця між ентропією до поділу та залишковою ентропією після поділу:

$$IG(d, D) = H(t, D) - rem(d, D), \quad (2.3)$$

де $H(t, D)$ — ентропія цільової ознаки t в наборі даних D ;

$rem(d, D)$ — залишкова ентропія після поділу даних за ознакою d .

Таким чином, приріст інформації є мірою, що визначає, наскільки ефективно певна ознака d зменшує невизначеність (ентропію) щодо цільової ознаки t в множині даних D .

Дерева рішень застосовуються як для класифікації, так і для прогнозування неперервних величин.

Розглянемо приклад для кращого розуміння. Тема прикладу – ринок нерухомості. Існує проблема прийняття рішення щодо того, чи варто інвестувати в нерухомість у певному районі. Якщо інвестувати, а згодом ринок нерухомості впаде, то це призведе до фінансових втрат; з іншого боку, якщо не інвестувати і ринок зросте, то буде втрачено потенційний прибуток.

Щоб чисельно описати цю ситуацію, використаємо ймовірності:

$$p(\text{Market} = \text{fall}) = 0.6, \quad (2.4)$$

$$p(\text{Market} = \text{rise}) = 0.4, \quad (2.5)$$

Користь визначається таким чином:

$$U(\text{invest}, \text{fall}) = -100, \quad (2.6)$$

$$U(\text{invest}, \text{rise}) = 500, \quad (2.7)$$

$$U(\text{no invest}, \text{fall}) = -100, \quad (2.8)$$

$$U(\text{no invest}, \text{rise}) = 50, \quad (2.9)$$

Цю ситуацію можна представити графічно у вигляді дерева рішень. Питання полягає в тому, чи слід інвестувати в нерухомість. Оскільки невідомо, що насправді станеться з ринком, обчислимо очікувану корисність кожного рішення:

$$U(\text{invest}) = U \sum_{\text{Market}} (\text{invest}, \text{Market}) \cdot P(\text{Market}), \quad (2.10.1)$$

$$U(\text{invest}) = (-100) \cdot 0.6 + 500 \cdot 0.4 = 140, \quad (2.10.2)$$

$$\begin{aligned} U(\text{no invest}) &= U \sum_{\text{Market}} (\text{no invest}, \text{Market}) \cdot P(\text{Market}) = \\ &= 0 \cdot 0.6 + 50 \cdot 0.4 = 20, \end{aligned} \quad (2.11)$$

Виходячи з очікуваної корисності, слід інвестувати в нерухомість, оскільки максимальна очікувана корисність становить 140.

$$\max_{\text{Market}} U(\text{invest}, \text{Market}) \cdot P(\text{Market}) = 140, \quad (2.12)$$

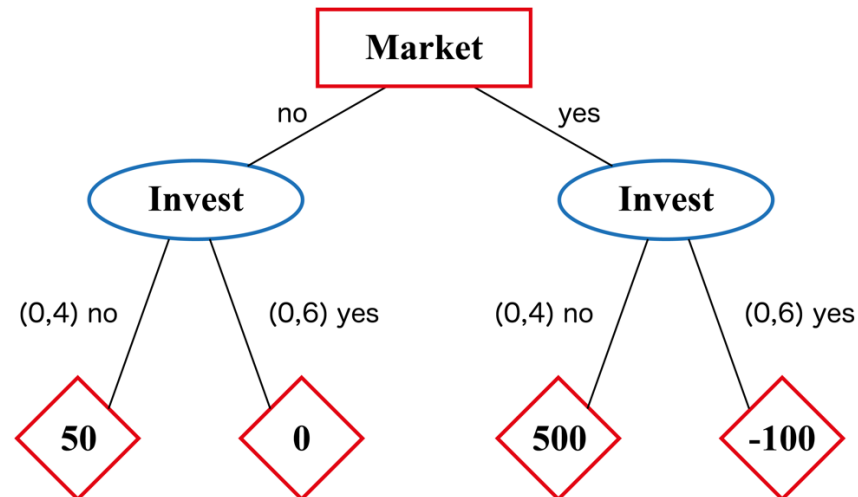


Рис. 2.1. Дерево рішень, яке складається з вузлів випадків (представлені у вигляді овалів), вузли рішення (квадрати) та корисних вузлів (ромби).

Джерело: побудовано автором

Варто зазначити, що дерево рішень не є графічним представленням мережі вірувань з додатковими вузлами. Швидше, дерево рішень є явним перерахуванням можливих варіантів вибору, які можна зробити, починаючи з самого лівого вузла вирішення, з ймовірностями на гілках поза "шансовими" вузлами.

2.2.1 Gradient Boosting Machines

Машина з підсиленням градієнта (GBM) – надзвичайно популярний алгоритм машинного навчання, який виявився успішним у багатьох сферах і є одним із провідних методів перемоги в змаганнях Kaggle. У той час як алгоритм Random Forrest (випадковий ліс) створюють ансамбль глибоких незалежних дерев, GBM будують ансамбль неглибоких і слабких послідовних дерев, кожне з яких навчається та вдосконалює попереднє. У поєднанні ці багато слабких послідовних дерев утворюють потужний «комітет», який часто важко перемогти іншими алгоритмами. Загалом методи на основі дерев, як правило, добре

працюють на необроблених даних (тобто без функцій нормалізації, центрування, масштабування) [36].

Бустінг - це техніка побудови ансамблевих моделей, в якій моделі прогнозування будуються не незалежно, а послідовно. Ця методика базується на ідеї, що наступна модель навчається на помилках попередньої, тобто кожна послідовна модель будує просту слабку модель, щоб трохи покращити залишкові помилки (рис. 2.2).

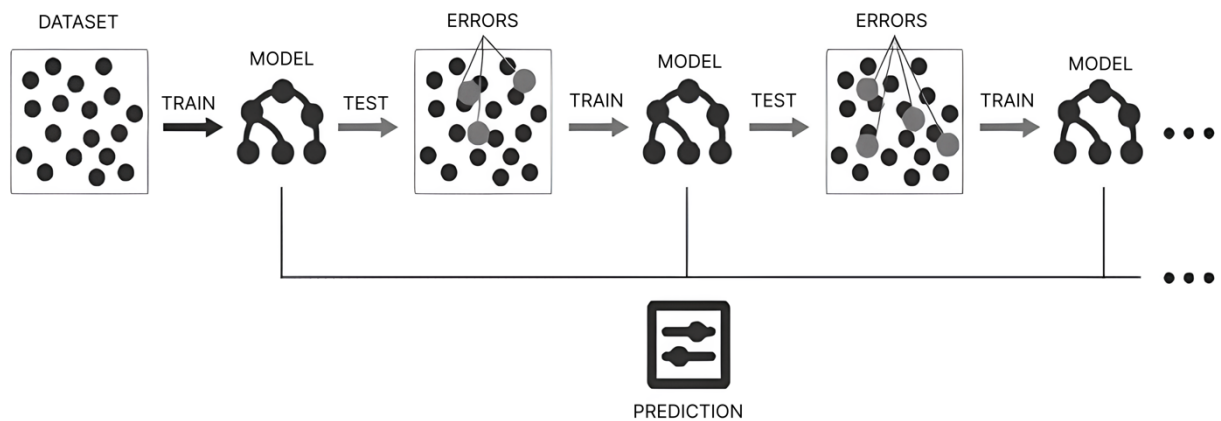


Рисунок 2.2. Принцип роботи алгоритму градієнтного бустингу

Джерело: [37]

Прогностичні моделі мають нерівні ймовірності включення в наступні етапи навчання: частіше використовуються ті, які допускають найбільші помилки. У бустінгу можуть бути застосовані різні типи моделей, такі як дерева рішень, регресійні моделі та класифікатори. Оскільки нові моделі навчаються на помилках попередніх, загальний час навчання зменшується, що дозволяє швидше досягти високої точності. Однак критерій зупинки слід обирати ретельно, щоб уникнути перенавчання. Метод "найшвидшого бустінгу" є прикладом бустінгу.

Градієнтний бустінг є технікою машинного навчання, яка використовується для розв'язання задач регресії та класифікації шляхом створення прогнозувальної моделі у вигляді ансамблю слабких моделей, зазвичай дерев рішень. Коли дерево рішень виступає в ролі слабого прогнозиста, такий алгоритм називається деревами з посиленням градієнтом, які зазвичай перевершують випадкові ліси. Цей метод будує модель поетапно, подібно до інших методів бустінгу, і

узагальнює їх, дозволяючи оптимізувати довільну диференційовану функцію втрат.

Метою будь-якого алгоритму навчання з учителем є визначення функції втрат та її мінімізація. Розглянемо математику градієнтного бустінгу. Припустимо, що функцією втрат є середньоквадратична помилка (MSE):

$$MSE = \sum (y_i - y_i^p)^2, \quad (2.13)$$

де y_i – цільове значення;

y_i^p – отримане.

Необхідно побудувати передбачення таким чином, щоб MSE була мінімальною. Використовуючи градієнтний спуск і оновлюючи передбачення на основі швидкості навчання (learning rate), ми шукаємо значення, за яких MSE досягає мінімуму. Оновлення передбачень здійснюється таким чином, щоб сума відхилень наближалася до нуля, а прогнозовані значення були близькими до реальних.

Поєднання багатьох слабких моделей (проти кількох сильних, як у Random Forest) має кілька переваг:

- швидкість: побудова слабких моделей обчислювально менш затратна;
- покращення точності: слабкі моделі дозволяють алгоритму повільно навчатися; внесення незначних коригувань у нових областях, де він не працює добре. Загалом статистичні підходи, які навчаються повільно, як правило, добре працюють;
- уникає перенавчання завдяки лише невеликим поступовим вдосконаленням кожної моделі в ансамблі, це дозволяє зупинити процес навчання, щойно буде виявлено перенавчання (зазвичай за допомогою перехресної перевірки).

2.2.2 Random Forest

Random forest (з англ. - «випадковий ліс») - це алгоритм машинного навчання, що базується на використанні ансамблю вирішальних дерев. Алгоритм поєднує дві ключові концепції: метод бегінга, розроблений Лео Брейманом, та метод випадкових підпросторів, запропонований Тін Кам Хо. Random forest застосовується для розв'язання задач класифікації, регресії та кластеризації. Основна ідея полягає у використанні великого ансамблю вирішальних дерев, кожне з яких окремо має невисоку якість класифікації, але завдяки їх великій кількості досягається висока точність результату [51].

RF (random forest) представляє собою сукупність вирішальних дерев. У задачах регресії їх відповіді усереднюються, тоді як у задачах класифікації рішення приймається шляхом голосування більшості. Усі дерева будуються незалежно за наступною схемою:

- Обирається підвибірка з навчальної вибірки розміру l для побудови кожного дерева (для кожного дерева - своя підвибірка).
- Для кожного розщеплення в дереві розглядається *max_features* випадкових ознак (для кожного нового розщеплення – свої випадкові ознаки).
- Обирається найкраща ознака для розщеплення за заздалегідь визначеним критерієм. Дерево будується до вичерпання підвибірки (поки в листі не залишаться лише представники одного класу), але сучасні реалізації мають параметри, що обмежують висоту дерева, кількість об'єктів в листі та кількість об'єктів в підвибірці, при якому проводиться розщеплення.

Ця схема побудови відповідає головному принципу ансамблювання: базові алгоритми повинні бути ефективними і різноманітними, тому кожне дерево будується на своїй навчальній вибірці та з використанням елемента випадковості при виборі розщеплення. Чим більше дерев у ансамблі, тим вища якість, однак час налаштування і роботи random forest також пропорційно збільшується. Зі збільшенням кількості дерев, якість на навчальній вибірці може навіть досягати 100%, тоді як якість на тестовій вибірці наближається до асимптоти [52].

