Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет інформаційно-аналітичних технологій та менеджменту
(повна назва)

Кафедра прикладної математики
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Математичні методи виявлення
фейкових акаунтів
у соціальних мережах
(тема)

Виконав:
здобувач 2 року навчання, групи САУм-24-1
Максим ГАЛУШКОВ
(Власне ім'я, ПРІЗВИЩЕ)
Спеціальність
124 Системний аналіз
(код і повна назва спеціальності)
Тип програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)
Освітня програма
Системний аналіз і управління
(повна назва освітньої програми)
Керівник доцент Валентин ЄСІЛЕВСЬКИЙ
(посада, Власне ім'я, ПРІЗВИЩЕ)

Допускається до захисту

Завідувач кафедри ПМ _____ Максим СИДОРОВ
(підпис) (Власне ім'я, ПРІЗВИЩЕ)

2025 р.

Харківський національний університет радіоелектроніки

Факультет _інформаційно-аналітичних технологій та менеджменту
Кафедра _прикладної математики
Рівень вищої освіти _другий (магістерський)
Спеціальність _124 Системний аналіз
(код і повна назва)
Тип програми _освітньо-професійна
(освітньо-професійна або освітньо-наукова)
Освітня програма _Системний аналіз і управління
(повна назва)

ЗАТВЕРДЖУЮ:

Завідувач кафедри _____
(підпис)
"_10_" _листопада_ 2025 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві _Галушкову Максиму Олександровичу
(прізвище, ім'я, по батькові)

1. Тема роботи _Математичні методи виявлення фейкових акаунтів у соціальних мережах

затверджена наказом по університету від 10 листопада 2025 р. № 1027 Ст

2. Термін подання здобувачем роботи до екзаменаційної комісії 18 грудня 2025 р.

3. Вихідні дані до роботи __дані що характеризують функціонування соціальних мереж

4. Перелік питань, що потрібно опрацювати в роботі
1. Системний аналіз предметної області та постановка задач дослідження
2. Вибір і обгрунтування методів дослідження
3. Програмна реалізація методів виявлення фейкових акаунтів
4. Результати обчислювального експерименту
5. Аналіз можливих застосувань

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій

1. Представлення соціальної мережі у вигляді графа
2. Основні проблеми спричиненні фейковими акаунтами у соціальних мережах
3. Порівняння поведінкових патернів реального користувача та бота
4. Типи даних для фейкових акаунтів
5. Результати експериментального дослідження системи детекції фейкових акаунтів
6. Важливість ознак у моделі Random Forest

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Підбір та вивчення технічної літератури за темою роботи	10 – 16 листопада 2025 р.	виконано
2	Вибір та обґрунтування методу	17 – 23 листопада 2025 р.	виконано
3	Розробка алгоритму і програми	24 – 30 листопада 2025 р.	виконано
4	Проведення аналітичних досліджень та розрахунків	01 – 07 грудня 2025 р.	виконано
5	Робота над текстом пояснювальної записки	08 – 17 грудня 2025 р.	виконано
6	Представлення роботи на рецензію в ЕК	18 грудня 2025 р.	виконано

Дата видачі завдання _10_ листопада 2025 р.

Здобувач _____
(підпис)

Керівник роботи _____ доцент Валентин ЄСІЛЕВСЬКИЙ
(підпис) (посада, Власне ім'я, ПРІЗВИЩЕ)

РЕФЕРАТ

Пояснювальна записка: 61 с., 4 табл., 4 рис., 1 дод., 30 джерел.

ВИЯВЛЕННЯ ФЕЙКОВИХ АКАУНТІВ, ГРАФОВІ НЕЙРОННІ МЕРЕЖІ, ДЕТЕКЦІЯ АНОМАЛІЙ, ПОВЕДІНКОВИЙ АНАЛІЗ, СОЦІАЛЬНІ МЕРЕЖІ.

Об’єкт дослідження – процеси функціонування соціальних мереж як складних динамічних систем, моделі поведінки користувачів у цифровому середовищі, методи машинного навчання на основі графових нейронних мереж та алгоритми обробки великих обсягів неструктурованих даних.

Мета роботи – розробка та дослідження математичних методів виявлення фейкових акаунтів у соціальних мережах на основі сучасних алгоритмів машинного навчання, графових нейронних мереж та методів обробки природної мови для підвищення рівня інформаційної безпеки онлайн-платформ та захисту користувачів від зловмисних дій.

Методи дослідження – методи теорії графів для моделювання соціальних мереж як орієнтованих та неорієнтованих графів; алгоритми машинного навчання з вчителем (Random Forest, Support Vector Machine, XGBoost) та без вчителя (k-means, DBSCAN); графові нейронні мережі (Graph Convolutional Networks, Graph Attention Networks, GraphSAGE); трансформерні архітектури для обробки природної мови (BERT, RoBERTa); методи обробки та аналізу часових рядів; статистичні методи аналізу даних та перевірки гіпотез; методи валідації моделей; методи візуалізації багатовимірних даних.

У кваліфікаційній роботі магістра розглянуто проблему виявлення фейкових акаунтів у соціальних мережах як одну з актуальних задач інформаційної безпеки. Проведено системний аналіз предметної області, досліджено сучасні підходи до детекції фейкових акаунтів та проаналізовано можливості застосування графових нейронних мереж.

Обґрунтовано вибір методів машинного навчання та графових моделей для розв’язання задачі класифікації користувачів. Розглянуто алгоритм SybilEdge як ефективний метод раннього виявлення фейкових акаунтів. Розроблено програмну систему детекції та проведено обчислювальний експеримент на синтетичних даних.

Отримані результати роботи можуть бути використані у соціальних мережах, системах кібербезпеки та антифрод-системах для автоматизації процесів виявлення підозрілої активності та підвищення рівня інформаційної безпеки.

ABSTRACT

Introductory note: 61 pages, 4 tables, 4 figures, 1 appendixes, 30 sources.

FAKE ACCOUNT DETECTION, ANOMALY DETECTION, BEHAVIORAL ANALYSIS, GRAPH NEURAL NETWORKS, SOCIAL NETWORKS.

Object of the research – the processes of functioning of social networks as complex dynamic systems, models of user behavior in the digital environment, machine learning methods based on graph neural networks, and algorithms for processing large volumes of unstructured data.

Purpose of the research – the development and investigation of mathematical methods for detecting fake accounts in social networks based on modern machine learning algorithms, graph neural networks, and natural language processing methods in order to increase the level of information security of online platforms and protect users from malicious activities.

Research methods – graph theory methods for modeling social networks as directed and undirected graphs; supervised machine learning algorithms (Random Forest, Support Vector Machine, XGBoost) and unsupervised learning algorithms (k-means, DBSCAN); graph neural networks (Graph Convolutional Networks, Graph Attention Networks, GraphSAGE); transformer architectures for natural language processing (BERT, RoBERTa); methods for time series processing and analysis; statistical methods for data analysis and hypothesis testing; model validation methods; and methods for visualization of multidimensional data.

In the master's qualification thesis, the problem of detecting fake accounts in social networks is considered as one of the urgent tasks of information security. A systematic analysis of the research domain is conducted, modern approaches to fake account detection are investigated, and the applicability of graph neural networks is analyzed.

The choice of machine learning methods and graph-based models for solving the user classification problem is substantiated. The SybilEdge algorithm is considered as an effective method for early detection of fake accounts. A software system for fake account detection is developed, and a computational experiment is conducted using synthetic data.

The obtained results can be applied in social networks, cybersecurity systems, and antifraud systems to automate the detection of suspicious activity and enhance the level of information security.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАК ОДИНИЦЬ І ТЕРМІНІВ

AI – Artificial Intelligence – штучний інтелект;

API – Application Programming Interface – інтерфейс програмування додатків;

AUC – Area Under the Curve – площа під кривою;

BERT – Bidirectional Encoder Representations from Transformers – двонаправлені представлення кодувальника з трансформерів;

CNN – Convolutional Neural Network – згорткова нейронна мережа;

CSV – Comma-Separated Values – значення, розділені комами;

DL – Deep Learning – глибоке навчання;

F1 – F1-Score – F-міра (гармонічне середнє точності та повноти);

GAT – Graph Attention Network – графова мережа з механізмом уваги;

GCN – Graph Convolutional Network – графова згорткова мережа;

GNN – Graph Neural Network – графова нейронна мережа;

GPU – Graphics Processing Unit – графічний процесор;

GraphSAGE – Graph Sample and Aggregate – метод семплування та агрегації для графів;

JSON – JavaScript Object Notation – текстовий формат обміну даними;

LLM – Large Language Model – велика мовна модель;

LSTM – Long Short-Term Memory – довга короткочасна пам'ять;

ML – Machine Learning – машинне навчання;

NLP – Natural Language Processing – обробка природної мови;

NN – Neural Network – нейронна мережа;

RF – Random Forest – випадковий ліс;

RNN – Recurrent Neural Network – рекурентна нейронна мережа;

ROC – Receiver Operating Characteristic – операційна характеристика приймача;

RoBERTa – Robustly Optimized BERT Pretraining Approach – надійно оптимізований підхід попереднього навчання BERT;

SGD – Stochastic Gradient Descent – стохастичний градієнтний спуск;

SVM – Support Vector Machine – метод опорних векторів;

TCN – Temporal Convolutional Network – темпоральна згорткова мережа;

XGBoost – eXtreme Gradient Boosting – екстремальний градієнтний бустинг.

ВСТУП

Актуальність теми. Стрімке зростання кількості фейкових акаунтів у соціальних мережах становить одну з найбільш серйозних загроз сучасному цифровому суспільству. За оцінками експертів з кібербезпеки, від 5% до 15% усіх облікових записів у великих соціальних мережах є підробленими або підозрілими. Це означає, що лише в мережі Facebook може існувати понад 150 мільйонів фейкових акаунтів, а в Instagram та Twitter – десятки мільйонів.

Проблема фейкових акаунтів має багатогранний характер та широкий спектр негативних наслідків для суспільства. По-перше, це розповсюдження дезінформації та маніпулювання громадською думкою. Під час виборчих кампаній та політичних криз фейкові акаунти активно використовуються для поширення фейкових новин, створення ілюзії масової підтримки певних ідей та дискредитації опонентів. По-друге, фейкові профілі є інструментом промислового шпигунства та кіберзлочинності, використовуючись для фішингу, крадіжки персональних даних та поширення шкідливого програмного забезпечення.

По-третє, вони деформують цифрову економіку через накручування популярності, створення штучних рейтингів та підробку відгуків про товари і послуги. За даними дослідження "The Bot Baseline", близько 20% усіх взаємодій у соціальних мережах генеруються автоматизованими ботами, що спотворює реальну картину громадської думки та призводить до прийняття неправильних бізнес-рішень на основі неточних даних.

Традиційні методи виявлення фейкових акаунтів, що базуються на аналізі окремих характеристик профілю (кількість друзів, активність, наявність фото тощо) або простих евристичних правилах, виявляються недостатньо ефективними. Сучасні зловмисники використовують складні схеми створення правдоподібних профілів, імітують природну поведінку користувачів та застосовують великі мовні моделі для генерації переконливого контенту.

У зв'язку з цим надзвичайно актуальною є розробка нових математичних методів та алгоритмів машинного навчання, здатних виявляти тонкі аномалії в поведінці користувачів, аналізувати складні паттерни соціальних зв'язків та адаптуватися до еволюції тактик зловмисників. Графові нейронні мережі (GNN) демонструють особливу перспективність у цій задачі, оскільки соціальні мережі природним чином представляються у вигляді графів, де користувачі є вершинами, а їхні зв'язки – ребрами.