Метод випадкового лісу [31] заснований на методі вирішальних дерев. Випадковий ліс – це сукупність вирішальних дерев, а клас об'єкта, що проходить класифікацію, визначається шляхом голосування більшості.

Згенеруємо випадкову вибірку S розміру l з вихідної навчальної вибірки $D = \{(x_i, y_i)\}_{i=1}^l$.

За вибіркою S індукуємо неусічене дерево рішень T_i з мінімальною кількістю спостережень в термінальних вершинах, рівним n_{\min} , рекурсивно виконуючи наступний підалгоритм:

- 1) Випадково вибираємо p ознак з вихідного набору n ознак.
- 2) Обираємо ознаку, яка забезпечує найкраще розщеплення.
- 3) Розщеплюємо вибірку, відповідну до оброблюваної вершини, на дві підвибірки.

В результаті отримуємо ансамбль дерев рішень $\{T_i\}_{i=1}^B$.

Класифікація нових спостережень здійснюється наступним чином: нехай $\hat{y}_i(x) \in \{y_1, \dots, y_l\}$, - клас, передбачений деревом рішень T_i , тобто $T_i(x) = \hat{y}_i(x)$. Тоді $\hat{y}_{rf}(x)$ - клас, який найчастіше зустрічається у множині $\{\hat{y}_b(x)\}_{i=1}^B$.

Для кращого розуміння як це працює для регресії можна подивитись на схему на рис. 2.3. На відміну від алгоритму градієнтного бустингу, випадкові дерева є паралельними і генерують окремі незалежні прогнози, які між собою не пов'язані і які потім в кінці враховуються в остаточному прогнозі.

Важливі особливості випадкового лісу:

- Різноманітність: під час створення окремого дерева враховуються не всі атрибути/змінні/функції; кожне дерево окреме і різне.
- Розпаралелювання: кожне дерево створюється незалежно з різних даних і атрибутів.
- Розподіл навчання та тестування: у випадковому лісі нам не обов'язково розділяти дані для навчання та тестування, оскільки завжди буде 30% даних, які не бачитимуть дерево рішень.
- Стабільність: стабільність виникає через те, що результат ґрунтується на більшості голосів/усередненні [39].

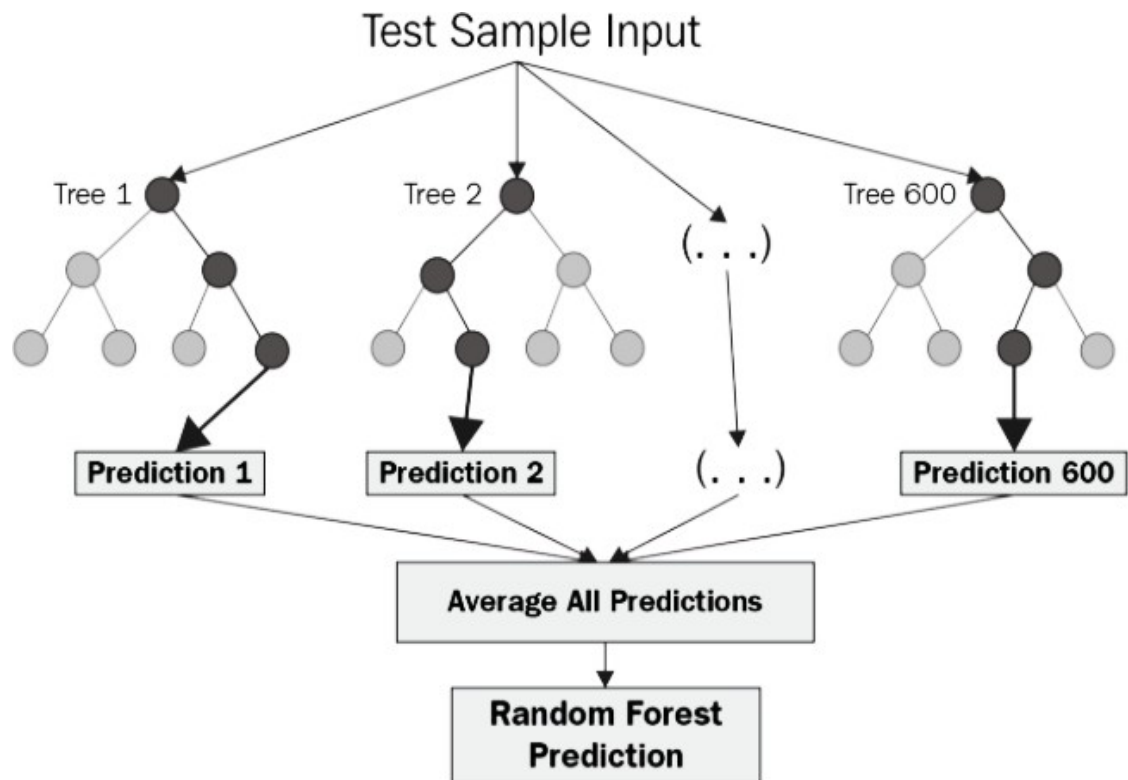


Рисунок 2.3. Принцип роботи алгоритму випадкового лісу в задачах регресії
Джерело: [43]

2.2.3 Model Stacking

Model Stacking (з англ. - «багаторівневе моделювання») - це техніка ансамблевого машинного навчання, що поєднує передбачення кількох базових моделей для покращення загальної продуктивності. Основна ідея полягає в тому, щоб навчити метамодель (або стекінг-модель) комбінувати передбачення від базових моделей з метою мінімізації похибок [44].

Використання стекування складається з двох рівнів. На першому рівні тренуються кілька базових моделей на одних і тих же даних, але з використанням різних алгоритмів або різних підвибірок даних. Базові моделі можуть включати будь-які алгоритми, такі як лінійна регресія, дерева рішень, нейронні мережі тощо.

На другому рівні метамодель навчається на передбаченнях, зроблених базовими моделями, як на вхідних даних. Метамодель приймає передбачення

базових моделей як свої ознаки та навчається комбінувати їх у такий спосіб, щоб зменшити загальну помилку ансамблю. Зазвичай для метамоделі обирають прості алгоритми, такі як лінійна регресія або логістична регресія, однак можуть використовуватися й інші складніші моделі.

Схема побудови Model Stacking наступна [45]:

1. Розділяємо початкову навчальну вибірку на кілька підвбірок.
2. Тренуємо базові моделі на кожній підвбірці.
3. Використовуємо базові моделі для передбачення на валідаційній підвбірці, отримуючи передбачення, що слугують новими ознаками для метамоделі.
4. Навчаємо метамодель на передбаченнях базових моделей, використовуючи їх як вхідні дані.
5. Для тестової вибірки отримуємо передбачення від базових моделей і використовуємо ці передбачення як вхідні дані для метамоделі, яка формує кінцеве передбачення.

Нехай X - початкова навчальна вибірка, y - вектор цільових значень. Базові моделі M_1, M_2, \dots, M_K тренуються на X і генерують передбачення $\widehat{y}_1, \widehat{y}_2, \dots, \widehat{y}_K$. Метамодель M_{meta} тренується на матриці передбачень $H = [\widehat{y}_1, \widehat{y}_2, \dots, \widehat{y}_K]$ і векторі цільових значень y .

Формально, для кожного базового алгоритму M_i отримуємо передбачення:

$$\widehat{y}_i = M_i(x), \quad (2.14)$$

де \widehat{y}_i - передбачення базової моделі M_i .

Метамодель M_{meta} тренується на передбаченнях базових моделей:

$$M_{meta}(H) = M_{meta}([\widehat{y}_1, \widehat{y}_2, \dots, \widehat{y}_K]) = \widehat{y}_{meta}, \quad (2.15)$$

де \widehat{y}_{meta} - кінцеве передбачення метамоделі.

Переваги Model Stacking полягають у тому, що він дозволяє об'єднувати різні моделі, що можуть вловлювати різні аспекти даних, тим самим підвищуючи

загальну точність прогнозування. Однак, даний метод потребує належного налаштування, включаючи правильний вибір базових моделей і метамоделі, а також механізму поділу даних на навчальні та валідаційні підвибірки.

Недоліками стекування є підвищена складність моделювання та обчислювальні витрати, адже потрібно навчати кілька моделей та додатково метамодель. Також важливо уникнути перенавчання, яке може виникнути через надмірну кількість базових моделей або неправильне налаштування метамоделі. Однак, при правильному використанні, *stacking* здатен значно покращити продуктивність системи машинного навчання.

Отже, у стекінгу моделей не використовується лише одна модель для прогнозування. Натомість прогнозування здійснюється за допомогою кількох різних моделей, а потім ці прогнози використовуються як вхідні дані для метамоделі вищого рівня. Цей підхід особливо ефективний, коли залучаються різні типи базових моделей, кожна з яких має свої переваги. Існує багато способів створення стеків моделей, і немає єдиного «правильного» підходу до їх використання. Базова схема стекінгу моделей, що використовується в третьому розділі, представлена на рис. 2.4:

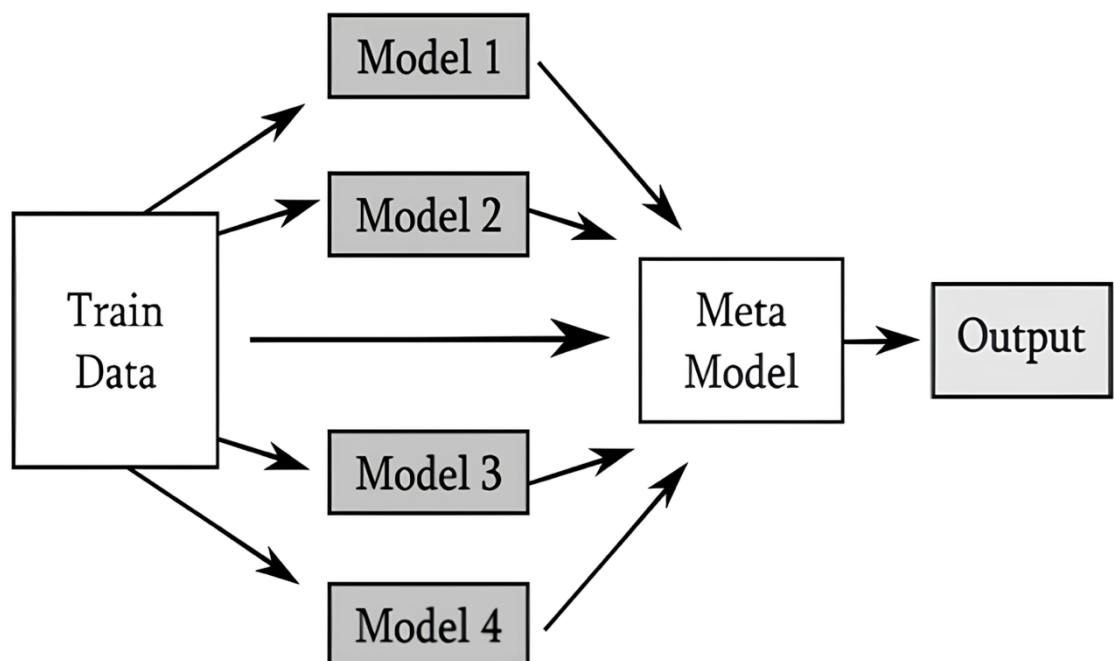


Рисунок 2.4. Принцип роботи алгоритму стекування моделей

Джерело: [45]

2.3 Розробка моделі машинного навчання для аналізу ринку нерухомості

Використаємо для аналізу ринку нерухомості програмне забезпечення SPSS Modeler та його вбудовану методику інтелектуального аналізу даних (Data Mining) – CRISP-DM.