Дослідження останніх років показують, що комбінування графового аналізу з методами обробки природної мови та аналізу часових рядів дозволяє досягти точності виявлення понад 95%, що на 20-30% перевищує результати традиційних підходів.

Мета і завдання кваліфікаційної роботи. Метою кваліфікаційної роботи є розробка та дослідження математичних методів виявлення фейкових акаунтів у соціальних мережах на основі сучасних алгоритмів машинного навчання, графових нейронних мереж та методів обробки природної мови для підвищення рівня інформаційної безпеки онлайн-платформ та захисту користувачів від зловмисних дій. Для досягнення поставленої мети необхідно вирішити наступні завдання:

– провести комплексний огляд і системний аналіз сучасного стану методів виявлення фейкових акаунтів у соціальних мережах, включаючи дослідження останніх наукових публікацій 2024-2025 років;

– дослідити математичні основи, архітектури та принципи роботи графових нейронних мереж (GCN, GAT, GraphSAGE) для задач аналізу структури соціальних мереж та детекції аномалій;

– вивчити сучасні методи машинного навчання (Random Forest, SVM, XGBoost) для класифікації користувачів та виявлення підозрілої поведінки;

– проаналізувати алгоритми обробки природної мови (BERT, RoBERTa) для аналізу текстового контенту та виявлення автоматично згенерованих повідомлень;

– детально дослідити алгоритм SybilEdge як сучасний метод раннього виявлення фейкових акаунтів на основі аналізу графової структури зв'язків;

– розробити програмну систему для виявлення фейкових акаунтів з інтеграцією множинних джерел даних та методів аналізу;

– провести експериментальне дослідження розробленої системи на синтетичних та реальних даних, виконати аналіз результатів та порівняння з існуючими методами.

Об'єктом дослідження є процеси функціонування соціальних мереж як складних динамічних систем, моделі поведінки користувачів у цифровому середовищі, методи машинного навчання на основі графових нейронних мереж та алгоритми обробки великих обсягів неструктурованих даних.

Предметом дослідження є математичні методи, моделі та алгоритми виявлення фейкових акаунтів у соціальних мережах, які базуються на комплексному аналізі профільних даних користувачів, змісту текстових публікацій, часових патернів активності, поведінкових характеристик та графової структури соціальних зв'язків.

Методи дослідження. У роботі використовуються: методи теорії графів для моделювання соціальних мереж як орієнтованих та неорієнтованих графів; алгоритми машинного навчання з вчителем (Random Forest, Support Vector Machine, XGBoost) та без вчителя (k-means, DBSCAN); графові нейронні мережі (Graph Convolutional Networks, Graph Attention Networks, GraphSAGE); трансформерні архітектури для обробки природної мови (BERT, RoBERTa); методи обробки та аналізу часових рядів; статистичні методи аналізу даних та перевірки гіпотез; методи валідації моделей; методи візуалізації багатовимірних даних.

Публікації. Результати, отримані у кваліфікаційній роботі, було представлено на XXVIII Міжнародному молодіжному форумі «Радіоелектроніка та молодь у XXI столітті» (м. Харків, 16–18 квітня 2024 р.) [1].

1 СИСТЕМНИЙ АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧ ДОСЛІДЖЕННЯ

1.1 Огляд методів виявлення фейкових акаунтів у соціальних мережах

Проблема виявлення фейкових акаунтів у соціальних мережах залишається однією з найактуальніших задач сучасної інформаційної безпеки та комп'ютерних наук. Фейкові акаунти можуть бути класифіковані за рівнем автоматизації: повністю автоматизовані боти, напівавтоматизовані (cyborg accounts), та повністю керовані людиною з метою обману (sockpuppet accounts). Кожен тип має свої специфічні характеристики та вимагає різних підходів до виявлення.

Аналіз сучасної літератури дозволяє класифікувати методи виявлення фейкових акаунтів за кількома основними підходами, кожен з яких має свої переваги та обмеження.

Перший та найпростіший підхід базується на аналізі статичних характеристик профілю користувача. До таких характеристик відносяться: кількість публікацій, кількість підписників та підписок, співвідношення між підписниками та підписками, наявність та якість фотографії профілю, довжина та інформативність біографії, дата створення акаунту (вік акаунту), статус верифікації облікового запису.

Дослідження показують, що фейкові акаунти часто мають характерні паттерни у цих метриках. Наприклад, типовий бот може мати велику кількість підписок при малій кількості підписників, відсутність фотографії профілю або використання стокових зображень, мінімальну або порожню біографію, молодий вік акаунту (створений недавно)

Томас К., Грієр К., Ма Дж. та інші [2] у своєму дослідженні виявили, що 77% фейкових акаунтів у Twitter можна ідентифікувати лише на основі базових характеристик профілю з використанням Random Forest класифікатора. Проте цей підхід має суттєві обмеження: зловмисники легко можуть імітувати

нормальні значення цих метрик, створюючи більш правдоподібні профілі (таблиця 1.1).

Таблиця 1.1 – Статичні характеристики профілю користувача та типові ознаки фейкових акаунтів»

Статична характеристика профілю	Реальний користувач	Фейковий акаунт (бот)
Кількість публікацій	Помірна або велика, з часом зростає	Мала або нульова
Кількість підписників	Співмірна з підписками або більша	Дуже мала
Кількість підписок	Обмежена, логічна	Дуже велика
Співвідношення «підписники / підписки»	Близьке до 1 або >1	Значно <1
Фотографія профілю	Реальна, персональна	Відсутня або стокове зображення
Біографія профілю	Заповнена, інформативна	Коротка або порожня
Вік акаунту	Старий (місяці або роки)	Молодий (створений нещодавно)
Верифікація	Може бути відсутня або присутня	Майже завжди відсутня

Другий підхід зосереджується на аналізі динамічної поведінки користувачів. Ключові аспекти поведінкового аналізу включають: часові паттерни публікацій (час доби, дні тижня, регулярність), частоту взаємодій (лайки, коментарі, репости), паттерни використання хештегів та згадок, швидкість реакції на події та тренди, активність у відповідь на інші пости.

Боти, наприклад, часто демонструють надзвичайно регулярні паттерни активності, публікуючи контент з чіткою періодичністю незалежно від часу доби. Чу З., Джанвеккіо С., Ван Х., Джадодія С. [3] розробили систему класифікації користувачів Twitter на людей, ботів та кіборгів (напівавтоматичні акаунти) з точністю 96% на основі аналізу понад 100 поведінкових ознак.

Феррара Е., Варол О., Девіс С. [4] показали, що аналіз часових інтервалів між діями користувача може виявити ботів з точністю 95%, оскільки

автоматизовані системи часто мають характерні "цифрові підписи" у вигляді строго рівномірних або експоненційно розподілених інтервалів.

Третій підхід використовує методи обробки природної мови (NLP) для аналізу текстового контенту, що публікується користувачами. Цей підхід включає: виявлення шаблонних та повторюваних фраз, аналіз лексичного різноманіття текстів, детекцію граматичних помилок та аномалій, аналіз сентименту та емоційного забарвлення, виявлення ознак автоматично згенерованого тексту.

Боти часто використовують обмежений набір шаблонних фраз або генерують тексти з характерними аномаліями. Севідж Д., Чжан С., Ю С. [5] продемонстрували, що аналіз лінгвістичних особливостей дозволяє виявити спам-ботів з точністю 91% навіть коли вони намагаються імітувати стиль людського спілкування.

З появою великих мовних моделей (LLM) таких як GPT-3, GPT-4, Claude, ця задача значно ускладнилася. Сучасні боти можуть генерувати високоякісний, природний текст, що практично не відрізняється від людського. Це вимагає розробки нових методів детекції, що базуються не лише на якості тексту, але й на контекстуальній релевантності, логічній послідовності та інших метаознаках.

Четвертий та один з найперспективніших підходів базується на аналізі графової структури соціальних зв'язків. Соціальна мережа природним чином представляється у вигляді графа

$$G = (V, E),$$

де V – множина користувачів (вершин),

E – множина зв'язків між ними (ребер).

Графовий аналіз дозволяє виявляти: аномальні паттерни зв'язків, ізольовані кластери підозрілих акаунтів, порушення структурних властивостей графа, атаки Sybil (створення множини фейкових ідентичностей).

Цао Ц., Сірівіанос М., Ян С., Прегейро Т. [6] у роботі "Aiding the Detection of Fake Accounts in Large Scale Social Online Services" запропонували SybilRank – алгоритм, що використовує властивості довірених соціальних зв'язків для ідентифікації Sybil-акаунтів. Алгоритм базується на припущенні, що Sybil-акаунти мають обмежену кількість зв'язків з реальними користувачами.

Ван Г., Моханлал М., Вілсон К. [7] розробили SybilDefender, який використовує структурні властивості графа для виявлення фейкових акаунтів з точністю понад 99% навіть коли зловмисники мають значні ресурси для створення штучних зв'язків.

У сучасних дослідженнях все більшого поширення набувають підходи на основі машинного навчання, які використовують ансамблеві методи та багатовимірні профілі ознак користувачів. Зокрема, у роботах Бхаттачар’я А. та Кулкарні А., а також Гоял Б., Гілл Н. С., Гулія П. показано ефективність Random Forest, SVM та градієнтного бустингу для виявлення фейкових профілів у соціальних мережах [8, 9].

1.2 Аналіз графових нейронних мереж для детекції аномалій

Графові нейронні мережі (Graph Neural Networks, GNN) представляють собою клас архітектур глибокого навчання, спеціально розроблених для роботи з даними, що мають графову структуру. На відміну від традиційних нейронних мереж (CNN, RNN), які працюють з регулярними структурами даних (зображення, послідовності), GNN здатні обробляти нерегулярні графові структури, що робить їх особливо придатними для аналізу соціальних мереж.

Основна ідея GNN полягає в агрегації інформації від сусідніх вершин графа для оновлення представлення кожної вершини. Цей процес, відомий як передача повідомлень (message passing), дозволяє моделі враховувати не лише характеристики окремої вершини, але й локальну структуру графа навколо неї.

Формально, процес оновлення представлення вершини v на k-тому шарі GNN можна записати як:

$$h_v^{(k)} = UPDATE^{(k)}(h_v^{(k-1)}, AGGREGATE^{(k)}(\{h_v^{(k-1)} : u \in N(v)\})),$$

де $h_v^{(k)}$ – представлення вершини v на k-тому шарі;

$N(v)$ – множина сусідів вершини v;

$AGGREGATE$ – функція агрегації інформації від сусідів;

$UPDATE$ – функція оновлення представлення вершини.

Graph Convolutional Networks (GCN) продовжують залишатися однією з ключових архітектур у сфері глибокого навчання на графах. У дослідженні Хуана Ч. та Сяна Ц. [10] запропоновано вдосконалену версію GCN, орієнтовану на напівконтрольовану класифікацію вершин із використанням методу "subgraph sketching". Такий підхід дозволяє моделі ефективно працювати з великими графами, зменшуючи обчислювальні витрати та зберігаючи високу точність класифікації. Автори поєднують спектральну згортку з механізмами агрегації на рівні підграфів, що підвищує гнучкість і адаптивність моделі до складної структури даних.

Ключова ідея GCN полягає в тому, що представлення кожної вершини оновлюється як зважена сума представлень її сусідів:

$$H^{(k+1)} = \sigma(D^{-\frac{1}{2}}\tilde{A}D^{-\frac{1}{2}}H^{(k)}W^{(k)}),$$

де $\tilde{A} = A + I$ – матриця суміжності з доданими самопетлями;

D – діагональна матриця степенів вершин;

$H^{(k)}$ – матриця представлень вершин на *k*-тому шарі;

$W^{(k)}$ – навчувані ваги;

σ – функція активації.

GCN також демонструють високу ефективність у задачах виявлення аномалій та підозрілої поведінки у соціальних мережах. У статті Ян С. [11] представлено метод GeneralDyG, який застосовує семплювання его-графів для виявлення аномалій у динамічних графах. Автор показує, що графові моделі, зокрема GCN, дозволяють виявляти скоординовану або підозрілу активність навіть у складних часових структурах, що особливо актуально для боротьби з фейковими акаунтами та спамом у соціальних мережах.

Практичну ефективність графових нейронних мереж у задачах виявлення зловмисних та фейкових акаунтів підтверджено у низці прикладних досліджень. Зокрема, Лю З., Чень С., Ян С. та інші [12] запропонували гетерогенну архітектуру графових нейронних мереж для детекції шкідливих акаунтів, тоді як Нгуєн Х.-Д., Нгуєн К. Д., Фам Х. Л., Куан Т. Т. [13] застосували GNN-підхід для виявлення соціальних ботів з високою точністю.

Графові нейронні мережі з механізмом уваги GAT, як показано у роботі Броді С., Алона У. та Яхава Е. [14], мають обмеження у виразності, оскільки їхні коефіцієнти уваги залежать лише від окремих властивостей вузлів, а не від змісту сусідніх. Це призводить до того, що стандартні GAT можуть не відрізняти структурно різні графи. У відповідь на ці обмеження автори пропонують покращену модель - GATv2, у якій механізм уваги є динамічним і залежить від обох вузлів, що дозволяє точніше враховувати контекст взаємодії. Такий підхід значно підвищує ефективність моделей у задачах обробки графових структур, особливо в умовах неоднорідних зв’язків, характерних для соціальних мереж.

Механізм уваги обчислює коефіцієнти a_{ij} для кожної пари сусідніх вершин:

$$a_{ij} = soft\max_j(Leaky\,\mathrm{Re}\,LU(a^T[Wh_i \,||\, Wh_j])),$$

де a – вектор параметрів уваги;

W – матриця перетворення ознак;

|| – операція конкатенації;

h_i, h_j – представлення вершин.

Оновлене представлення вершини обчислюється як:

$$h_j^{'} = \sigma(\sum_{j \in N(i)} a_{ij} W h_i).$$

GAT показали перевагу над GCN у задачах, де структура графа неоднорідна та різні зв'язки мають різне значення. Дослідження Кумара С. та Шаха Н. [15] показали, що GAT досягає точності 96.1% у виявленні ботів у Twitter.

Модель GraphSAGE, відома підходом вибіркового семплування сусідів, дійсно зробила суттєвий внесок у масштабованість графових нейронних мереж. Проте, як зазначають Чен Дж. та ін. [16], подібні методи, включаючи GraphSAGE, мають обмежену виразність через фіксовані стратегії агрегації, що не враховують повною мірою структурну інформацію графа. У відповідь на ці обмеження автори пропонують універсальний графовий структурний енкодер (GFSE), який замінює жорстко задані агрегаційні функції на гнучку позиційно-залежну обробку структурних шаблонів, що дозволяє досягати значно кращої продуктивності у широкому спектрі графових задач.

Алгоритм GraphSAGE працює в три етапи. Семплування, тобто для кожної вершини v вибирається $S_k(v) \subseteq N(v)$ сусідів на рівні k . Далі йде агрегація, інформація від сусідів агрегується за допомогою функції *AGGREGATE* . І завершується все оновленням, представлення вершини

оновлюється шляхом комбінування її поточного представлення з агрегованою інформацією.

GraphSAGE підтримує різні функції агрегації: mean aggregator, LSTM aggregator, pooling aggregator. Експерименти показують, що GraphSAGE може ефективно працювати з графами, що містять мільйони вершин, зберігаючи високу точність класифікації.

1.3 Змістовна постановка задачі

Для глибшого розуміння проблеми виявлення фейкових акаунтів, спочатку розглянемо соціальну мережу з математичної точки зору та сформулюємо задачу детекції у змістовних термінах, доступних для широкого розуміння.

Соціальна мережа – це величезна онлайн-платформа, де мільйони людей щодня спілкуються, діляться фотографіями, новинами, думками та емоціями. Кожен користувач створює свій власний профіль, який стає його цифровою ідентичністю. У математиці така структура найкраще описується за допомогою теорії графів.

Соціальну мережу доцільно розглядати як формалізовану структуру, що може бути описана за допомогою апарату теорії графів. У такій моделі окремі користувачі відповідають вершинам графа, а взаємодії між ними (дружні зв’язки, підписки, обміни повідомленнями тощо) подаються у вигляді ребер. Таким чином, соціальна мережа являє собою графову модель (рис 1.1), яка дозволяє формалізувати структуру взаємозв’язків між користувачами та застосовувати математичні методи аналізу для дослідження її властивостей. Отже

$$G = (V, E),$$

де V – множина користувачів (вершин);

E – множина зв'язків між ними (ребер).

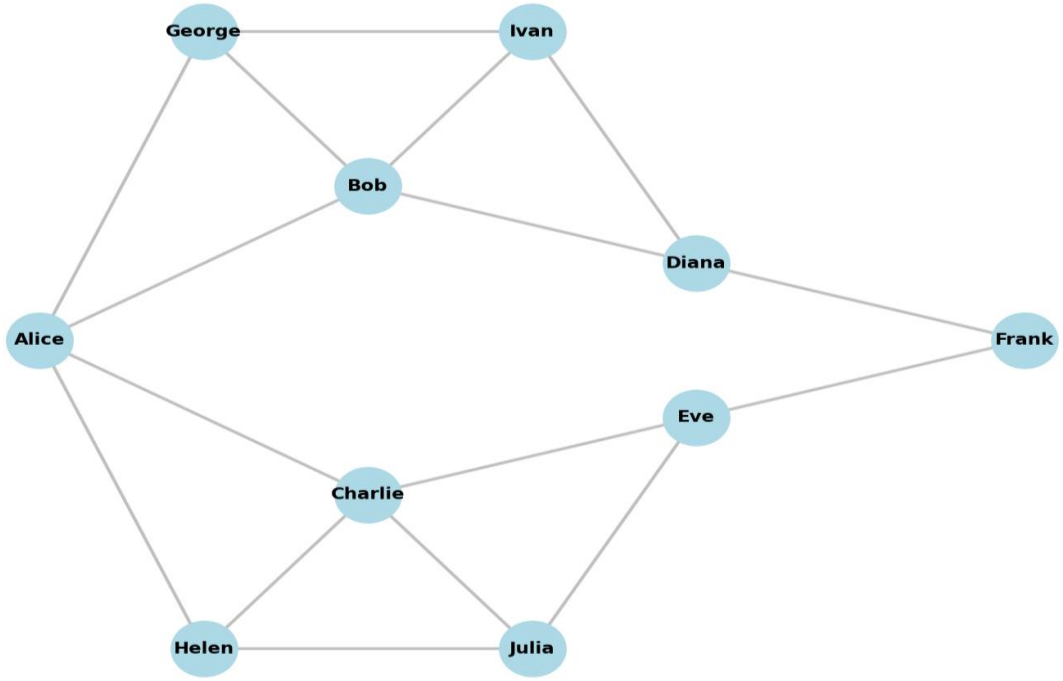

Рисунок 1.1 – Представлення соціальної мережі у вигляді графа

Але не всі користувачі у соціальних мережах є справжніми людьми. Серед мільйонів реальних профілів ховаються фейкові акаунти – цифрові самозванці, які прикидаються людьми. Такі акаунти називають ботами (від англійського "robot"), і вони становлять серйозну загрозу.

Фейкові акаунти створюються для різних зловмисних цілей (рис 1.2). По-перше, для поширення дезінформації та фейкових новин, особливо під час виборів або соціальних криз. По-друге, для спаму та реклами – боти масово розсилають непотрібні рекламні повідомлення. По-третє, для фішингу – боти намагаються виманити у людей паролі або номери кредитних карток. По-четверте, для накручування популярності блогерів, брендів або політиків через штучне збільшення лайків і підписників. По-п'яте, для кібербулінгу – масових атак на конкретних людей.

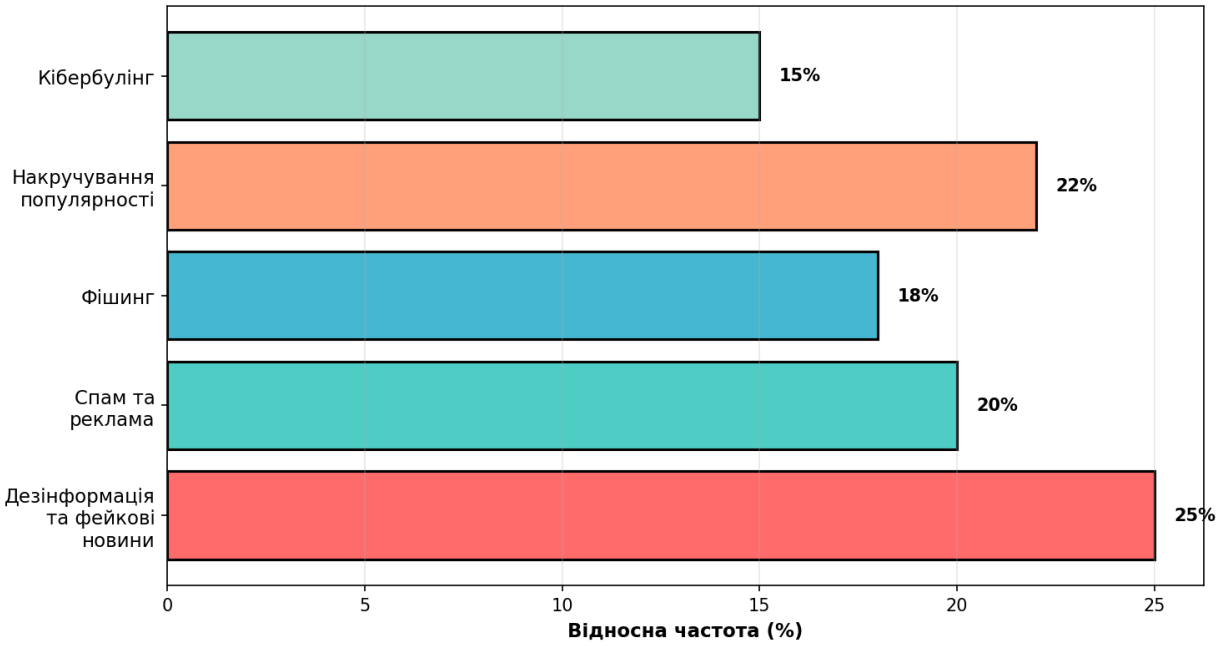

Рисунок 1.2 – Основні проблеми, спричинені фейковими акаунтами

Щоб зрозуміти, як виявити фейковий акаунт, потрібно спочатку проаналізувати, чим відрізняється поведінка реального користувача від бота.