Методологія CRISP-DM включає 6 етапів, де дані є центральним елементом, що об'єднує їх у ітеративний процес.

1. *Розуміння бізнесу*: Визначення результатів проекту, оцінка поточної ситуації та розробка плану проекту з врахуванням бізнес-потреб.

2. *Розуміння даних*: Опис та вивчення доступних даних, перевірка їх якості, виявлення можливих проблем, таких як неправильне чи неповне збереження даних у хмарі або базі даних.

3. *Підготовка даних*: Відбір та очищення даних (включаючи заповнення пропущених значень та видалення нетипових спостережень), об'єднання різних джерел даних та агрегація значень.

4. *Моделювання*: Вибір технік моделювання (наприклад, регресія, нейронні мережі), розбиття даних на набори для навчання (train, що містить 75% даних) та тестування (test, що містить 25% даних), побудова моделі та оцінка її ефективності.

5. *Оцінювання*: Перевірка результатів моделі на нових даних протягом 2-3 місяців, перегляд та затвердження моделей.

6. *Деплоймент (розгортання)*: Інтеграція результатів моделей у щоденну ділову активність.

Діаграма процесу підкреслює той факт, що ітерація є нормою, а не винятком. Це означає, що може бути необхідно пройти через процес кілька разів, щоб досягти оптимального результату. Наприклад, під час етапу моделювання може виявитися, що не вистачає розуміння певного бізнес-аспекту, який доцільно обговорити з власником процесу або іншими ключовими учасниками.

CRISP-DM (Cross-Industry Standard Process for Data Mining) є стандартним процесом, який може бути застосований для аналізу ринку нерухомості. Підхід CRISP-DM передбачає кілька етапів (Табл. 2.1)

Таблиця 2.1

Етапи аналізу ринку нерухомості

№	Назва етапу	Сутність етапу
1	Формулювання та аналіз бізнес-проблеми	Першим кроком у процесі аналізу ринку нерухомості є ретельне вивчення бізнес-проблеми або цілей. Це може включати в себе визначення потреб клієнтів, прогнозування тенденцій на ринку або ідентифікацію можливостей для інвестицій
2	Збір та аналіз первинних даних для формування датасету	Наступним кроком є збір і аналіз наявних даних про ринок нерухомості. Це може включати в себе цінову інформацію, характеристики властивостей, географічні дані, дані про населення та економічні показники.
3	Підготовка даних	Після збору даних їх потрібно очистити, перетворити та підготувати для подальшого аналізу. Це може включати видалення аномальних даних, кодування категоріальних змінних та масштабування чи нормалізація числових даних.
4	Моделювання	На цьому етапі використовуються різні методи аналізу даних, такі як регресійний аналіз, класифікація або кластерний аналіз, для розробки моделей, які можуть передбачати ціни на нерухомість, ідентифікувати тенденції на ринку та ідентифікувати потенційні можливості
5	Оцінка результатів моделювання	Після побудови моделей їх ефективність оцінюється за допомогою відповідних метрик, таких як середньоквадратична помилка або точність прогнозування.
6	Впровадження	Результати аналізу ринку нерухомості можуть бути використані для прийняття бізнес-рішень. Наприклад, модель прогнозування цін на нерухомість може бути використана для оптимізації цінової стратегії компанії або для прийняття рішень щодо інвестицій
7	Моніторинг та підтримка	Після впровадження моделей їх ефективність слід моніторити та необхідність оновлення та підтримки. Також можуть здійснюватися корективи в залежності від змін на ринку нерухомості або бізнес-потреб

Джерело: складено автором

Оцінка ефективності класифікатора є основною проблемою в галузі машинного навчання. Модель, отриману під час навчання, можна оцінити за допомогою різних метрик. Ця оцінка має вирішальне значення для вимірювання точності прогнозів і визначення необхідних коригувань обробки даних, вибору гіперпараметрів та інших процесів прийняття рішень.

На додаток до таких факторів, як обчислювальна складність і зрозумілість, у цьому підрозділі в першу чергу розглядатимуться критерії, які використовуються для оцінки прогнозної точності математичних моделей.

Щоб оцінити продуктивність моделі та переконатися в її ефективності, необхідно встановити методологію для розділення початкового набору даних для подальшого аналізу. Цього можна досягти шляхом поділу даних на набори для навчання та тестування або використання методів перехресної перевірки.

Концепція поділу наборів даних для навчання та тестування (train test split), відома як метод поділу тестування, передбачає поділ початкового набору даних певним чином. Отриману модель потім перевіряють з окремими даними, щоб оцінити її здатність узагальнювати нові, невідомі дані. Однак виникає компроміс між розміром набору для тестування та набору для навчання. Збільшення розміру набору для тестування дозволяє точніше оцінити продуктивність моделі на незнайомих даних, але це відбувається за рахунок зменшення розміру набору для навчання, що потенційно може вплинути на ефективність навчання моделі.

Крос-валідація, яка також називається перехресною перевіркою, охоплює серію методологій, призначених для оцінки ефективності прогнозної моделі при застосуванні до автентичних ненавчальних даних. Серед цих методів часто використовується k -кратна перехресна перевірка (k -fold cross-validation), коли вибірка випадковим чином розбивається на k рівних сегментів з метою ітераційного навчання та оцінки моделі. Під час кожної ітерації один із k сегментів позначається як набір для тестування, а решта $k-1$ сегментів використовуються для навчання. Цей процес гарантує, що кожен сегмент служить тестовим набором рівно один раз. Згодом зазвичай береться до уваги середнє значення показників продуктивності моделі, отримане з кожної ітерації. Рисунок 2.5 демонструє ідею k -кратної крос-валідації з $k=5$.

Під час навчання моделей доцільно застосувати крос-валідацію, щоб оцінити якість моделі на всьому обсязі навчальних даних перед порівнянням її результатів з іншими моделями на тестовій вибірці. Таким чином,

використовується комбінація обох підходів. Цей процес проілюстровано на рисунку 2.6.



Рисунок 2.5. K-fold cross-validation

Джерело: побудовано автором на основі [11]

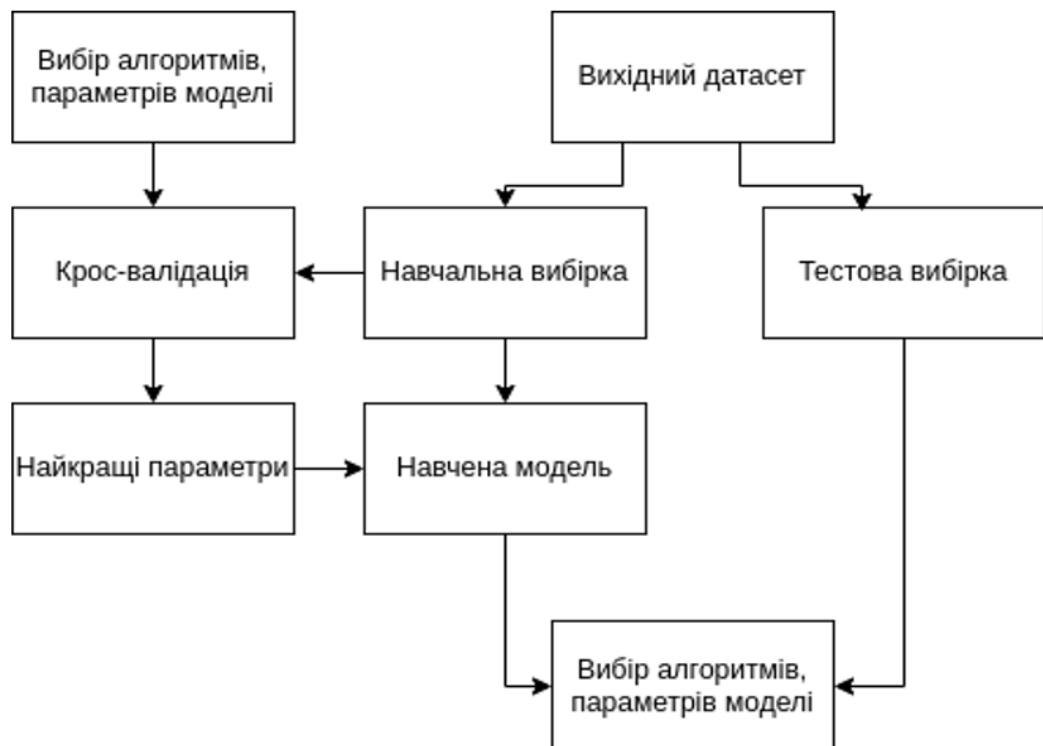


Рисунок 2.6 Стратегія навчання моделі

Джерело: побудовано автором на основі [11]

У свою чергу класифікаційна матриця (confusion matrix) служить інструментом для структурування та відображення результатів моделі класифікації, що дозволяє візуально оцінити ефективність використаного алгоритму. Це дозволяє ідентифікувати кількість точних позитивних прогнозів (True Positive - TP), точних негативних прогнозів (True Negative - TN), неточних позитивних прогнозів (False Positive - FP) і неточних негативних прогнозів (False Negative - FN). Схему заповнення матриці помилок подано на рисунку 2.7.

		PREDICTIVE VALUES	
		POSITIVE (1)	NEGATIVE (0)
ACTUAL VALUES	POSITIVE (1)	TP	FN
	NEGATIVE (0)	FP	TN

Рисунок 2.7 Схema заповнення матриці помилок

Джерело: побудовано автором на основі [11]

Цей інструмент дозволяє оцінити помилки першого роду (FN) та другого роду (FP), а також розрахувати метрики, які більш детально відображають характеристики роботи навченої моделі, на відміну від традиційної точності. Серед таких метрик можна виокремити наступні (2.16-2.18):

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2.16)$$

$$\text{True positive rate (Recall)} = \frac{TP}{FN+TP} \quad (2.17)$$

$$\text{False positive rate} = \frac{FP}{TN+FP} \quad (2.18)$$

Точність є одним з показників для оцінки класифікації моделі. Вона являє собою відношення правильних прогнозів моделі до загальної кількості її прогнозів (2.19).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (2.19)$$

Цей показник відносно простий для розуміння, однак важливо зазначити, що він має обмеження. Слід визнати, що він не підходить для оцінки продуктивності моделі на незбалансованих наборах даних. У ситуаціях, коли один клас непропорційно представлений у наборі даних порівняно з іншим класом (міноритарним класом), показники точності можуть здаватися високими, навіть якщо передбачувана здатність моделі для екземплярів меншинного класу є потенційно недостатньою.

У контексті класифікації транзакцій, зокрема у сфері виявлення шахрайських транзакцій, загальна проблема виникає через дисбаланс даних, коли випадки шахрайства зазвичай є меншістю порівняно з законними транзакціями. Отже, для оцінки ефективності системи виявлення шахрайства для прогнозування шахрайської діяльності необхідно вибрати альтернативний показник.

2.3.1 Показники ROC, AUC

Ще одним показником якості роботи моделі є ROC (Receiver operating characteristic) криві, що ілюструють співвідношення між істинно позитивними та хибно позитивними прогнозами моделі [13].

Ще одним важливим показником якості роботи моделі є ROC (Receiver Operating Characteristic) криві, які ілюструють співвідношення між істинно позитивними та хибно позитивними прогнозами моделі. На рисунку 2.8 представлено графік ROC кривої, де вісь абсцис відповідає FPR (False Positive Rate), а вісь ординат – TPR (True Positive Rate), згадані у попередньому підпункті. Найкращою точкою на цій площині є (0, 1), що означає правильну

класифікацію всіх позитивних прикладів без жодної помилкової класифікації негативних прикладів як позитивних. ROC крива також є надійним методом для визначення потенційно оптимальних класифікаторів [14].

Зазвичай, класифікатор вважається прийнятним, якщо його крива переважно або повністю знаходиться вище лінії $y=x$.

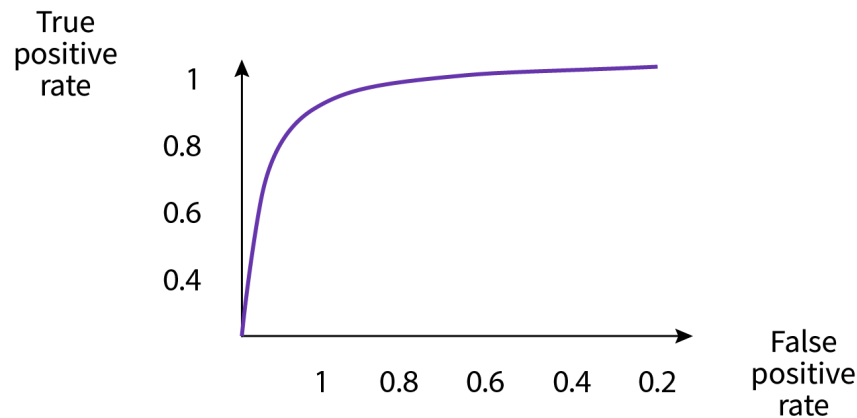


Рисунок. 2.8 Зображення ROC кривої

Джерело: побудовано автором на основі [11]

AUC (Area Under Curve) є площею під графіком ROC кривої. Показник AUC, що дорівнює одиниці, свідчить про те, що модель ідеально класифікує вхідні дані.

Під час інтерпретації результатів роботи моделей варто враховувати, що значення AUC, близьке до 0.5, вказує на те, що прогнози цієї моделі є еквівалентними випадковому розподілу від 0 до 1 і не залежать від вхідних даних.

2.3.2 Покращення результатів роботи моделей

Обравши алгоритми машинного навчання, що можуть бути застосовані до вирішення задачі бінарної класифікації, необхідно також розуміти, що якість їх роботи може також залежати від таких факторів, як попередня підготовка даних та задані гіперпараметри моделі.

Питання прогнозування надійності продавця нерухомості шляхом вивчення відгуків, деталей конкретної забудови (порівняння вартості кв. м житла з

аналогічними об'єктами тощо) та облікових даних продавця є загальною проблемою класифікації. Подібні завдання вирішуються за допомогою систем підтримки прийняття рішень, які використовуються в банківській галузі, зокрема в кредитуванні, які не тільки допомагають зменшити потенційні ризики, але й допомагають визначити відповідний кредитний рейтинг. Так само дана система дозволяє визначити шахрайські дії в процесі інвестування нерухомості, аналізуючи умови фінансування – розмір першого внеску, відсоткова ставка, платіжні системи тощо.

При вирішенні проблем класифікації важливо досягти високої точності тестових зразків. Згідно з методологією CRISP-DM, початковий крок включає підготовку даних, включаючи перевірку та оцінку повноти набору даних. Згодом виконується попередня обробка даних, щоб підготувати набір даних для побудови та навчання аналітичної моделі за допомогою методів машинного навчання. Третій крок передбачає побудову аналітичної моделі для аналізу нових даних, результати якої залежать від обраних методів навчання та аналізу даних. Останнім кроком є перевірка моделі шляхом тестування на окремих даних. Щоб зібрати дані про продавця нерухомості (забудовник чи ріелторська компанія, інвестиційний фонд тощо) використовувався скрапер даних для збору інформації для навчання моделей, наприклад дати реєстрації, статус юридичної документації, номери та категорії об'єктів нерухомості, а також відгуки клієнтів, що містять рейтинги та деталі купівлі. Традиційні методи вирішення таких проблем включають нейронні мережі, логістичну регресію, опорні векторні машини, дерева рішень, випадкові ліси та метод k-найближчих сусідів. Завдяки глобальним дослідженням і експериментам було визначено CRISP-DM як найбільш прийнятний підхід для бінарної класифікації багатовимірних даних з обмеженою вибіркою дослідження.

РОЗДІЛ 3 ПРАКТИЧНЕ ЗАСТОСУВАННЯ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ ДЛЯ ВИЯВЛЕННЯ ШАХРАЙСЬКИХ ДІЙ НА РИНКУ НЕРУХОМОСТІ

3.1 Процес розробки моделі

Розроблена модель побудована на основі методики машинного навчання та аналізу даних Crisp-DM пакету SPSS Modeler 18.1. Cross Industry Standard Process for Data Mining (CRISP-DM) - стандарт, що описує загальні процеси і підходи до аналітики даних, що використовуються в промислових data-mining проектах незалежно від конкретного завдання та індустрії.

Розглянутий у роботі датасет для навчання моделі взято з сервісу датасетів Kaggle. Мною було обрано датасет France Weekly Real Estate Listings 2022-2023 для аналізу ринку нерухомості за 2 роки [12].

Таблиця 3.1

Опис атрибутів датасету

Атрибут	Опис
Unnamed: 0	Індексний стовпець.
bedrooms	Кількість спалень у нерухомості.
businessUnit	Підрозділ, пов'язаний з оголошенням про нерухомість.
city	Місто, де розташована нерухомість.
country	Країна, де розташована нерухомість.
date_of_data_generation	Дата створення даних.
id	Унікальний ідентифікатор оголошення про нерухомість.
isExclusiveness	Вказує, чи є оголошення ексклюзивним (True/False).
isIndividual	Вказує, чи оголошення для індивідуальної нерухомості (True/False).
isNew	Вказує, чи є нерухомість новою (True/False).
latitude	Широта координати розташування нерухомості.
longitude	Довгота координати розташування нерухомості.
price	Ціна нерухомості.
priceVariation.date	Дата останньої зміни ціни.
priceVariation.newPrice	Нова ціна після останньої зміни.
priceVariation.oldPrice	Стара ціна перед останньою зміною.

Атрибут	Опис
publicationId	Ідентифікатор публікації оголошення.
realtyType	Тип нерухомості (наприклад, квартира, будинок).
rooms	Кількість кімнат у нерухомості.
title	Назва оголошення про нерухомість.
transactionType	Тип угоди (наприклад, продаж, оренда).
zipCode	Поштовий індекс місцезнаходження нерухомості.

Джерело: складено автором на основі [12]

Розмірність даних: 1806 записів і 23 стовпці, що дозволяє стверджувати, що обсяг обробленої інформації є досить значним і його результати будуть репрезентативними. Підготовка даних включала наступні етапи:

1. Перевірка, чи всі стовпці мають правильні типи даних.
2. Перетворення колонок з датами на відповідний формат (datetime).
3. Перетворення текстових та булевих даних на числові або категоріальні.
4. Виявлення та видалення дублікатів записів.
5. Виявлення та обробка нетипових значень.

Після виконання наведених вище кроків, дані підготовлені для моделювання.

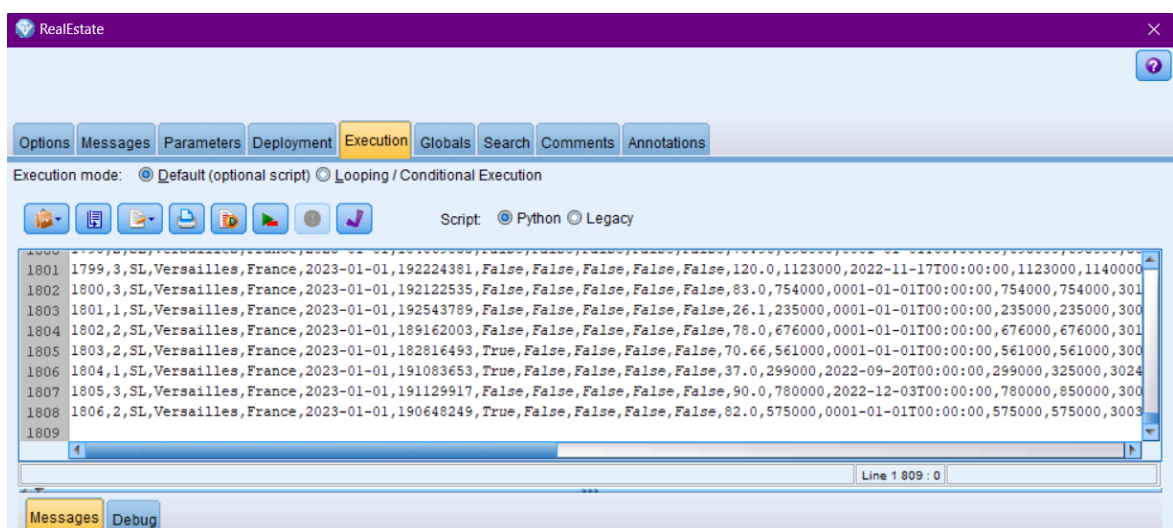


Рис. 3.1. Вигляд завантаженого файлу

Джерело: створено автором у середовищі SPSS Modeler

Розглянемо поведінку деяких змінних датасету. Зокрема цікавим питанням є залежність статусу (шахрайська чи ні) операції від типу квартири. Як видно з рисунку 3.2, дещо частіше шахрайства відбуваються з квартирами з групи С.

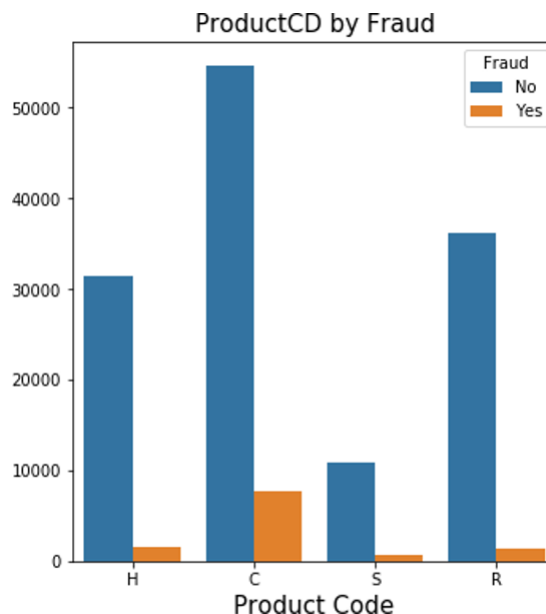


Рисунок 3.2 Належність до шахрайства в залежності від типу об'єкта нерухомості

Джерело: створено автором у середовищі SPSS Modeler

Додавши до цієї картини дані про суму транзакції, з рисунку 3.3 бачимо, що квартири категорії С - це квартири, які в середньому мають більш низьку ціну.

Змінна `transactionDT` представляє дату і час кожної транзакції. Проблема аналізу цієї змінної полягає в тому, що її значення відображають кількість часу (timestamp), що минув не з 1.1.1970, як це зазвичай прийнято в найпоширеніших мовах програмування, а з певного випадкового моменту часу. Визначення цього початкового моменту може сприяти врахуванню сезонності в моделюванні.

Як показано на рисунку 3.4, транзакції з тренувальної та тестової вибірок відбуваються в часових відрізках, що не перетинаються. При цьому всі «тестові» транзакції датовані пізніше, ніж «тренувальні».