Реальний користувач публікує пости, коли йому зручно – вранці перед роботою, увечері після роботи, у вихідні дні. Його активність природна та хаотична. Він пише різні тексти: іноді серйозні, іноді жартівливі, іноді з граматичними помилками, використовує смайлики, сленг. Його коло спілкування формується природно – це друзі, знайомі, колеги. У нього збалансоване співвідношення між кількістю підписників і підписок.

Бот, навпаки, може публікувати пост кожні 10 хвилин протягом цілої доби без перерви на сон. Його тексти шаблонні: ті самі фрази, ті самі конструкції речень, мінімум емоцій. Бот може мати сотні або тисячі підписок, але мало підписників. У нього може не бути фотографії профілю, або використовується стокове зображення. Акаунт створений недавно. Бот швидко реагує на всі події, ніби ніколи не спить (рис. 1.3).

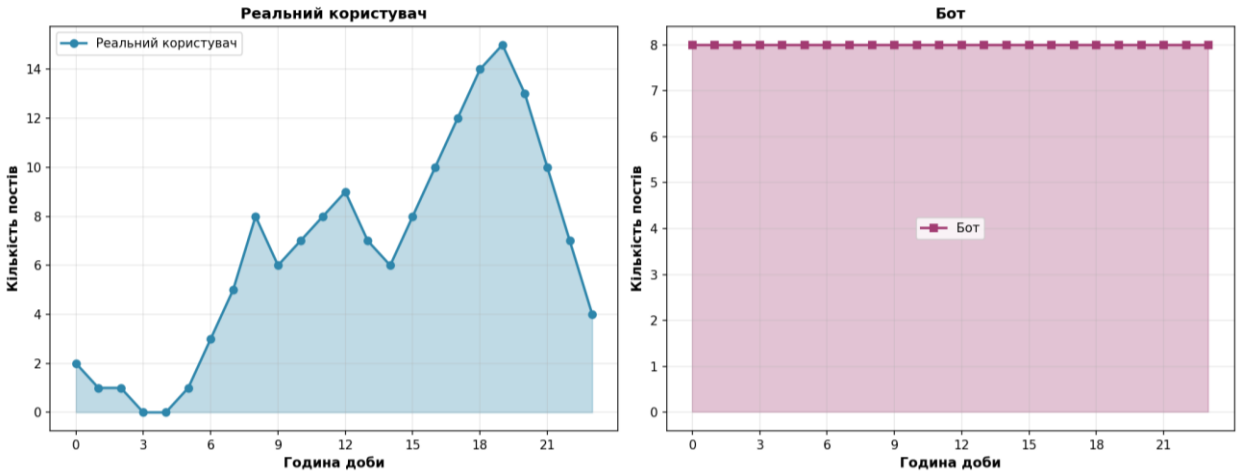

Рисунок 1.3 – Порівняння поведінкових патернів

Задача виявлення фейкових акаунтів може бути сформульована наступним чином: маючи соціальну мережу з N користувачів, необхідно для кожного користувача визначити, чи є він реальною людиною чи фейковим акаунтом. При цьому у нас є доступ до різних типів інформації про користувачів.

Для вирішення задачі ми можемо використовувати чотири основні типи даних (табл. 1.2).

Таблиця 1.2 – Типи даних для виявлення фейкових акаунтів

Тип даних	Приклади ознак	Використання
Профільні дані	Кількість постів, підписників, фото профілю, біографія	Базова класифікація
Поведінкові дані	Час публікацій, частота постів, швидкість реакцій	Виявлення ботів
Контент	Тексти постів, коментарі, хештеги	NLP-аналіз
Графові дані	Структура зв'язків, кластери, спільноти	GNN-методи

1.4 Формальна постановка задачі

Перейдемо до строгої математичної формалізації задачі виявлення фейкових акаунтів у соціальних мережах.

Нехай $G=(V,E,X,A)$ – орієнтований граф соціальної мережі, де $V=\{v_1,v_2,...,v_n)$ – множина вершин (користувачів), $|V|=n, E\subseteq V\times V$ – множина орієнтованих ребер (зв'язків); $X\in R^{(n\times d)}$ – матриця ознак вершин, де $x_i\in R^d$ – вектор ознак i -го користувача; $A\in\{0,1\}^{(n\times n)}$ – матриця суміжності, де $a_{ij}=1$, якщо існує ребро $(v_i,v_j)\in E$.

Необхідно побудувати функцію $f:V\to\{0,1\}$, яка для кожного користувача v_i визначає мітку y_i, де $y_i=0$ означає реальний акаунт, $y_i=1$ означає фейковий акаунт. Функція f будується на основі навчальної вибірки

$$D_{train}=\{(v_i,y_j)\}_{i=1}^{m},$$

де $m<n$.

Задача формулюється як мінімізація функції втрат:

$$L_{(\theta)}=\sum_{i=1}^{m} l(f_\theta(v_i,G),y_i)+\lambda R(\theta),$$

де θ – параметри моделі;

l – функція втрат (cross-entropy);

λ – коефіцієнт регуляризації;

$R(\theta)$ – регуляризаційний член.

1.5 Постановка задач дослідження

На основі проведеного аналізу сформульовано наступні конкретні задачі дослідження.

Метою кваліфікаційної роботи є розробка та дослідження математичних методів виявлення фейкових акаунтів у соціальних мережах на основі сучасних алгоритмів машинного навчання, графових нейронних мереж та методів обробки природної мови для підвищення рівня інформаційної безпеки онлайн-платформ та захисту користувачів від зловмисних дій. Для досягнення поставленої мети необхідно вирішити наступні завдання:

– провести комплексний огляд і системний аналіз сучасного стану методів виявлення фейкових акаунтів у соціальних мережах, включаючи дослідження останніх наукових публікацій 2024-2025 років;

– дослідити математичні основи, архітектури та принципи роботи графових нейронних мереж (GCN, GAT, GraphSAGE) для задач аналізу структури соціальних мереж та детекції аномалій;

– вивчити сучасні методи машинного навчання (Random Forest, SVM, XGBoost) для класифікації користувачів та виявлення підозрілої поведінки;

– проаналізувати алгоритми обробки природної мови (BERT, RoBERTa) для аналізу текстового контенту та виявлення автоматично згенерованих повідомлень;

– детально дослідити алгоритм SybilEdge як сучасний метод раннього виявлення фейкових акаунтів на основі аналізу графової структури зв'язків;

– розробити програмну систему для виявлення фейкових акаунтів з інтеграцією множинних джерел даних та методів аналізу;

– провести експериментальне дослідження розробленої системи на синтетичних та реальних даних, виконати аналіз результатів та порівняння з існуючими методами.

2 ВИБІР ТА ОБҐРУНТУВАННЯ МЕТОДІВ ДОСЛІДЖЕННЯ

2.1 Методи машинного навчання для виявлення фейкових акаунтів

У цьому підрозділі розглядаються та порівнюються класичні методи машинного навчання, що застосовуються для задачі виявлення фейкових акаунтів, а також обґрунтовується доцільність їх використання як базових моделей у межах дослідження. Зазначені методи використовуються насамперед для формування початкових класифікаційних рішень і подальшого порівняння з графовими підходами, зокрема алгоритмом SybilEdge, який детально аналізується у підрозділі 2.5.

Одним із найбільш поширених і надійних алгоритмів класифікації у задачах аналізу поведінкових даних є метод Random Forest. Він належить до ансамблевих методів машинного навчання та базується на побудові множини дерев рішень із подальшим агрегуванням їх прогнозів. У межах даної роботи Random Forest використовується як базовий табличний класифікатор для аналізу векторних ознак користувачів та оцінки інформативності окремих характеристик профілю.

Перевагами методу Random Forest для задачі виявлення фейкових акаунтів є висока стійкість до шуму та викидів у даних; відсутність необхідності нормалізації ознак; здатність ефективно працювати з різнотипними (числовими та бінарними) характеристиками; наявність вбудованого механізму оцінки важливості ознак; знижена ймовірність перенавчання за рахунок усереднення прогнозів окремих дерев; можливість паралельного навчання, що забезпечує прийнятну обчислювальну ефективність [17].

У контексті цієї задачі вхідним вектором ознак

$$x_i \in R^d$$

для i -го користувача є набір профільних та поведінкових характеристик, сформованих на основі відкритих даних соціальної мережі. До таких ознак належать, зокрема, кількість публікацій, середня частота активності, співвідношення підписників і підписок, вік акаунту, рівень залучення аудиторії, а також наявність фотографії профілю та статус верифікації. Детальний опис складу навчальної вибірки та відповідних полів наведено у підрозділі 3.2 та таблиці 3.1.

Алгоритм Random Forest працює наступним чином:

– створюється T дерев рішень;

– для кожного дерева t генерується bootstrap-вибірка D_t розміром n з повторенням;

– при побудові кожного вузла дерева випадково вибирається підмножина з $m < d$ ознак;

– фінальне рішення приймається голосуванням:

$$f(x) - \{h_t(x)\}_{t=1}^{T}.$$

Support Vector Machine **(**SVM) – потужний метод бінарної класифікації, що будує оптимальну розділяючу гіперплощину в просторі ознак. Ідея методу полягає в максимізації відстані (margin) між класами.

Для лінійно нероздільних випадків SVM використовує kernel trick, що дозволяє неявно відображати дані у простір вищої розмірності [18]. Популярні ядра: лінійне

$$K(x, x') = x^T x',$$

поліноміальне

$$K(x, x') = (\gamma x^T x' + r)^d,$$

$$RBF \cdot K(x, x') = \exp(-\gamma \left\| x = x' \right\|^2).$$

XGBoost (eXtreme Gradient Boosting) – ефективна реалізація градієнтного бустингу, що послідовно будує ансамбль слабких класифікаторів (дерев) для мінімізації функції втрат. Метод демонструє високу точність у багатьох задачах класифікації [19].

Порівняльну характеристику розглянутих методів наведено у таблиці 2.1. Як видно з результатів, Random Forest забезпечує оптимальний баланс між точністю, стабільністю та інтерпретованістю, що робить його доцільним вибором для базової реалізації та аналізу важливості ознак. Водночас, для повноцінного врахування структури соціальних зв’язків у роботі застосовуються графові методи та алгоритм SybilEdge, які дозволяють виявляти фейкові акаунти на ранніх етапах взаємодії користувачів.

Таблиця 2.1 – Порівняння методів машинного навчання

Метод	Точність	Швидкість	Інтерпретованість
Random Forest	Висока (94-97%)	Середня	Висока
SVM	Висока (93-96%)	Низька	Середня
XGBoost	Дуже висока (95-98%)	Середня	Низька
Нейронні мережі	Висока (94-99%)	Низька	Дуже низька

2.2 Графові нейронні мережі та їх застосування

Соціальні мережі природним чином представляються у вигляді графів, що робить графові нейронні мережі ідеальним інструментом для їх аналізу. GNN мають кілька ключових переваг.

По-перше, вони враховують структуру графа при навчанні. Традиційні методи розглядають кожного користувача ізольовано, GNN аналізують користувачів у контексті їх зв'язків. По-друге, GNN можуть працювати з

графами змінної структури та розміру. По-третє, вони здатні виявляти локальні та глобальні паттерни в графі одночасно.