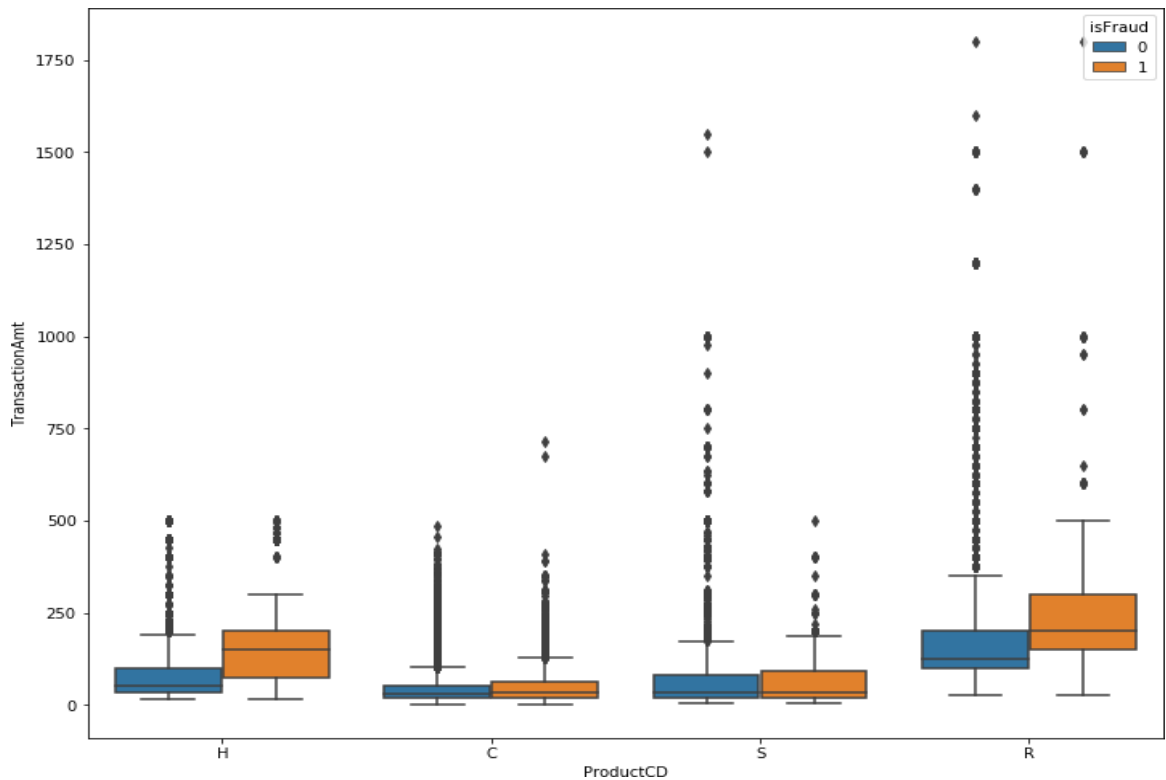


Рис. 3.3 Належність до шахрайства в залежності від суми та розташування
Джерело: створено автором у середовищі SPSS Modeler

Усі транзакції відбуваються впродовж 396 днів, що приблизно дорівнює 13 місяцям. Розглядаючи піки з обох боків графіку, можна припустити, що вони пов'язані із сезонними змінами, коли люди здійснюють більше покупок квартир. Спробуємо перевірити дану гіпотезу. Схожий підхід описаний у [6].

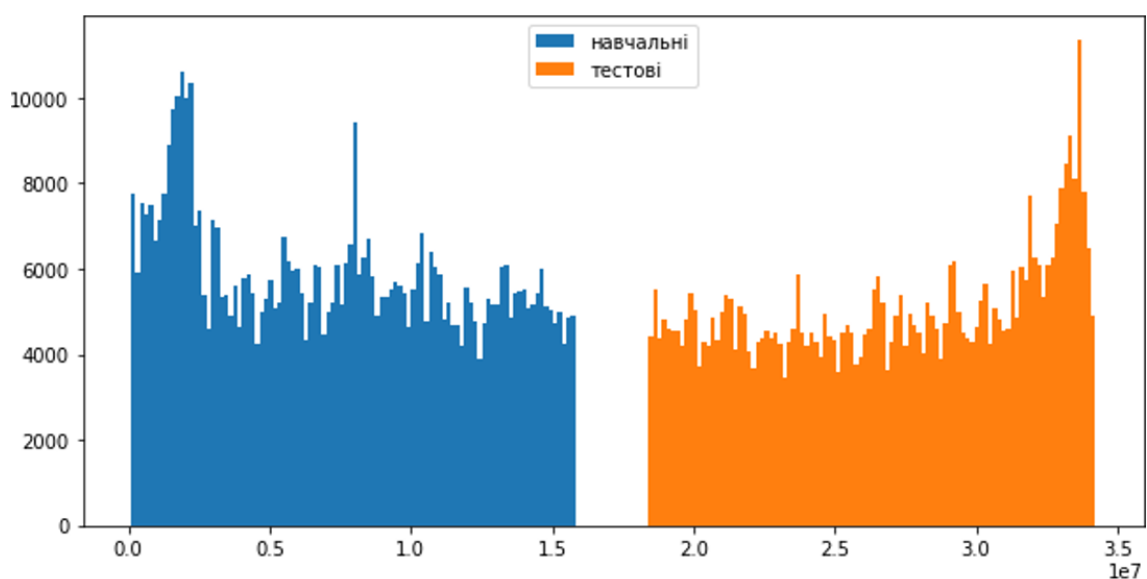


Рисунок 3.4 Дата транзакції в тренувальній та тестовій вибірках
Джерело: створено автором у середовищі SPSS Modeler

Мінімальне значення змінної дорівнює 86400, що збігається із кількістю секунд у одній добі, що підтверджує: transactionDT вимірюється у секундах.

Провівши первинну підготовку даних і проаналізувавши роботу навчених на них моделей, можна імплементувати інформаційну технологію для виявлення шахрайських операцій.

Навчання складається з епох, після кожної з них відбувається оцінка основних параметрів точності нейромережі, тому можна дивитись, як саме відбувається навчання та чи не наступив параліч нейромережі. Як видно, при перших ітераціях навчання нейромережу сильно осцилює, але потім настає впевнений тренд що заключається в постійному падінні середньоквадратичної помилки.

```
Epoch 1/200
14/14 [=====] - 1s 42ms/sample - loss: 0.1179 - accuracy: 0.4007 - val_loss: 0.1008 - val
accuracy: 0.2000
Epoch 2/200
14/14 [=====] - 0s 16ms/sample - loss: 0.1179 - accuracy: 0.4268 - val_loss: 0.1006 - val
accuracy: 0.1902
Epoch 3/200
14/14 [=====] - 0s 16ms/sample - loss: 0.1176 - accuracy: 0.4530 - val_loss: 0.1003 - val
accuracy: 0.1902
Epoch 4/200
14/14 [=====] - 0s 16ms/sample - loss: 0.1176 - accuracy: 0.4564 - val_loss: 0.1001 - val
accuracy: 0.1902
Epoch 5/200
14/14 [=====] - 0s 17ms/sample - loss: 0.1173 - accuracy: 0.4774 - val_loss: 0.0999 - val
accuracy: 0.2000
Epoch 6/200
14/14 [=====] - 0s 15ms/sample - loss: 0.1176 - accuracy: 0.4564 - val_loss: 0.0997 - val
```

Рис. 3.5. Подання процесу навчання у системі

Джерело: створено автором у середовищі SPSS Modeler

На рисунках 3.6 – 3.8 зображені графіки навчання мережі для різних умов, по графіку видно, що навчальна і контрольна помилки із зростанням кількості епох зменшуються.

Отже, розглянувши декілька прикладів навчання моделей на датасеті, можна сказати, що система визначає об'єкт загалом вірно, а отже, модель дійсно гарно навчилася та може бути використана для автоматичного визначення потенційно шахрайських дій.

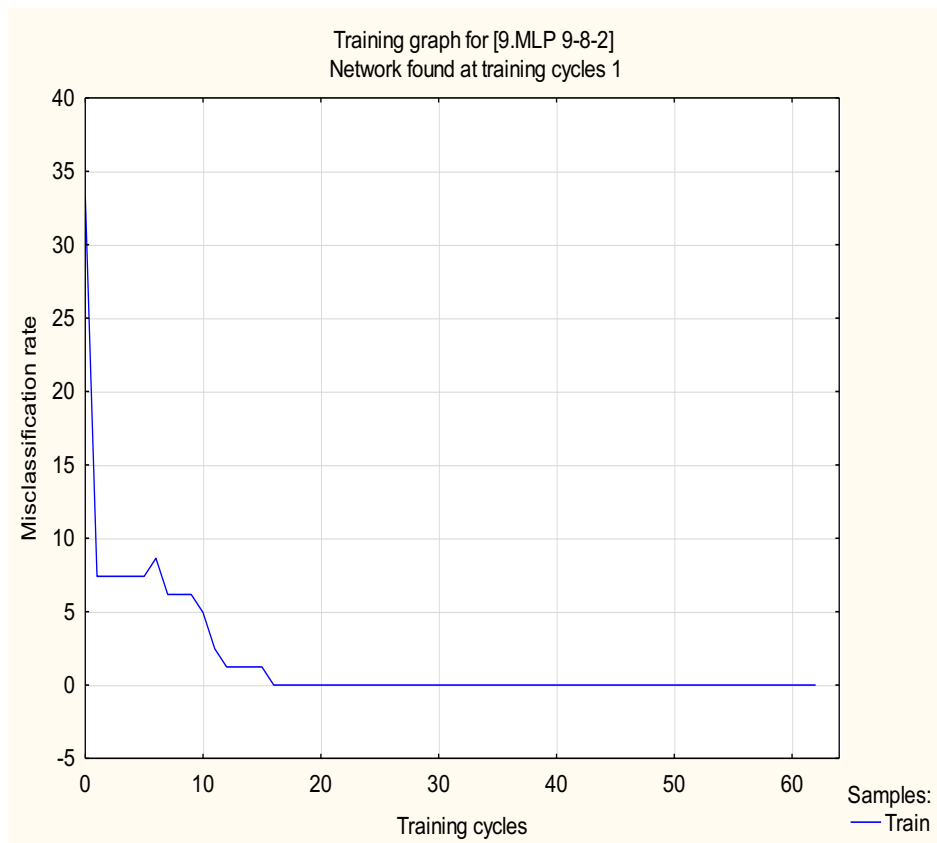


Рис. 3.6. – Графік навчання мережі №3

Джерело: створено автором



Рис. 3.7. – Графік навчання мережі №2

Джерело: створено автором

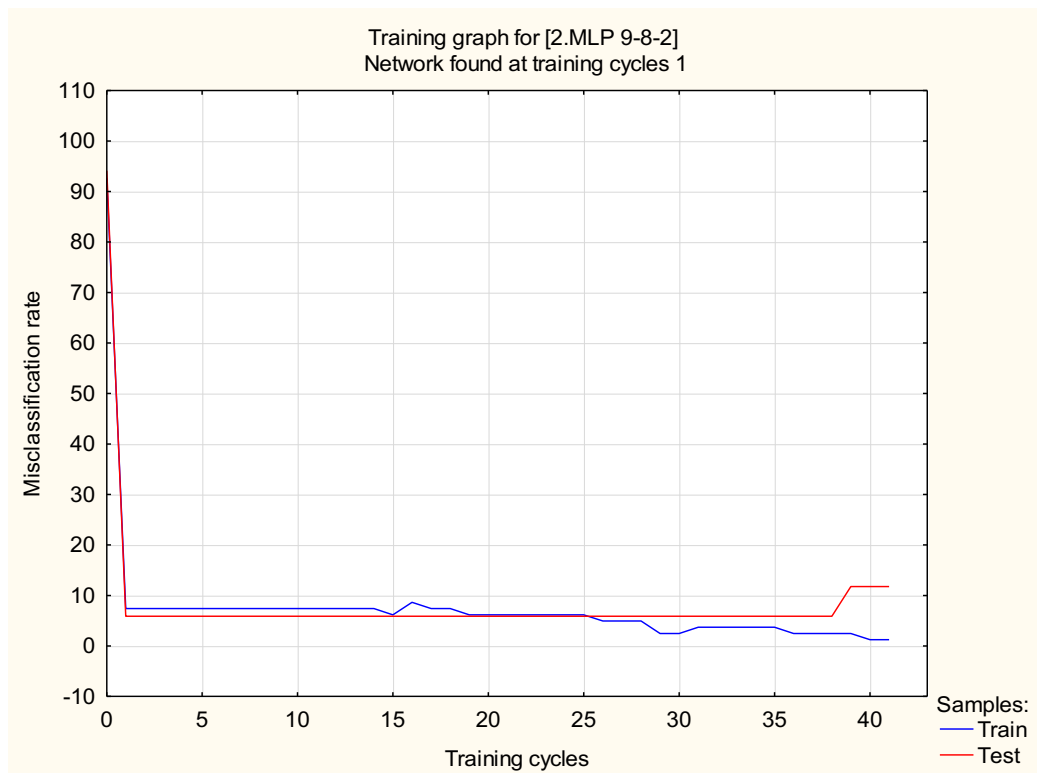


Рис. 3.8. – Графік навчання мережі №1

Джерело: створено автором

3.2 Оцінка ефективності моделей машинного навчання

Класифікація може бути визначена як групування речей за спільними ознаками, характеристиками та якостями. У пакеті SPSS Modeler для класифікації використовується категоріальна вихідна змінна – різні її значення відповідають різним класам.