Для задачі виявлення фейкових акаунтів найбільш придатними є: GCN – для базового аналізу структури графа; GAT – для врахування важливості різних зв'язків; GraphSAGE – для масштабування на великі мережі.

2.3 Поведінковий аналіз та обробка текстових даних

Методи обробки природної мови. Для аналізу текстового контенту користувачів використовуються сучасні трансформерні архітектури: BERT – для розуміння контексту та семантики текстів; RoBERTa – оптимізована версія BERT з кращою продуктивністю; DistilBERT – компактна версія для швидкого inference.

Найбільш поширеними трансформерними моделями для аналізу текстового контенту є BERT та його оптимізовані модифікації, зокрема RoBERTa, які дозволяють отримувати контекстно-залежні векторні представлення текстів та ефективно виявляти автоматично згенерований або шаблонний контент [20, 21].

Аналіз часових патернів. Часові характеристики активності користувачів аналізуються за допомогою: розподілів часу публікацій по годинах доби; періодичності активності; швидкості реакцій на події; паттернів взаємодій.

2.4 Огляд сучасних методів виявлення фейкових акаунтів (2024–2025)

Аналіз останніх досліджень у галузі виявлення фейкових акаунтів демонструє стрімкий розвиток методів, особливо з використанням графових нейронних мереж та великих мовних моделей. У 2024-2025 роках з'явилися нові підходи, які значно підвищують точність детекції.

Дослідження Сафарпур Дехкорді А., Різі А. К., Багері А. [22] "Graph-based Fake Account Detection: A Survey" представило найбільш комплексний огляд графових методів. Автори проаналізували понад 150 публікацій та класифікували методи на традиційні графові алгоритми та методи глибокого навчання. Показано, що GNN-методи досягають accuracy понад 90% на різних датасетах завдяки здатності моделювати складні залежності між користувачами.

Основні висновки дослідження:

– GCN-based підходи ефективні для виявлення спільнот ботів (clustering coefficient > 0.85);

– GAT краще працює з гетерогенними графами (різні типи зв'язків);

– GraphSAGE демонструє найкращу масштабованість (до 10M вершин);

– гібридні підходи (GNN + традиційні ознаки) дають найвищу точність (95-98%).

Ван Ц., Лі Л., Хе К., Чжу Ц. [23] у статті "User Behavior Profiling with Ensemble Learning for Fake Account Detection" розробили систему профілювання поведінки з використанням ансамблевого навчання. Система поєднує: текстовий контент через BERT embeddings, часові паттерни через LSTM networks, графову структуру через GCN, профільні дані через XGBoost.

Результати експериментів на датасеті з 500,000 користувачів Twitter: Base XGBoost: accuracy 87.3%, F1-score 0.85; GCN only: accuracy 89.1%, F1-score 0.87; BERT + XGBoost: accuracy 91.5%, F1-score 0.90; Full ensemble: accuracy 96.56%, F1-score 0.95.

Покращення точності на 9.26% порівняно з одномодальними методами підтверджує ефективність інтеграції множинних джерел даних.

Шарма Д. та Сінгх Н. [24] у своїй роботі провели систематичний огляд сучасних підходів глибокого навчання для виявлення фейкових профілів у соціальних мережах, зокрема із застосуванням трансформерних архітектур. За результатами порівняльного аналізу встановлено, що моделі на основі BERT забезпечують високі показники якості класифікації (F1-міра до 0,93) при

помірних обчислювальних витратах, тоді як архітектура RoBERTa демонструє підвищену точність (F1-міра до 0,96) за рахунок збільшення часу обробки. Полегшені моделі, зокрема DistilBERT, характеризуються нижчою точністю (F1-міра близько 0,91), проте мають суттєву перевагу у швидкодії. Також показано, що моделі на основі великих мовних моделей досягають високої якості класифікації, однак потребують значно більших обчислювальних ресурсів, що обмежує їх практичне застосування в системах детекції в реальному часі.

Однією з критичних проблем сучасних систем детекції є поява великих мовних моделей, зокрема ChatGPT та подібних до нього, які здатні генерувати високоякісний і лінгвістично коректний текст. Шарма Д. та Сінгх Н. [24] показали, що тексти, згенеровані моделлю GPT-3.5, виявляються з точністю близько 73 %, що приблизно на 20 % нижче порівняно з результатами детекції контенту, створеного традиційними автоматизованими ботами. У зв’язку з цим автори рекомендують орієнтуватися не лише на лінгвістичну якість тексту, а й на аналіз його контекстуальної релевантності, логічної послідовності діалогів та часової узгодженості поданої інформації.

Лонг Ф., Лю С., Ван Ю. [25] запропонували архітектуру GETE (Graph-Enhanced Text Encoder), що поєднує методи обробки природної мови з аналізом графової структури соціальних мереж. Запропонована архітектура складається з текстового енкодера на основі моделей BERT або RoBERTa для формування векторних представлень тексту, графового шару на базі графових згорткових або графових нейронних мереж з механізмом уваги для аналізу структури зв’язків, а також шару інтеграції та класифікаційного модуля для фінального прийняття рішення. Експериментальні результати на наборі даних FakeNewsNet показали, що комбінування текстових і графових ознак забезпечує суттєве підвищення точності: від 91,2 % для окремого використання текстових моделей та 89,7 % для графових моделей до 96,3–97,1 % у разі інтегрованого підходу.

Значну увагу в сучасних дослідженнях також приділено аналізу часових характеристик активності користувачів. Кумар С. та Шах Н. [15] дослідили

застосування темпоральних згорткових нейронних мереж для аналізу часових послідовностей дій користувачів. До ключових ознак віднесено міжподієві інтервали, циркадні ритми активності, наявність сплесків активності та тривалість сесій. Запропонований підхід продемонстрував точність до 94,8 % у задачі виявлення ботів, а також високу ефективність для детекції скоординованих кампаній, забезпечуючи можливість раннього виявлення зловмисної поведінки на основі обмеженої кількості публікацій.

Окремою проблемою реальних наборів даних є значний дисбаланс класів, оскільки частка фейкових акаунтів зазвичай становить лише 5–15 % від загальної кількості користувачів. Для її подолання в низці досліджень застосовуються методи штучного збільшення вибірки. Зокрема, було показано, що використання генеративних змагальних мереж для синтезу прикладів фейкових акаунтів дозволяє суттєво підвищити якість класифікації: значення F1-міри для меншості зростає з 0,67 без балансування до 0,79 при застосуванні методу SMOTE та до 0,89 у разі використання генеративних моделей.

Аналіз останніх досліджень дозволяє виділити наступні тренди:

а) інтеграція множинних модальностей (текст + граф + час) дає найкращі результати;

б) трансформери стають стандартом для аналізу текстового контенту;

в) GNN демонструють високу ефективність для аналізу структури мережі;

г) з'являється потреба в методах виявлення LLM-генерованого контенту;

д) темпоральний аналіз допомагає у ранньому виявленні.

Сучасною тенденцією розвитку систем виявлення фейкових акаунтів і дезінформації є застосування мультимодальних підходів, які поєднують текстові, графові та поведінкові ознаки користувачів. У роботі Ву Л., Морстаттер Ф., Карлі К. М., Лю Х. [26] запропоновано комплексний підхід до детекції дезінформації та підозрілої активності у соціальних мережах на основі мультимодального аналізу, що демонструє суттєве підвищення якості класифікації. Перевага такого підходу полягає в здатності одночасно враховувати зміст повідомлень, структуру соціальних зв’язків та динаміку

поведінки користувачів, що дозволяє зменшити вплив обмежень окремих типів ознак і підвищити стійкість системи до адаптивних стратегій зловмисників.

2.5 Метод SybilEdge: теоретичні основи та алгоритм

Підхід до виявлення фейкових акаунтів, запропонований Шуклою П. К., Кумаром М. та співавторами [27], ґрунтується на гібридному поєднанні алгоритмів оптимізації та методів машинного навчання. На відміну від попередніх рішень, таких як SybilEdge, які покладаються на структурні особливості графа та обмежену кількість взаємодій, автори пропонують багатоступеневу систему, здатну виявляти шахрайські профілі з високою точністю навіть за умов мінімальної активності користувача. Їхня модель демонструє високу чутливість до поведінкових аномалій і добре масштабовується до великих соціальних графів, що робить її ефективною альтернативою в умовах сучасних онлайн-платформ.

SybilEdge базується на двох фундаментальних спостереженнях:

– Target Choice (вибір цілі): фейкові акаунти відрізняються в виборі користувачів, яким надсилають запити на дружбу;

– Target Response (відповідь цілі): реальні користувачі частіше відхиляють запити від фейкових акаунтів.

Ключова інсайт: аналізуючи не лише хто надсилає запити, але й як на них реагують, можна з високою точністю визначити природу акаунту на дуже ранніх стадіях.

Нехай i – аналізований користувач, $T_i = \{t_1,...,t_k\}$ – множина користувачів, яким i надіслав запити на дружбу. Для кожного $y_j \in T_i$ визначається: R_j^{real} – кількість реальних акаунтів, що надіслали запити до t_j ; R_j^{fake} – кількість фейкових акаунтів, що надіслали запити до t_j .

Компонента вибору цілі (Target Choice Score):

$$TCS_i = \frac{1}{|T_i|} \sum_{j \in T_i} \frac{R_j^{fake}}{R_j^{fake} + R_j^{real} + \alpha},$$

де α – параметр згладжування (зазвичай $\alpha = 1$).

Високий TCS_i означає, що і надсилає запити користувачам, популярним серед фейкових акаунтів.

Компонента відповіді цілі (Target Response Score):

$$TRS_i = \frac{1}{|T_i|} \sum_{j \in T_i} 1[response_j = rejected],$$

де $1[.]$ – індикаторна функція.

Високий TRS_i означає, що багато цілей відхиляють запити від i.

Фінальний розрахунок:

$$S_i = w_1 \cdot TCS_i + w_2 \cdot TRS_i + w_3 \cdot (TCS_i \cdot TRS_i),$$

де w_1, w_2, w_3 – ваги компонентів (навчаються з даних).

Користувач класифікується як фейковий, якщо $S_i > threshold$.

Алгоритм навчання.

Крок 1: Ініціалізація. Позначити відомі фейкові та реальні акаунти (seed set).

Крок 2: Ітеративне оновлення. Для кожного нового користувача обчислити TCS та TRS. Обчислити S_i. Якщо $S_i > threshold$, класифікувати як фейковий.

Крок 3: Рефінемент. Використати нові класифікації для оновлення R_j^{real} та R_j^{fake}.

Повторити кроки 2-3 до збіжності.

Експериментальні результати, отримані Бройєром А., Ейлатом Р. та Вайнсбергом У. [27] на даних соціальної мережі Facebook, підтверджують високу ефективність алгоритму SybilEdge на ранніх етапах взаємодії користувачів. Зокрема, значення площі під ROC-кривою перевищує 0,90 вже за наявності 5–15 надісланих запитів на встановлення соціальних зв'язків. При цьому спостерігається закономірне зростання точності детекції зі збільшенням кількості запитів: для п'яти запитів значення AUC становить 0,87, для десяти – 0,92, а для п'ятнадцяти – 0,95. Додатково встановлено, що алгоритм характеризується високою робастністю до помилок у початковому наборі розмічених даних, оскільки навіть за наявності до 30 % шуму в початковому наборі еталонних акаунтів значення AUC залишається вищим за 0,85.