Ми можемо використовувати різні типи методів класифікації та регресії в SPSS Modeler, в тому числі:

- дерево загальної класифікації та регресії;
- загальна модель CHAID;
- Boosted Tree;
- Random Forest тощо.

Зазвичай, інтерпретація результатів, підсумованих у дереві, дуже проста. Ця простота корисна для швидкої класифікації нових прикладів (об'єктів). Крім того, при аналізі будь-яких проблем (наприклад, бізнес-задач), набагато простіше

представити керівництву кілька простих логічних «якщо-тоді» висловлювань, ніж деякі складні рівняння.

Модуль "Загальні класифікації та регресії дерев" (CART) в SPSS Modeler дозволяє створювати дерева класифікації та регресії для категоричних змінних предикатів (класифікація) та прогнозування неперервних залежностей змінних (регресії). Цей модуль підтримує класичний алгоритм CART і включає в себе різні методи для обрізки та перехресної перевірки, а також потужні крос-перевірки методів, v-fold.

Категоріальні змінні в пакеті SPSS Modeler представляються в бінарному вигляді чи 1 з N варіантів. При проведенні дослідження економічного стану підприємства існують лише два стани класифікації показників: задовольняє нормативові і не задовольняє. Отже, в даному випадку буде використана категоріальна змінна в бінарному вигляді.

Крім того, пакет SPSS Modeler використовує довірчі інтервали як засіб інтерпретації отриманих вихідних значень. Ці порогові значення можна змінити, щоб маніпулювати схильністю мережі до затвердження певного класу з різним ступенем ймовірності.

Для бінарної категоріальної змінної довірчий інтервал впливає на значення цієї змінної, тобто, якщо вихідне значення елемента перевищує поріг прийняття, то вибирається клас 1.0. Якщо вихідне значення лежить нижче порогу, вибирається клас 0.0. Якщо вихідне значення лежить між порогами, клас вважається не визначеним.

У пакеті SPSS Modeler вибір елемента з найвищим вихідним сигналом залежить від умови, що його вихідний сигнал перевищує або дорівнює вказаному порогу прийняття, тоді як вихідні сигнали всіх інших елементів падають нижче визначеного порогу відхилення. Отже, класифікація, визначена цим конкретним елементом, буде обрана як кінцевий результат. Ці принципи справедливі для різних типів мереж, таких як MLP, RBF, лінійні мережі та PNN (хоча в мережах PNN двійкове кодування є неможливим і натомість замінюється кодуванням 1 на N).

Побудова моделі нейронної мережі ведеться адаптивно, виключаючи залучення експертів до процесу навчання. Однак такий підхід вимагає використання спеціального програмного забезпечення. Моделі нейронних мереж мають високий рівень складності та невизначеності, їх часто називають «чорним ящиком», оскільки логіка прийняття рішень залишається прихованою від експертів. Для вирішення цієї проблеми використовується алгоритм «вилучення знань з нейронних мереж», щоб формалізувати набір логічних правил, що дозволяє створити експертну систему на основі мережі [10]. Дані та деякі підходи до розрахунку описані у [16].

На основі отриманої моделі перевіримо операцію з нерухомістю на ринку України на її підозрілість у шахрайстві.

Адреса об'єкта: Київ, Оболонський пр-т, 7-Б, кв. 27. Власник: Дмитрунь Н.І. Довідка БТІ № 22332242. Квартира – приватизована.

Об'єкт розташований біля синьої лінії метро, транспортна мережу розвинена добре. До станції метро «Оболонь» 7 хв. пішки. Поруч розташовуються торгово-розважальний центр «Дрім-таун», школа № 225, поштове відділення № 205, дитсадок № 208, у десятих хвилинах ходьби – магазин «Сільпо», районна стоматологічна поліклініка.

Найближчий мікроринок знаходиться у 7 хвилинах ходьби біля метро «Оболонь».

Аналіз району розташування дозволяє дати гарну оцінку розміщення об'єкту оцінки. Близькість до магазинів, навчальних закладів оцінюється задовільно. Якість комунальних послуг оцінюється задовільно.

Характеристика будинку.

Матеріал стін – панельний. Загальна кількість поверхів – 9. Висота поверхів 2,5 м. Будинок віддалений від проїзної частини, розташоване у дворах.

Рік будівлі – 1975 рік. Загальний вид будинку – хороший.

Водопостачання – центральне, опалення – центральне, вентиляція – природна, енергопостачання – центральне.

Інформації по квартирі:

Двохкімнатна квартира розташована на 5 поверсі будинку. Загальна площа квартири – 47,5 кв. м, житлова площа – 28 кв. м (одна кімната – 18 кв. м, друга – 10 кв. м). Площа кухні і коридору становить по 8 кв. м, санвузол займає 3,5 кв. м. У квартирі є балкон, проведено телефон. Орієнтація вікон – захід. Загальний вид квартири – задовільний.

Отже, аналіз розташування і стан об'єкта у першому наближенні дозволяють казати про досить високу вартість об'єкта.

Оцінка зроблена відповідно до таких допущень:

- спеціальна експертиза інженерних мереж не проводилася;
- об'єкт оцінки не має певними прихованими дефектами, зокрема дефектами несе каркаса будинку, які б спричинити його;
- об'єкт нерухомості відповідає всім санітарним і за екологічними мірками;
- в оцінці використані достовірні джерела інформації.

У табл. 3.2 наведена інформація по ринку про нещодавно проданих об'єктах нерухомості, порівнянних з оцінюваним нежитловим приміщенням.

Таблиця 3.2

Дані про продажі нежитлових приміщень

Аналоги	Розташування	Площа, м ²	Матеріал стін	Стан приміщення	Вартість додаткового обладнання, тис. грн.
A ₁	с/в	49,8	Цегла	Ср	35
A ₂	с/в	46,8	Цегла	Пг	22
A ₃	с/в	46,3	Панель	Ср	18
A ₄	В	48,5	Панель	Ср	25
A ₅	В	47,0	Цегла	Пг	0
A ₆	В	49,5	Цегла	Ср	15
A ₇	Ц	46,0	Цегла	Ср	23
A ₈	Ц	49,3	Цегла	Пг	0
A ₉	Ц	46,3	Панель	Ср	30

Джерело: складено автором на основі [8]

На основі порівняльного аналізу зіставних об'єктів отримаємо скориговані ціни продажу, тобто дев'ять варіантів оціночної вартості розглянутого житлового приміщення.

Результати проведення коректувальних розрахунків відображені в табл. 3.3.

Таблиця 3.3

Корегування цін порівнянних нежитлових приміщень

Об'єкт порівняння	Розташування	Площа, м ²	Матеріал стін	Стан приміщення	Вартість додаткового обладнання, тис. грн.	Ціна продажу*, тис. грн.	Ц _{кх} ** , тис. грн.	Поправка в долях			Підсумкова скоригована ціна, тис. грн..
								За розташуванням	за матеріалом стін	за станом приміщення	
1	2	3	4	5	6	7	8	9	10	11	12
О	с/в	47,5	Цегла	Пг	0	-	-	-	-	-	-
А ₁	с/в	49,8	Цегла	Ср	35	615	553,77	0,85	1,09	1,12	574,63
А ₂	с/в	46,8	Цегла	Пг	22	520	505,99	0,85	1,09	1,12	525,05
А ₃	с/в	46,3	Панель	Ср	18	545	541,24	0,85	1,09	1,12	561,64
А ₄	В	48,5	Панель	Ср	25	465	430,93	0,72	1,09	1,12	378,78
А ₅	В	47,0	Цегла	Пг	0	443	447,71	0,72	1,09	1,12	393,53
А ₆	В	49,5	Цегла	Ср	15	530	494,19	0,72	1,09	1,12	434,38
А ₇	Ц	46,0	Цегла	Ср	23	655	652,61	0,9	1,09	1,12	717,03
А ₈	Ц	49,3	Цегла	Пг	0	580	559,39	0,9	1,09	1,12	614,61
А ₉	Ц	46,3	Панель	Ср	30	605	590,54	0,9	1,09	1,12	648,84

Джерело: складено автором на основі [16]

* Ціни продажів житлових приміщень, порівнянних з оцінюваним об'єктом нерухомості

** Ц_{кх} – ціна продажу аналога, приведена до оцінюваного об'єкта за кількісними характеристиками (за площею та вартістю додаткового обладнання).

$$C_{кхi} = \frac{P_i - C_{обi}}{S_{Pi}} \cdot S_{Pi} \quad (3.1)$$

де Ц_{кхi} – ціна продажу і-го аналога, приведена до оцінюваного об'єкта за кількісними характеристиками (за площею та вартістю обладнання), тис. грн.;

P_i – ціна продажу і-го аналога, тис. грн.;

$C_{обі}$ – вартість обладнання і-го аналога, тис. грн.;

S_{ni} – площа і-го аналога, м².

$$Ц_{кк1} = (615 - 35)/49,8*47,5 = 553,77$$

$$Ц_{кк2} = (520 - 22)/46,8*47,5 = 505,99$$

$$Ц_{кк3} = (545 - 18)/46,3*47,5 = 541,24$$

$$Ц_{кк4} = (465 - 25)/48,5*47,5 = 430,93$$

$$Ц_{кк5} = (443 - 0)/47,0*47,5 = 447,71$$

$$Ц_{кк6} = (530 - 15)/49,5*47,5 = 494,19$$

$$Ц_{кк7} = (655 - 23)/46,0*47,5 = 652,61$$

$$Ц_{кк8} = (580 - 0)/49,3*47,5 = 559,39$$

$$Ц_{кк9} = (605 - 30)/46,3*47,5 = 590,54$$

Дані та розрахунки, які потенційно можуть бути використані при моделюванні описані у [16]. Практика застосування порівняльного методу оцінки нерухомості показує, що розходження в цінах порівнянних нежитлових приміщень за рахунок розташування складають:

1. Між розташованими у віддаленому районі та районі середньої віддаленості – 15%.

2. Між розташованими в центрі міста і районі середньої віддаленості – 10%.