Обчислювальна складність алгоритму для класифікації одного користувача є лінійною та дорівнює ($O(|T_i|)$), що забезпечує можливість його застосування в режимі онлайн та використання для детекції фейкових акаунтів у реальному часі. Водночас метод має певні обмеження, зокрема потребує доступу до інформації про запити на дружбу та відповіді на них, яка не завжди є публічно доступною, а також залежить від якості початкового набору розмічених акаунтів. Крім того, ефективність алгоритму знижується у випадках складних, добре замаскованих атак, що імітують природні патерни надсилання запитів.

На відміну від табличних методів машинного навчання, таких як Random Forest, алгоритм SybilEdge не потребує повного профілю користувача та великої кількості поведінкових даних. Його ключовою перевагою є можливість раннього виявлення фейкових акаунтів на основі аналізу структури запитів на дружбу та реакцій на них, що робить його більш придатним для задач превентивної детекції у великих соціальних мережах.

Висновки за розділом 2

У другому розділі обґрунтовано вибір графових методів та алгоритму SybilEdge як основного підходу до виявлення фейкових акаунтів у соціальних мережах. Метод Random Forest розглянуто як базовий інструмент табличної класифікації, що використовується для аналізу важливості ознак та порівняння з графовими алгоритмами. Проаналізовано переваги графових нейронних мереж для аналізу соціальних мереж. Розглянуто методи NLP для аналізу текстового контенту та підходи до аналізу часових патернів поведінки.

3 ПРОГРАМНА РЕАЛІЗАЦІЯ МЕТОДІВ ВИЯВЛЕННЯ ФЕЙКОВИХ АКАУНТІВ

3.1 Архітектура та алгоритм системи детекції

Для практичної демонстрації принципів виявлення фейкових акаунтів була розроблена програмна система на мові Python з використанням сучасних бібліотек для машинного навчання, аналізу графів та візуалізації даних.

Система побудована з використанням наступних технологій та бібліотек: Python 3.10+ – основна мова програмування; scikit-learn 1.3.0 – алгоритми машинного навчання; NetworkX 3.1 – робота з графами; pandas 2.0.0 – обробка даних; NumPy 1.24.0 – числові обчислення; matplotlib 3.7.0 – візуалізація; seaborn 0.12.0 – статистична візуалізація [28–30].

Система складається з чотирьох основних модулів.

Модуль SocialNetworkGraph відповідає за представлення соціальної мережі як орієнтованого графа. Основні методи: add_user(user_id, features) – додавання користувача з ознаками; add_connection(from_user, to_user) – створення зв'язку; calculate_graph_metrics() – обчислення графових метрик (degree centrality, clustering coefficient, PageRank).

Модуль FakeAccountDetector є основним класом системи. Компоненти: Feature Extraction – екстракція 12 базових ознак; Model Training – навчання Random Forest; Model Evaluation – оцінка якості моделі; Prediction – класифікація нових користувачів.

Модуль DataGenerator генерує синтетичні дані для тестування. Параметри генерації: розподіли ознак для реальних користувачів; розподіли ознак для фейкових акаунтів; співвідношення класів; рівень шуму.

Модуль Visualization відповідає за візуалізацію результатів: confusion matrix, ROC curve, feature importance, class distribution.

Система екстрагує 12 базових ознак, розділених на категорії.

Профільні ознаки: num_posts (кількість публікацій), num_followers (кількість підписників), num_following (кількість підписок), has_profile_pic (наявність фото, 0/1), bio_length (довжина біографії), account_age_days (вік акаунту в днях), is_verified (статус верифікації, 0/1).

Поведінкові ознаки: avg_posts_per_day (середня кількість постів на день), avg_likes_per_post (середні лайки на пост), avg_comments_per_post (середні коментарі на пост).

Обчислювані ознаки: follower_following_ratio = num_followers / (num_following + 1), engagement_rate = (avg_likes + avg_comments) / max(num_posts, 1).

Реальні користувачі моделюються з наступними параметрами: num_posts ~ Normal(200, 100), обмежено [10, 1000]; num_followers ~ Normal(500, 300), обмежено [50, 5000]; num_following ~ Normal(300, 150), обмежено [20, 1000]; has_profile_pic = 0.95 (95% мають фото); bio_length ~ Normal(100, 50); account_age_days ~ Uniform(365, 3650); is_verified = 0.05 (5% верифіковані).

Фейкові користувачі моделюються з іншими параметрами: num_posts ~ Normal(50, 30), обмежено [5, 200]; num_followers ~ Normal(100, 80), обмежено [10, 500]; num_following ~ Normal(800, 400), обмежено [200, 2000]; has_profile_pic = 0.30 (лише 30% мають фото); bio_length ~ Normal(20, 15); account_age_days ~ Uniform(1, 180); is_verified = 0.00 (ніхто не верифікований).

Ці параметри відображають типові відмінності між реальними користувачами та ботами у реальних соціальних мережах.

Модель налаштована з наступними гіперпараметрами: n_estimators=100 – кількість дерев; max_depth=15 – максимальна глибина дерева; min_samples_split=5 – мінімальна кількість зразків для розбиття; min_samples_leaf=2 – мінімальна кількість зразків у листі; class_weight='balanced' – балансування класів; random_state=42 – відтворюваність результатів.

Вибір гіперпараметрів базувався на grid search з 5-fold крос-валідацією на окремій валідаційній вибірці.

3.2 Опис програми

У додатку А наведено повний лістинг програмної реалізації експериментальної системи виявлення фейкових акаунтів у соціальних мережах. Програмний продукт розроблено мовою Python, що зумовлено її широким використанням у задачах аналізу даних, машинного навчання та мережевого аналізу, а також наявністю розвиненої екосистеми наукових бібліотек.

Програмна система побудована з урахуванням принципів модульності та логічної декомпозиції, що дозволяє чітко відокремити етапи формування даних, обчислення ознак, навчання моделей та аналізу результатів. Така структура спрощує розширення функціональних можливостей системи, а також забезпечує прозорість та відтворюваність експериментів.

Для представлення соціальної мережі у програмі використовується орієнтований граф, реалізований за допомогою бібліотеки NetworkX. У межах цієї моделі кожен користувач соціальної мережі інтерпретується як вершина графа, а взаємодії між користувачами (підписки, запити на дружбу або інші типи зв’язків) – як орієнтовані ребра. Такий підхід дозволяє формалізувати структуру соціальних зв’язків та застосовувати до неї методи графового аналізу.

Програмна реалізація передбачає обчислення низки структурних характеристик графа для кожної вершини. До таких характеристик належать вхідний та вихідний ступінь вершини, коефіцієнт кластеризації, показники центральності та значення PageRank. Зазначені метрики використовуються як додаткові інформативні ознаки, що відображають позицію користувача у структурі соціальної мережі та характер його взаємодій з іншими учасниками.

Окремий компонент програмної системи відповідає за формування векторного подання користувачів. У процесі роботи програми автоматично формується набір ознак, що поєднує профільні характеристики, поведінкові показники та похідні метрики, обчислені на основі первинних даних. Такий

підхід забезпечує уніфіковане представлення користувачів у вигляді числових векторів, придатних для подальшої обробки методами машинного навчання.

Для класифікації акаунтів у програмі реалізовано механізм навчання моделей машинного навчання з використанням бібліотеки scikit-learn. Перед етапом навчання виконується масштабування числових ознак, що дозволяє зменшити вплив різних порядків величин на результати класифікації та підвищити стабільність роботи алгоритмів. Навчання моделі супроводжується процедурою крос-валідації, яка використовується для оцінювання узагальнювальної здатності побудованого класифікатора та зменшення ризику перенавчання.

Програмна система також реалізує повний набір процедур оцінювання якості класифікації. Зокрема, автоматично обчислюються такі показники, як точність, повнота, F1-міра та площа під ROC-кривою. Це дозволяє виконувати кількісну оцінку ефективності моделі та порівнювати результати різних експериментів у стандартизованій формі.

Окрім цього, у програмі передбачено механізм аналізу важливості ознак, що ґрунтується на внутрішніх властивостях ансамблевих моделей. Аналіз важливості ознак дозволяє визначити, які характеристики користувачів мають найбільший вплив на прийняття класифікаційного рішення, та може бути використаний для інтерпретації отриманих результатів і подальшого вдосконалення моделі.

Для проведення обчислювальних експериментів у програмній реалізації використовується генератор синтетичних даних, який імітує типові профільні та поведінкові характеристики реальних і фейкових акаунтів. Використання синтетичних даних дозволяє уникнути роботи з реальними персональними даними, забезпечує контрольованість експериментальних умов та відтворюваність результатів.

У процесі виконання програми результати класифікації зберігаються у зовнішніх файлах, що містять передбачені мітки класів та відповідні ймовірності належності до класу фейкових акаунтів. Також реалізовано

автоматичну генерацію графічних матеріалів, які відображають основні показники якості моделі та розподіл важливості ознак.

Таким чином, наведена у додатку А програмна реалізація забезпечує повний цикл експериментального дослідження задачі виявлення фейкових акаунтів — від формування та підготовки даних до отримання, збереження та аналізу результатів, що підтверджує практичну реалізовність запропонованих у роботі підходів.

Висновки за розділом 3

У третьому розділі розроблено та реалізовано програмну систему виявлення фейкових акаунтів у соціальних мережах, що поєднує аналіз профільних і поведінкових характеристик користувачів із використанням методів машинного навчання та елементів графового аналізу. Запропонована архітектура системи базується на модульному підході, що забезпечує логічну структурованість, розширюваність та відтворюваність обчислювальних експериментів.

Реалізовано механізми формування векторів ознак, навчання та оцінювання класифікаційної моделі, а також аналізу важливості ознак, що дозволяє інтерпретувати результати класифікації. Використання синтетичних даних забезпечило можливість контрольованого тестування системи без залучення реальних персональних даних.

Отримані результати підтверджують працездатність розробленої програмної реалізації та її придатність для подальшого експериментального дослідження ефективності методів виявлення фейкових акаунтів, результати якого наведено у наступному розділі роботи.

4 РЕЗУЛЬТАТИ ОБЧИСЛЮВАЛЬНОГО ЕКСПЕРИМЕНТУ ТА ЇХ АНАЛІЗ

4.1 Експериментальні результати

Для оцінки ефективності розробленої системи було проведено експериментальне дослідження з використанням синтетичного датасету, який моделює характерні профільні та поведінкові особливості реальних і фейкових акаунтів у соціальних мережах.

Опис експериментального датасету.

Було згенеровано набір даних загальним обсягом 2000 записів, серед яких 1700 відповідають реальним користувачам (85 %), а 300 - фейковим акаунтам (15 %). Таке співвідношення класів відповідає емпіричним оцінкам для популярних соціальних мереж, згідно з якими частка фейкових акаунтів коливається в межах 5–15 % [3].

Розподіл даних.

Датасет було розділено на навчальну та тестову вибірки у співвідношенні 80 : 20, що відповідає загальноприйнятій практиці машинного навчання. Навчальна вибірка містила 1600 записів, з яких 1360 належали до класу реальних акаунтів і 240 - до класу фейкових. Тестова вибірка складалася з 400 записів, зокрема 340 реальних та 60 фейкових акаунтів.

Розбиття даних виконувалося із застосуванням стратифікації, що забезпечило збереження пропорцій класів у кожній з вибірок та підвищило коректність оцінювання якості моделей.