Поправка в частках за матеріалом стін отримана наступним чином:

$$\text{Для району с/в: } 553,77/541,24 = 1,02$$

$$\text{Для району в: } 494,19/430,93 = 1,15$$

$$\text{Для центру: } 652,61/590,54 = 1,11$$

Загальна поправка по місту визначена як середня арифметична отриманих величин і дорівнює 1,09.

Поправка в частках за станом приміщення отримана наступним чином:

$$\text{Для району с/в: } 553,77/505,99 = 1,09$$

$$\text{Для району в: } 494,19/447,71 = 1,1$$

$$\text{Для центру: } 652,61/559,39 = 1,17$$

Загальна поправка по місту визначена як середня арифметична отриманих величин і дорівнює 1,12. У [16] наведено релевантні розрахунки.

Підсумкова скоригована ціна і-го аналога (гр.12, табл.3) визначаються як добуток величини ціни продажу і-го аналога, скоригованої за кількісними характеристиками (C_{kxi}) і загальних поправок по місцю розташування, матеріалу стін і станом приміщення.

Підсумкова скоригована ціна

$$1\text{-го аналога: } 553,77 * 0,85 * 1,09 * 1,12 = 574,63$$

$$2\text{-го аналога: } 505,99 * 0,85 * 1,09 * 1,12 = 525,05$$

$$3\text{-го аналога: } 541,24 * 0,85 * 1,09 * 1,12 = 561,64$$

$$4\text{-го аналога: } 430,93 * 0,72 * 1,09 * 1,12 = 378,78$$

$$5\text{-го аналога: } 447,71 * 0,72 * 1,09 * 1,12 = 393,53$$

$$6\text{-го аналога: } 494,19 * 0,72 * 1,09 * 1,12 = 434,38$$

$$7\text{-го аналога: } 652,61 * 0,9 * 1,09 * 1,12 = 717,03$$

$$8\text{-го аналога: } 559,39 * 0,9 * 1,09 * 1,12 = 614,61$$

$$9\text{-го аналога: } 590,54 * 0,9 * 1,09 * 1,12 = 648,84$$

Розрахуємо середню ціну:

$$\text{Ср.ціна} = (574,63+525,05+561,64+378,78+393,53+434,38+717,03+614,61+648,84)/9 = 538,72 \text{ тис. грн.}$$

Оскільки запропонована агенцією ціна майже вдвічі менша за оціночну, то можна з високою ймовірністю вважати, що дана операція може бути шахрайською і даний об'єкт нерухомості має якісь додаткові характеристики, які впливають на зниження ціни.

Запропонована модель була оцінена за допомогою метрики оцінки точності, яка вимірює частку правильно класифікованих прогнозів. Цей показник вважає істинні позитивні та істинні негативні результати правильно класифікованими, а хибнопозитивні та хибнонегативні результати класифікованими неправильно. Модель рекурентної нейронної мережі оцінювалася за допомогою традиційних класифікаторів машинного навчання, зокрема SVM і Random Forest, у контексті використання кібербезпеки, що включає алгоритми, створені доменом (DGA).

Три моделі було навчено та випробувано на наборі даних, в яких були наявні шахрайські операції. У таблиці 3.4 представлено докладні результати моделей RNN та інших моделей машинного навчання для цього конкретного сценарію.

Таблиця 3.4

Результати тесту різних видів алгоритмів виявлення шахрайських дій

№	Алгоритми	Точність, %	AUC	Gini	Чутливість	Специфічність
1	SVM	0,985	0,989	0,954	0,89	0,89
2	Random forest	0,994	0,961	0,930	0,99	0,95
3	RNN з LSTM	0,997	0,989	0,969	0,99	1,00
4	RNN (шарова нормалізація)	0,994	0,881	0,853	0,88	0,99
5	RNN (пакетна нормалізація)	0,946	0,854	0,822	0,94	0,94

Джерело: складено автором

Тому було проведено дослідження щодо застосування повторюваних нейронних мереж (RNN) у безпеці мережі, зокрема в області виявлення шкідливих програм. Ефективність RNN порівнювали з традиційними класифікаторами машинного навчання, такими як Support Vector Machines (SVM) і Random Forest, для класифікації шкідливих програм у сфері кібербезпеки. Результати цього дослідження демонструють, що RNN перевершує класичні класифікатори машинного навчання з точки зору точності. Це можна пояснити властивою RNN здатністю пам'яті, що дозволяє зберігати та згадувати кілька попередніх станів і витягувати важливі характеристики, а також складні послідовні зв'язки в даних. Отже, RNN має значення для розробки програм реального часу, спрямованих на аналіз шкідливих дій у мережах.

ВИСНОВКИ

Робота присвячена виявленню потенційно шахрайських дій на ринку нерухомості за допомогою методів машинного навчання.

В роботі було поставлено і виконано наступні завдання:

- проведено аналіз поточного стану ринку нерухомості;
- здійснено огляд літератури з поточного стану дослідження проблеми виявлення шахрайських дій на ринку нерухомості;
- визначено типологію підозрілих дій на ринку нерухомості: основні ознаки та критерії виявлення;
- проведено огляд існуючих методик виявлення шахрайських дій на ринку нерухомості;
- досліджено використання моделей машинного навчання для шахрайських дій на ринку нерухомості;
- оглянуто підходи до розробки моделі машинного навчання для аналізу ринку нерухомості;
- здійснено аналіз функціонування розробленої моделі;
- надано оцінку ефективності моделей машинного навчання.

Отже, для роботи з даними про ринок нерухомості було обрано такий інструмент, як нейронні мережі. Процес навчання в нейронних мережах передбачає використання прикладів даних. Користувач нейронної мережі вибирає репрезентативні дані та застосовує алгоритм навчання, який автоматично визначає основні закономірності в даних. Важливо, щоб користувач володів деякими евристичними знаннями щодо вибірки та підготовки даних, вибору архітектури мережі та інтерпретації результатів. Тим не менш, рівень досвіду, необхідний для ефективного використання нейронних мереж, як правило, скромніший порівняно з використанням традиційних статистичних методів.

Нейронні мережі є високоефективним методом моделювання, який може точно фіксувати складні зв'язки. Однією з помітних характеристик нейронних мереж є їхня нелінійна природа. Лінійне моделювання було домінуючим

підходом у різних галузях промисловості протягом багатьох років, оскільки виграє від добре встановлених методів оптимізації. Однак лінійні моделі виявляються неадекватними в ситуаціях, коли лінійна апроксимація не виконується, що є звичайним явищем. Крім того, нейронні мережі вирішують «розмірну проблему», коли лінійні залежності не можуть бути адекватно змодельовані, коли існує велика кількість змінних.

Виконання завдань за допомогою нейронної мережі передбачає підвищений рівень складності. Початковий етап передбачає створення набору даних у форматі .csv, який використовує коми для розділення значень і полегшує організацію табличних даних. Перш ніж завантажувати набір даних у цьому форматі, дуже важливо видалити будь-які небажані елементи шляхом очищення тексту.

По суті, робота з нейронною мережею для вирішення завдань вимагає ряду кроків, таких як створення набору даних, очищення тексту, векторизація, проектування архітектури нейронної мережі та навчання. Ретельне виконання цих кроків може призвести до оптимізації нейронних мереж і отримання точних і надійних результатів.

У процесі виконання практичної частини роботи вихідний датасет було підготовлено для навчання моделей, налаштовано гіперпараметри для алгоритмів машинного навчання, а також порівняно результати роботи отриманих моделей на основі метрики AUC. Результати моделювання використано в запропонованій інформаційній технології для виявлення шахрайських транзакцій на ринку нерухомості.

Подальші дослідження можуть бути спрямовані на покращення прогнозної потужності існуючих моделей і використання інших алгоритмів машинного навчання у наявній інформаційній технології. Запропонована інформаційна технологія також може бути застосована для імплементації системи підтримки прийняття рішень.

Використання інструментів нейромережевого опрацювання текстів дозволяє в автоматизованому режимі оцінити датасет на наявність підозрілої інформації з

точки зору шахрайства (неіснуючі об'єкти, об'єкти з заниженою вартістю, сумнівними правами тощо), а також порівняти дані з раніше опрацьованими на предмет співпадіння. Це дозволяє знайти об'єкти нерухомості, які вже були продані нещодавно, тобто один і той же об'єкт продається надто часто.

Нейромережеві технології дозволяють масштабування, тому в процесі тестування нейромережі можна виділити раніше неопрацьовані механізми ідентифікації і додати їх до переліку інструментів визначення шахрайських дій.

Отже, ми дослідили методики та технології виявлення шахрайських дій та пошуку об'єктів нерухомості за сформованими критеріями в результаті тестування нейромережі. Оскільки технології маскування шахрайських дій постійно змінюються, то нейромережа має бути масштабованою під нові критерії пошуку та механізми обробки даних. Ці питання можуть стати предметом наступних наукових розвідок на цю тематику.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Сорочан А.В. Теоретичні та праксеологічні основи розслідування шахрайства на первинному ринку нерухомості. Дис. к.ю.н. спец. 12.00.09. Дніпро, ДДУВС, 2021. 236 с.
2. Чередник К.О., Чаплинський К.О. Теоретичні засади методики розслідування шахрайства на ринку нерухомості. Херсон, Видавничий дім «Гельветика», 2020. 244 с.
3. Кушнарьова О., Конопельський В. Шахрайство на ринку нерухомості, узагальнення його видів та невід'ємність правопорушення від покарання. 2023. Публічне право № 3 (51). С. 110-116.
4. Дикий О.В. До питання вивчення злочинності на сучасному ринку житлової нерухомості в Україні. Актуальні проблеми держави і права. 2012. С. 586-592.
5. Попова І.М. Розслідування шахрайств, пов'язаних із залученням коштів громадян на будівництво житла: дис канд. юрид. наук. 12.00.09. Нац. акад. внутр. справ. К., 2011. 250 с.
6. Кузьменко С.С. Розслідування шахрайства, пов'язаного з інвестуванням коштів у будівництво об'єктів нерухомості: автореф. дис. ... канд. юрид. наук. 12.00.09. Дніпр. Держ. ун-т. внутр. справ. Д., 2019. 20 с.
7. Опанасенко Н.О. Криміналістична характеристика та основні положення розслідування шахрайства, вчиненого організованими злочинними групами у сфері житлового будівництва: автореф. дис канд. юрид. наук. 12.00.09. Академія адвокатури України. Київ., 2018. 20 с.
8. Портал нерухомості Дім.Ріа. URL: <https://dom.ria.com/uk/> (дата звернення 02.05.2024)
9. Оцінка ринку нерухомості станом на 2023 рік. URL: <https://opendatabot.ua/analytics/realty-2023> (дата звернення 03.05.2024)
10. Кравець П., Киркало Р. Системи прийняття рішень з нечіткою логікою. Вісник НУ «Львівська політехніка». 2009. № 650. С. 115-123.

11. Chornous G., Farenjuk Y., Didenko I. Data Mining for Economists: Course book. – Kyiv: Publishing Lira-K, 2023. – 290 p.
12. Черняк О.І. Інтелектуальний аналіз даних у бізнесі з використанням IBM SPSS Modeler : навч. посібник / О.І. Черняк, Г.О. Чорноус. – Київ : Київський університет, 2020. – 263 с. :
13. France Weekly Real Estate Listings 2022-2023. Kaggle: your machine learning and data science community. URL: <https://www.kaggle.com/datasets/arturdragunov/france-weekly-real-estate-listings-2022-2023/data> (дата звернення: 05.05.2024)
14. Асаул А. М., Брижань І. А., Чевганова В. Я. Економіка нерухомості: підручник. Київ : «Лібра», 2004. 304 с.
15. Гордієнко І.В. Інформаційні системи і технології в менеджменті. 5-ге вид., перероб. і допов. К.: КНЕУ, 2013. 279 с.
16. Оцінка житлової нерухомості. http://4ua.co.ua/economy/yb2ac78b4c53b89521216c37_1.html (дата звернення: 05.05.2024)
17. Свистун Є. А., Даниленко Т. І. Особливості та проблеми функціонування ринку нерухомості. Економіка та регіон. 2011. № 3. С. 34-38.
18. Housingwire: How scammers are using AI to commit new fraud in real estate. URL: <https://www.housingwire.com/articles/how-scammers-are-using-ai-to-commit-new-fraud-in-real-estate/> (дата звернення: 22.05.2024)
19. Transparency International: Corruption in the Real Estate Sector. URL: <https://www.transparency.org/en> (дата звернення: 22.05.2024)
20. International Real Estate Federation (FIABCI): "Real Estate Fraud: A Global Perspective". URL: <https://www.fiabci.org/> (дата звернення: 22.05.2024)
21. Sciencedirect: "The Socio-Economic Determinants of Real Estate Fraud". URL: <https://www.sciencedirect.com/science/article/pii/0094119074900138>
22. National Association of Realtors (NAR) . Fraud Alert: Email Scam Not From NAR Leaders. URL: <https://www.nytimes.com/2023/12/27/realestate/national-association-realtors-real-estate.html> (дата звернення: 23.05.2024)

23. Nguyen J. Is there a correlation between inflation and home prices?. Investopedia. URL:<https://www.investopedia.com/ask/answers/correlation-inflation-houses.asp> (дата звернення: 03.05.2024)
24. The neuro–fuzzy diagnostic model synthesis with hashed transformation in the sequence and parallel mode. URL: <https://ric.zntu.edu.ua/article/view/101022/96247>(дата звернення: 02.05.2024).
25. Usability Determination Using Multistage Fuzzy System. Sciencedirect: web–site. URL: <https://www.sciencedirect.com/science/article/pii/S1842>(дата звернення: 08.05.2024).
26. A novel fuzzy decision–making system for CPU scheduling algorithm. URL: <https://link.springer.com/article/10.1007/s00521-015-1987-8>(дата звернення: 12.05.2024).
27. Network–based output tracking control for T–S fuzzy systems using an event–triggered communication scheme. URL: <https://www.sciencedirect.com/science/article/pii/S0165011032>(дата звернення: 21.05.2024).
28. Expert evaluation model of the computer system diagnostic features. URL: <https://ieeexplore.ieee.org/document/7027101/metrics>(дата звернення: 29.05.2024).
29. Diagnosing computer hardware failures using expert system (rule–based technique). URL: https://www.researchgate.net/publication/79205502_DIAGNOSING_COMPUTER_HARDWARE_FAILURES_USING_EXPERT_SYSTEM_RULEBASED_TECHNIQUE (дата звернення: 01.05.2024).
30. Computer Aided Development of Fuzzy, Neural and Neuro–Fuzzy Systems. URL: https://www.researchgate.net/publication/312590719_Computer_Aided_Development_of_Fuzzy_Neural_and_Neuro-Fuzzy_Systems (дата звернення: 03.06.2024).
31. A fuzzy expert system for automatic seismic signal classification. URL: <https://www.sciencedirect.com/science/article/pii/S09574114053> (дата звернення: 01.06.2024).

32. Classification of Network Traffic Using Fuzzy Clustering for Network Security. URL: https://link.springer.com/chapter/10.1007/978-3-319-62701-4_22 (дата звернення: 02.06.2024).

33. Abadeh M., Habibi L., Kortos N. Intrusion Detection Using a Fuzzy Genetics-Based Learning. Deli, 2007. P. 314–318.

34. Simulation and Model-Based Design. Mathworks: web-site. URL: <https://www.mathworks.com/simulink.html>(дата звернення: 16.05.2024).Rabotiahov, A., Kobylin, O., Dudar, Z., & Lyashenko, V. (2018, February). Bionic image segmentation of cytology samples method. In 2018 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET) (pp. 665-670). IEEE.

35. Lyashenko, V., Mohammad, A., & Kobylin, O. (2015). Experiments with Fusion of Images with Use of Wavelet Transformation in Problems of the Text Information Analysis.

36. Kobylin, O., Vyskrebentseva, S., & Petrova, R. (2019). Обробка даних, що містять пропуски в задачах кластеризації. Системи управління, навігації та зв'язку. Збірник наукових праць, 5(57).

37. Kobylin, O. et al. Video Clustering via Multidimensional Time-Series Analysis. In Proceedings of the 9th International Conference on Information Management and Engineering. 2017. pp. 60-63.

38. Mashtalir S., Mashtalir V., Stolbovyi M.. Representative Based Clustering of Long Multivariate Sequences with Different Lengths. In 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP). 2018. pp. 545-548.

39. Bodyanskiy Y., Kobylin I., Rashkevych Y., Vynokurova O., Peleshko D. Hybrid fuzzy-clustering algorithm of unevenly and asynchronously spaced time series in computer engineering. In 2018 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET). 2018. pp. 930-935.

40. Bodyanskiy Y., Vynokurova O., Kobylin, I, Kobylin, O. Adaptive fuzzy clustering of short time series with unevenly distributed observations in Data Stream Mining tasks. *Information Technology and Management Science*, 2016. 19(1), 23-28.
41. Lyashenko V., Kobylin O., Selevko O. (2020) Wavelet Analysis and Contrast Modification in the Study of Cell Structures Images. *International Journal of Advanced Trends in Computer Science and Engineering*. 9(4). - 4701-4706.
42. Mashtalir, V., Ruban, I., & Levashenko, V. (Eds.). (2019). *Advances in Spatio-Temporal Segmentation of Visual Data (Vol. 876)*. Springer Nature.
43. 58. Daradkeh Y.I., Gorokhovatskyi V., Tvoroshenko I., and Zeghid M. (2022) Cluster representation of the structural description of images for effective classification, *Computers, Materials & Continua*, 73 (3), pp. 6069-6084..
44. Tvoroshenko I.S., and Gorokhovatsky V.O. (2020) Effective tuning of membership function parameters in fuzzy systems based on multi-valued interval logic, *Telecommunications and Radio Engineering*, 79(2), pp. 149-163.
45. Tvoroshenko I.S., and Kramarenko O.O. (2019) Software determination of the optimal route by geoinformation technologies, *Radio Electronics Computer Science Control*, 3, pp. 131-142.
46. Gorokhovatskyi V., and Tvoroshenko I. (2020) Image Classification Based on the Kohonen Network and the Data Space Modification, In *CEUR Workshop Proceedings: Computer Modeling and Intelligent Systems (CMIS-2020)*, 2608, pp. 1013-1026.
47. Tvoroshenko I., and Tkachenko D. (2020) Mechanisms of image classification based on descriptors of local features, *Abstracts of IV International Scientific and Practical Conference «Integration of scientific bases into practice» (October 12-16, 2020)*. Stockholm, Sweden, pp. 443-448.
48. Творошенко, І. С. (2018). Особливості застосування сучасних принципів штучного інтелекту до розробки ефективних механізмів моделювання складних систем. *Science and Technology of the Present Time: Priority Development Directions of Ukraine and Poland*, 118-121.

49. Gorokhovatskyi, V., Rusakova, N., and Tvoroshenko, I. (2020) The application of image analysis methods and predicate logic in applied problems of magnetic monitoring, *Telecommunications and Radio Engineering*, 79(20), pp. 1801-1811.

50. Гороховатський В.О., Творошенко І.С. (2021). Методи інтелектуального аналізу та оброблення даних: навч. посібник.

51. Tvoroshenko Irina, Ahmad M. Ayaz, Mustafa Syed Khalid, Lyashenko Vyacheslav, and Alharbi Adel R. (2020) Modification of Models Intensive Development Ontologies by Fuzzy Logic, *International Journal of Emerging Trends in Engineering Research*, 8(3), pp. 939-944.

52. Random forest. URL: https://ru.wikipedia.org/wiki/Random_forest.

53. Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(01), 20-28.

Київський національний університет імені Тараса Шевченка

Економічний факультет

Кафедра економічної кібернетики

ЗАВДАННЯ

на кваліфікаційну роботу бакалавра

студентки 4 курсу спеціальності 051 «Економіка», ОПП «Економічна кібернетика»

Дікої Данії Василівни

1. Тема роботи: виявлення шахрайських дій на ринку нерухомості за допомогою моделей машинного навчання.
2. Термін завершення роботи: 02.06.2024.
3. Попередній захист роботи: 03.06.2024.
4. Об'єкт дослідження: процеси управління на ринку нерухомості.
5. Предмет дослідження: системи виявлення шахрайських дій на ринку нерухомості на основі нейромереж.
6. Мета дослідження: розробити методіку виявлення шахрайських дій на ринку нерухомості за допомогою методів машинного навчання.
7. Завдання дослідження:
 - 1.1 Провести аналіз поточного стану ринку нерухомості;
 - 1.2 Здійснити огляд літератури з поточного стану дослідження проблеми виявлення шахрайських дій на ринку нерухомості;
 - 1.3 Визначити типологію підозрілих дій на ринку нерухомості: основні ознаки та критерії виявлення;
 - 1.4 Провести огляд існуючих методик виявлення шахрайських дій на ринку нерухомості;
 - 1.5 Дослідити використання моделей машинного навчання для шахрайських дій на ринку нерухомості;
 - 1.6 Оглянути підходи до розробки моделі машинного навчання для аналізу ринку нерухомості;

- 1.7 Здійснити аналіз функціонування розробленої моделі;
- 1.8 Надати оцінку ефективності моделей машинного навчання.

Науковий керівник: доктор економічних наук, професор Черноус Галина
Олександрівна

Студент: Дікая Данія Василівна

Затверджено на засіданні кафедри економічної кібернетики
протокол № 4 від 22.11.2023 р.

Календарний план виконання кваліфікаційної роботи бакалавра

№	Етапи роботи	Терміни виконання	Відмітка керівника про виконання
1	Вибір теми кваліфікаційної роботи бакалавра.	До 22.11.2023	Виконано
2	Розробка та затвердження завдання кваліфікаційної роботи бакалавра.	До 22.11.2023	Виконано
3	Аналіз стану ринку нерухомості Написання підрозділу 1.1.	До 08.03.2024	Виконано
4	Дослідження поточного стану дослідження проблеми виявлення шахрайських дій на ринку нерухомості. Написання підрозділу 1.2.	До 15.03.2024	Виконано
5	Опис типології підозрілих дій на ринку нерухомості: основні ознаки та критерії виявлення Написання підрозділу 1.3.	До 22.03.2024	Виконано
6	Огляд існуючих методик виявлення шахрайських дій на ринку нерухомості. Написання підрозділу 2.1.	До 12.04.2024	Виконано
7	Дослідження моделей машинного навчання що використовуються для виявлення шахрайських дій на ринку	До 19.04.2024	Виконано

	нерухомості. Написання підрозділів 2.2, 2.3.		
8	Опис процесу розробки моделі. Написання підрозділу 3.1.	До 16.05.2024	Виконано
9	Оцінка ефективності моделей машинного навчання. Написання розділу 3.2.	До 23.05.2024	Виконано
13	Оформлення вступу та висновків, джерел, додатків. Форматування файлу роботи.	До 30.05.2024	Виконано
14	Подання роботи для перевірки на плагіат.	До 31.05.2024	Виконано
15	Подання роботи до попереднього захисту.	До 03.06.2024	Виконано

Науковий керівник: доктор економічних наук, професор

Чорноус Галина Олександрівна

Студент: Дікая Данія Василівна