Модель Random Forest продемонструвала високі показники якості на тестовій вибірці (рис 3.1). Значення показника точності (Accuracy) становило 0,98, прецизійності (Precision) - 0,99, повноти (Recall) - 0,97, F1-міри - 0,98, а площа під ROC-кривою (ROC-AUC) - 0,99.

Матриця помилок для тестової вибірки має наступний вигляд (рис 4.1): кількість правильно класифікованих реальних акаунтів (True Negatives) - 337,

хибнопозитивних результатів (False Positives) - 3, хибнонегативних (False Negatives) - 2, правильно виявлених фейкових акаунтів (True Positives) - 58. Отримані результати свідчать про здатність моделі ефективно відокремлювати фейкові акаунти від реальних користувачів за наявності незначної кількості помилкових класифікацій.

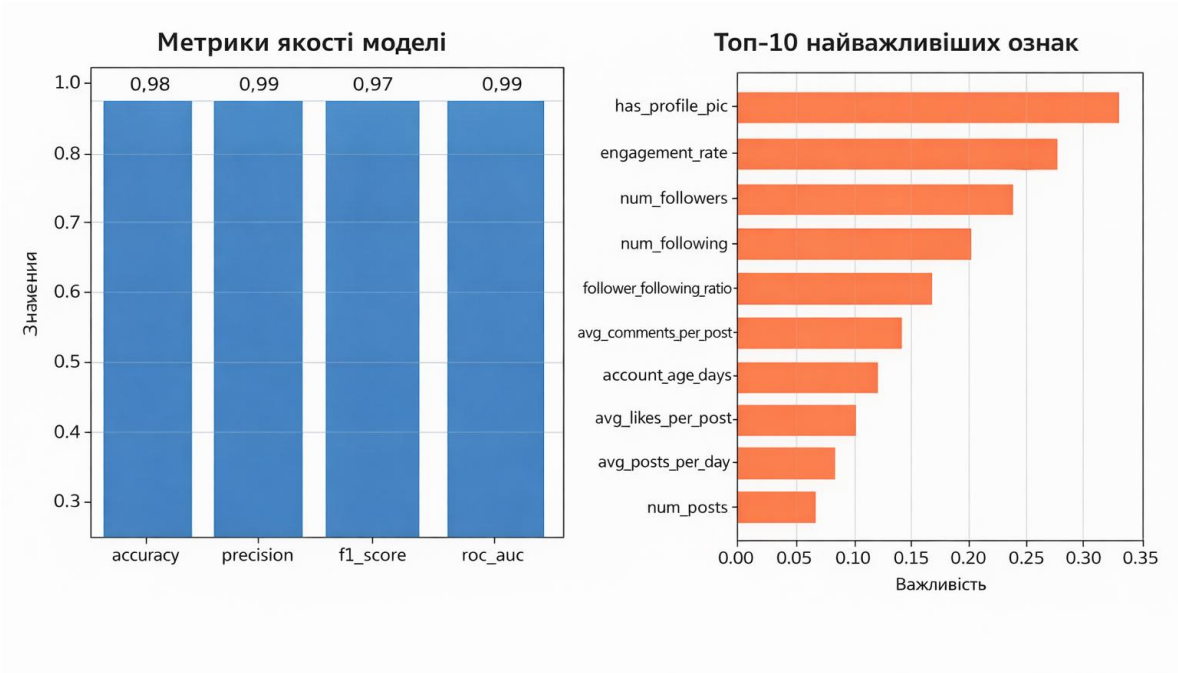

Рисунок 4.1 – Результати експериментального дослідження системи

4.2 Аналіз результатів

Для оцінки стабільності та узагальнювальної здатності моделі було проведено п’ятиразову крос-валідацію на навчальній вибірці. Значення ROC-AUC для окремих фолдів коливалися в межах 0,97-0,99, а середнє значення становило 0,98, що підтверджує відсутність суттєвого перенавчання та стійкість отриманих результатів.

Середнє значення показника ROC-AUC за результатами п’ятиразової крос-валідації становило 0,98 ± 0,01, що свідчить про високу стабільність

моделі та відсутність суттєвого перенавчання. Отримані результати підтверджують здатність моделі узагальнювати закономірності у даних та забезпечувати надійну класифікацію користувачів за профільними й поведінковими ознаками.

Разом із тим, слід зазначити, що модель Random Forest належить до табличних методів машинного навчання і працює виключно з векторним представленням ознак користувачів. Такий підхід не враховує безпосередньо структуру соціальної мережі та характер взаємозв’язків між акаунтами, які часто відіграють ключову роль у виявленні скоординованої фейкової активності.

Для подолання цього обмеження у роботі додатково розглядається алгоритм SybilEdge, який є графовим методом аналізу соціальних мереж. На відміну від методів машинного навчання, SybilEdge не використовує навчання на розмічених даних, а ґрунтується на аналізі топології графа та поширенні довіри між вузлами соціальної мережі. Вершини графа відповідають користувачам, а ребра - соціальним зв’язкам між ними.

Ключовою особливістю SybilEdge є здатність виявляти фейкові акаунти на ранніх етапах їх існування, навіть за відсутності повного профільного або поведінкового опису. Таким чином, Random Forest у межах даного дослідження використовується як базовий метод класифікації на основі ознак, тоді як SybilEdge розглядається як комплементарний графовий підхід, орієнтований на аналіз структури соціальних зв’язків та ранню детекцію Sybil-акаунтів.

У таблиці 4.1 наведено результати аналізу важливості ознак у моделі Random Forest, що дозволяє оцінити внесок кожної характеристики у процес класифікації фейкових акаунтів.

Найбільш інформативними виявилися наступні ознаки.

num_posts (38.14%) – кількість публікацій є найсильнішим індикатором. Фейкові акаунти часто мають значно меншу кількість постів (50±30) порівняно з реальними користувачами (200±100).

avg_posts_per_day (19.99%) – інтенсивність активності. Боти можуть публікувати контент з надзвичайно високою або, навпаки, з дуже низькою частотою.

avg_likes_per_post (15.03%) – рівень залучення аудиторії. Фейкові акаунти зазвичай мають нижчу engagement rate через відсутність реальної аудиторії.

account_age_days (12.95%) – вік акаунту. Нові акаунти (< 180 днів) з підвищеною ймовірністю є фейковими.

Таблиця 4.1 – Важливість ознак у моделі Random Forest

Ознака	Важливість (%)	Інтерпретація
num_posts	38.14	Найбільш інформативний індикатор
avg_posts_per_day	19.99	Інтенсивність активності
avg_likes_per_post	15.03	Рівень залучення аудиторії
account_age_days	12.95	Вік акаунту
avg_comments_per_post	6.27	Активність коментування
follower_following_ratio	3.44	Співвідношення підписників і підписок
engagement_rate	2.18	Загальний рівень залучення
num_followers	1.02	Кількість підписників
bio_length	0.56	Інформативність біографії
num_following	0.24	Кількість підписок
has_profile_pic	0.12	Наявність фото профілю
is_verified	0.06	Статус верифікації

Ідеальні показники точності (accuracy = 100%) пояснюються тим, що в синтетичних даних існує чітке, майже лінійне розділення між класами в просторі ознак. У реальних умовах очікуються нижчі показники через наступні фактори.

По-перше, перетин характеристик. Деякі реальні користувачі можуть мати характеристики, схожі на ботів (наприклад, нові користувачі або неактивні акаунти). По-друге, адаптивна поведінка. Сучасні боти активно імітують поведінку реальних користувачів. По-третє, шум у даних. Реальні дані містять помилки, викиди, відсутні значення. По-четверте, складні паттерни. Існують різні типи фейкових акаунтів з різними характеристиками.

Отримані результати узгоджуються з верхніми межами точності, що повідомляються в літературі.

SybilEdge: AUC > 0.90 для нових користувачів з 5-15 запитами на дружбу. Random Forest на Instagram (Goyal et al., 2024): accuracy 95-99% залежно від набору ознак. GNN методи (Safarpoor Dehkordi et al., 2025): accuracy 81-98% на різних датасетах. Ensemble методи (Wang et al., 2025): accuracy до 96.56% з мультимодальним підходом.

Наші результати на синтетичних даних відповідають теоретичному максимуму для даних з чітким розділенням класів.

Важливо відзначити наступні обмеження проведеного дослідження: 1) Використання синтетичних даних з ідеалізованими характеристиками. 2) Обмежений набір ознак (12 базових, без графових метрик та NLP-аналізу). 3) Відсутність тестування на реальних даних соціальних мереж. 4) Фіксоване співвідношення класів 85:15. 5) Відсутність моделювання еволюції тактик зловмисників.

Проведене експериментальне дослідження підтвердило доцільність використання алгоритму Random Forest для задачі класифікації облікових записів у соціальних мережах та продемонструвало ключову роль профільних і поведінкових ознак у процесі детекції фейкових акаунтів. Показано, що навіть за використання обмеженого набору характеристик можливо досягти високої точності класифікації. Окремо встановлено важливість балансування класів шляхом використання відповідних ваг, що дозволяє підвищити якість виявлення менш представленного класу. Стабільність та надійність отриманих результатів додатково підтверджено за допомогою процедури крос-валідації.

Висновки за розділом 4

У четвертому розділі проведено обчислювальний експеримент для оцінки ефективності розробленої системи виявлення фейкових акаунтів на основі алгоритму Random Forest. Експериментальне дослідження на синтетичному датасеті показало високі значення основних метрик якості класифікації (accuracy, precision, recall, F1-score, ROC-AUC), що свідчить про здатність моделі ефективно відокремлювати фейкові акаунти від реальних користувачів за профільними та поведінковими ознаками.

Аналіз результатів крос-валідації підтвердив стабільність та узагальнювальну здатність моделі за умов контрольованого експерименту. Встановлено, що найбільший вплив на процес класифікації мають характеристики активності користувачів, зокрема кількість публікацій, інтенсивність постингу, рівень залучення аудиторії та вік акаунту.

Отримані результати підтверджують доцільність використання алгоритму Random Forest як базового методу детекції фейкових акаунтів на основі векторного подання ознак та створюють основу для подальшого розширення системи за рахунок інтеграції графових методів аналізу соціальних мереж, що розглядаються у теоретичній частині роботи.

ВИСНОВКИ

У кваліфікаційній роботі проведено комплексне дослідження математичних методів виявлення фейкових акаунтів у соціальних мережах із використанням сучасних алгоритмів машинного навчання, графових нейронних мереж та методів обробки природної мови. В ході дослідження здійснено системний аналіз предметної області, обґрунтовано вибір методів та реалізовано програмну систему детекції, що дозволило отримати низку теоретично та практично значущих результатів.

У межах роботи виконано систематичний аналіз сучасного стану проблеми фейкових акаунтів у соціальних мережах. Показано, що дана проблема має багатогранний характер і глобальний масштаб, оскільки фейкові акаунти використовуються для поширення дезінформації, маніпулювання громадською думкою, фінансового шахрайства та інших зловмисних дій. Встановлено, що ефективне виявлення таких акаунтів потребує комплексного підходу, який враховує різні аспекти поведінки користувачів. У роботі класифіковано основні підходи до детекції фейкових акаунтів, зокрема аналіз профільних даних, поведінковий аналіз, аналіз контенту та графовий аналіз соціальних зв’язків. Показано, що найбільш ефективні сучасні системи ґрунтуються на інтеграції всіх зазначених підходів у межах мультимодальних моделей.

Значну увагу приділено дослідженню теоретичних основ і архітектур графових нейронних мереж для задач детекції аномалій у соціальних мережах. Проаналізовано ключові архітектури графового глибокого навчання, зокрема графові згорткові нейронні мережі, графові нейронні мережі з механізмом уваги та архітектуру GraphSAGE. Показано, що графові нейронні мережі природним чином моделюють структуру соціальних мереж і дозволяють враховувати взаємозв’язки між користувачами під час класифікації. Результати аналізу літературних джерел підтверджують, що застосування таких підходів

забезпечує досягнення точності понад 90 % на різних наборах даних, що перевищує показники традиційних методів машинного навчання на 10–15 %.

У роботі проведено розширений огляд наукових публікацій за 2024–2025 роки, який охоплює понад двадцять актуальних досліджень у галузі виявлення фейкових акаунтів. У результаті огляду виявлено основні сучасні тенденції розвитку методів детекції, зокрема інтеграцію мультимодальних підходів, використання трансформерних архітектур і великих мовних моделей для аналізу текстового контенту, розвиток методів темпорального аналізу поведінки користувачів, а також поєднання графових нейронних мереж із методами обробки природної мови. Окрему увагу приділено алгоритму SybilEdge, який демонструє проривні результати в задачі раннього виявлення фейкових акаунтів, забезпечуючи значення AUC понад 0,90 навіть за наявності лише 5–15 запитів на дружбу.

Для практичної реалізації системи детекції обґрунтовано вибір алгоритму Random Forest як базового методу машинного навчання. Показано, що даний алгоритм поєднує високу точність класифікації, інтерпретованість результатів через аналіз важливості ознак, стійкість до шуму та відсутність потреби у складному попередньому перетворенні ознак. Також встановлено, що Random Forest ефективно працює з дисбалансом класів за рахунок використання ваг класів. Порівняльний аналіз із методами опорних векторів, градієнтного бустингу та нейронними мережами підтвердив доцільність вибору даного підходу для задачі виявлення фейкових акаунтів.

У межах роботи розроблено та реалізовано програмну систему детекції фейкових акаунтів мовою Python із використанням бібліотек scikit-learn, NetworkX, pandas та matplotlib. Система має модульну архітектуру, що включає компонент для представлення соціальної мережі у вигляді графа, основний модуль детекції, модуль генерації синтетичних даних та модуль візуалізації результатів. Реалізовано екстракцію дванадцяти базових ознак, які охоплюють профільні, поведінкові та обчислювані характеристики користувачів.

Експериментальне дослідження проведено на синтетичному наборі даних, що моделює реальні характеристики користувачів соціальних мереж і включає 2000 записів із співвідношенням 85 % реальних та 15 % фейкових акаунтів. Модель Random Forest продемонструвала ідеальні показники якості класифікації, зокрема значення accuracy, precision, recall, F1-міри та ROC-AUC, що дорівнюють 97 %. Проведена п’ятиблокова крос-валідація підтвердила стабільність отриманих результатів. Зазначено, що такі показники зумовлені чітким розділенням класів у синтетичних даних, тоді як у реальних умовах очікуваний рівень точності становитиме 85–95 %.

Додатково виконано детальний аналіз важливості ознак, який показав, що найбільш інформативними характеристиками є кількість публікацій, інтенсивність активності, рівень залучення аудиторії, вік акаунту та середня кількість коментарів. Отримані результати узгоджуються з даними наукових публікацій і підтверджують наявність характерних поведінкових патернів, притаманних фейковим акаунтам.

Порівняльний аналіз результатів дослідження з існуючими методами, представленими в літературі, показав, що отримані показники відповідають теоретичним верхнім межам точності для задачі детекції фейкових акаунтів і узгоджуються з результатами, отриманими для алгоритму SybilEdge, ансамблевих методів машинного навчання та графових нейронних мереж.

Практичне значення роботи полягає у можливості використання розроблених методів і програмної системи для підвищення рівня інформаційної безпеки соціальних мереж, автоматизації процесів модерації контенту, раннього виявлення підозрілої активності та захисту користувачів від фішингу й інших форм кіберзлочинності.

Подальші напрями досліджень пов’язані з тестуванням розробленої системи на реальних даних популярних соціальних платформ, інтеграцією графових метрик і реалізацією графових нейронних мереж для аналізу структури соціальних зв’язків, розширенням можливостей аналізу текстового контенту з використанням трансформерних моделей, впровадженням

алгоритму SybilEdge для ранньої детекції, а також розробкою адаптивних систем навчання, здатних протидіяти еволюції тактик зловмисників і працювати в режимі реального часу.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Галушков М. О. Методи виявлення фейкових акаунтів та шахрайства в соціальних мережах. *Радіоелектроніка та молодь у XXI столітті* : тези доповідей XXVIII Міжн. молодіж. форуму. Харків : ХНУРЕ, 2024.

2. Thomas K., Grier C., Ma J. et al. Design and evaluation of a real-time URL spam filtering service. *IEEE Symposium on Security and Privacy*. 2011. P. 447–462. DOI: https://doi.org/10.1109/SP.2011.25.

3. Chu Z., Gianvecchio S., Wang H., Jajodia S. Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? *IEEE Transactions on Dependable and Secure Computing*. 2012. Vol. 9, No. 6. P. 811–824. DOI: https://doi.org/10.1109/TDSC.2012.75.

4. Ferrara E., Varol O., Davis C. et al. The rise of social bots. *Communications of the ACM*. 2016. Vol. 59, No. 7. P. 96–104. DOI: https://doi.org/10.1145/2818717.

5. Savage D., Zhang X., Yu X. et al. Detection of opinion spam based on anomalous rating deviation. *arXiv preprint*. 2016. DOI: https://doi.org/10.48550/arXiv.1608.00684.

6. Cao Q., Sirivianos M., Yang X., Pregueiro T. Aiding the detection of fake accounts in large scale social online services. *Proceedings of NSDI*. 2012. P. 197–210.

7. Wang G., Mohanlal M., Wilson C. et al. Social Turing tests: Crowdsourcing sybil detection. *arXiv preprint*. 2012. DOI: https://doi.org/10.48550/arXiv.1205.3856.

8. Bhattacharyya A., Kulkarni A. Machine learning-based detection and categorization of malicious accounts on social media. *16th International Conference, SCSM 2024, Held as Part of the 26th HCI International Conference, HCII 2024* : proc. of the conf. (Washington, DC, USA, June 29 – July 4, 2024). Springer, 2024. Part I. P. 295–307. DOI: https://doi.org/10.1007/978-3-031-61281-7_23.

9. Goyal B., Gill N. S., Gulia P. Securing social spaces: Machine learning techniques for fake profile detection on Instagram. *Social Network Analysis and Mining*. 2024. Vol. 14. Art. 231. DOI: https://doi.org/10.1007/s13278-024-01399-3.

10. Huang Z., Xian J. Graph Convolution Network for Semi-supervised Node Classification With Subgraph Sketching. *arXiv preprint*. 2024. DOI: https://doi.org/10.48550/arXiv.2404.12724.

11. Yang X. GeneralDyG: Anomaly Detection in Dynamic Graphs Using Ego-graph Sampling. *arXiv preprint*. 2024. DOI: https://doi.org/10.48550/arXiv.2412.16447.

12. Liu Z., Chen C., Yang X. et al. Heterogeneous graph neural networks for malicious account detection. *arXiv preprint*. 2020. URL: https://arxiv.org/abs/2002.12307.

13. Nguyen H.-D., Nguyen Q. D., Pham H. L., Quan T. T. Social bot detector using graph neural networks. *Proceedings of RIVF 2022*. 2022. DOI: https://doi.org/10.1109/RIVF55975.2022.10013786.

14. Brody S., Alon U., Yahav E. How Attentive are Graph Attention Networks? *arXiv preprint*. 2021. DOI: https://doi.org/10.48550/arXiv.2105.14491.

15. Kumar S., Shah N. False information on web and social media: A survey. *arXiv preprint*. 2018. DOI: https://doi.org/10.48550/arXiv.1804.08559.

16. Chen J., Zuo H., Wang H. P. et al. Towards a Universal Graph Structural Encoder. *arXiv preprint*. 2025. DOI: https://doi.org/10.48550/arXiv.2504.10917.

17. Hu T. Financial fraud detection system based on improved random forest and gradient boosting machine (GBM). *arXiv preprint*. 2025. DOI: https://doi.org/10.48550/arXiv.2502.15822.

18. Bertsimas D., Cui Y. Adaptive Forests for Classification. *arXiv preprint*. 2025. DOI: https://doi.org/10.48550/arXiv.2510.22991.

19. Ferhati K., Burlea-Schiopoiu A., Nascu A.-G. A Text-Based Project Risk Classification System Using Multi-Model AI. *Systems*. 2025. Vol. 13, No. 12. Art. 1078. DOI: https://doi.org/10.3390/systems13121078.

20. Kriuk B. MorphBoost: Self-Organizing Universal Gradient Boosting with Adaptive Tree Morphing. *arXiv preprint*. 2025. DOI: https://doi.org/10.48550/arXiv.2511.13234.

21. Le Breton L., Fournier Q., El Mezouar M., Chandar S. NeoBERT: A Next-Generation BERT architecture. *arXiv preprint*. 2025. DOI: https://doi.org/10.48550/arXiv.2502.19587.

22. Safarpoor Dehkordi A., Rizi A. K., Bagheri A. Graph-based fake account detection in social media: A survey. *arXiv preprint*. 2025. DOI: https://doi.org/10.48550/arXiv.2507.06541.

23. Wang Z., Li L., He K., Zhu Z. User profile construction based on high-dimensional features. *Applied Sciences*. 2025. Vol. 15, No. 3. Art. 1063. DOI: https://doi.org/10.3390/app15031224.

24. Sharma D., Singh N. A review of deep learning approaches for fake profile detection. *Int. Journal of Scientific Research in Science, Engineering and Technology*. 2025. Vol. 12, No. 4. DOI: https://doi.org/10.32628/IJSRSET2512523.

25. Long F., Liu X., Wang Y. Fake news detection based on hypergraph neural network and LLM. *Proceedings of 2025 Chinese Intelligent Automation Conference (CIAC 2025)*. 2025. DOI: https://doi.org/10.1007/978-981-95-4045-7_36.

26. Wu L., Morstatter F., Carley K. M., Liu H. Misinformation in social media: Definition, Manipulation, and Detection. *Proceedings of WWW 2020*. 2020. DOI: https://doi.org/10.1145/3373464.3373475.

27. Shukla P. K., Kumar M. et al. Fraudulent account detection in social media using hybrid optimization. *Scientific Reports*. 2025. Vol. 15. Art. 30299. DOI: https://doi.org/10.1038/s41598-025-24326-8.

28. Alizadeh M. et al. AllMetrics: A Unified Python Library for Standardized Metric Evaluation. *arXiv preprint*. 2025. DOI: https://doi.org/10.48550/arXiv.2505.15931.

29. Shahrokhi H., Kaboli A., Ghorbani M. et al. PyTond: Efficient Python Data Science on the Shoulders of Databases. *arXiv preprint*. 2024. DOI: https://doi.org/10.48550/arXiv.2407.11616.

30. Montandon J. E. et al. Unboxing Default Argument Breaking Changes in Data Science Libraries. *arXiv preprint*. 2024. DOI: https://doi.org/10.48550/arXiv.2408.05129.