МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ

НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ КІБЕРБЕЗПЕКИ ТА ЗАХИСТУ ІНФОРМАЦІЇ
КАФЕДРА УПРАВЛІННЯ КІБЕРБЕЗПЕКОЮ ТА ЗАХИСТОМ ІНФОРМАЦІЇ

КВАЛІФІКАЦІЙНА РОБОТА

на тему: “МЕТОДИКА АНАЛІЗУ ФЕЙКОВИХ АКАУНТІВ ТА ЇХНЬОГО ВПЛИВУ НА ПОШИРЕННЯ ДЕЗІНФОРМАЦІЇ”

на здобуття освітнього ступеня бакалавра
зі спеціальності 125 Кібербезпека
освітньої програми Управління інформаційною та кібернетичною безпекою

Кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

_____ Яків КОСТЕНКО
(підпис) *Ім’я, ПРІЗВИЩЕ здобувача*

Виконав: здобувач вищої освіти гр. УБД-42
Яків КОСТЕНКО
Ім’я, ПРІЗВИЩЕ

Керівник: Віталій ТИЩЕНКО
Ім’я, ПРІЗВИЩЕ

Рецензент: _____
К.т.н., доцент Ім’я, ПРІЗВИЩЕ

Київ 2025

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ

Навчально-науковий інститут кібербезпеки та захисту інформації

Кафедра управління кібербезпекою та захистом інформації

Ступінь вищої освіти бакалавр

Спеціальність 125 Кібербезпека

Освітня програма Управління інформаційною та кібернетичною безпекою

ЗАТВЕРДЖУЮ
Завідувач кафедри УКБЗІ
_____ Світлана ЛЕГОМІНОВА
"____" _____2025 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

Костенку Якову Олександровичу

(прізвище, ім'я, по батькові здобувача)

1. Тема кваліфікаційної роботи "Методика аналізу фейкових акаунтів та їхнього впливу на поширення дезінформації",
керівник кваліфікаційної роботи ТИЩЕНКО Віталій,
(ІПРІЗВИЩЕ, Ім'я,, науковий ступінь, вчене звання)

затверджені наказом Державного університету інформаційно-комунікаційних технологій від "07" березня 2025 р. №195.

2. Строк подання кваліфікаційної роботи "25" травня 2025р.

3. Вихідні дані до кваліфікаційної роботи: *дезінформація в цифровому середовищі, соціальні мережі як платформа для інформаційного впливу, фейкові акаунти та бот-мережі, методи аналізу даних у соціальних мережах.*

4. Перелік питань, які мають бути розроблені:
4.1. Проаналізувати основні характеристики фейкових акаунтів у соціальних мережах та їхню роль у поширенні дезінформації, включаючи типи, поведінкові моделі та канали впливу.
4.2. Дослідити механізми поширення дезінформації через фейкові акаунти та представити схему інформаційного впливу, що здійснюється через ці акаунти з урахуванням типових цілей і цільових аудиторій.
4.3. Визначити напрями, методи та технології виявлення фейкових акаунтів, а також розробити рекомендації щодо мінімізації їхнього впливу на інформаційний простір, з урахуванням сучасних тенденцій та інструментів автоматизованого аналізу.
Перелік ілюстративного матеріалу: *презентація PowerPoint*

5. Дата видачі завдання "5" березня 2025 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Етапи кваліфікаційної роботи	Термін виконання етапів роботи	Примітка
1.	Визначення об’єкту, предмету, мети та завдань дослідження.	18.03.2025	
2.	Збір та аналіз літератури.	29.03.2025	
3.	Аналіз особливостей управління інформаційною безпекою підприємства	08.04.2025	
4.	Дослідження основних характеристик технологій формування обізнаності й навчання персоналу.	15.04.2025	
5.	Вивчення інструментів та методів формування обізнаності й навчання персоналу з інформаційної безпеки	22.04.2025	
6.	Формулювання висновків за результатами проведеного дослідження.	29.04.2025	
7.	Оформлення роботи.	06.05.2025	
8.	Оформлення презентації.	09.05.2025	
9.	Отримання рецензії на роботу.	12.06.2025	
10.	Захист в ЕК.	__.06.2025	

Здобувач вищої освіти _____ Яків КОСТЕНКО
(підпис) *(Ім’я, ПРІЗВИЩЕ)*

Керівник
кваліфікаційної роботи _____ Віталій ТИЩЕНКО
(підпис) *(Ім’я, ПРІЗВИЩЕ)*

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ КІБЕРБЕЗПЕКИ ТА ЗАХИСТУ ІНФОРМАЦІЇ

ПОДАННЯ
ГОЛОВІ ЕКЗАМЕНАЦІЙНОЇ КОМІСІЇ
ЩОДО ЗАХИСТУ КВАЛІФІКАЦІЙНОЇ РОБОТИ
на здобуття освітнього ступеня бакалавра

Направляється здобувач Костенко Я. О. до захисту кваліфікаційної роботи
(прізвище та ініціали)
за спеціальністю 125 Кібербезпека
(код, найменування спеціальності)
освітньої програми Управління інформаційною та кібернетичною безпекою
(назва)
на тему: "Методика аналізу фейкових акаунтів та їхнього впливу на поширення дезінформації"

Кваліфікаційна робота і рецензія додаються.

Директор ННІКБЗІ _____ Євгенія ІВАНЧЕНКО
(підпис) *(Ім'я, ПРІЗВИЩЕ)*

Висновок керівника кваліфікаційної роботи

Здобувач КОСТЕНКО Яків у кваліфікаційній роботі проаналізував природу фейкових акаунтів у соціальних мережах, дослідив механізми їхнього впливу на поширення дезінформації, вивчив сучасні підходи до виявлення таких акаунтів з використанням інструментів машинного навчання та аналізу великих даних, а також розробив практичні рекомендації щодо зменшення негативного впливу дезінформаційних кампаній.

КОСТЕНКО Яків продемонстрував глибоке розуміння теми дослідження, володіння теоретичними знаннями та практичними навичками аналізу соціальних мереж, вміння застосовувати наукові методи для вирішення актуальних завдань інформаційної безпеки. Результати дослідження апробовані під час участі в двох наукових конференціях.

Все це дозволяє оцінити кваліфікаційну роботу здобувача КОСТЕНКА Якова на оцінку "_____" та присвоїти йому кваліфікацію бакалавра з кібербезпеки за освітньою програмою Управління інформаційною та кібернетичною безпекою.

Керівник кваліфікаційної роботи _____ Віталій ТИЩЕНКО
(підпис) *(Ім'я, ПРІЗВИЩЕ)*

"____" _____2025 року

Висновок кафедри про кваліфікаційну роботу

Кваліфікаційна робота розглянута. Здобувач Костенко Я.О. допускається до захисту даної роботи в Екзаменаційній комісії.

Завідувач кафедри
управління кібербезпекою та
захистом інформації _____ Світлана ЛЕГОМІНОВА
(підпис) *(Ім'я, ПРІЗВИЩЕ)*

ВІДГУК РЕЦЕНЗЕНТА
на кваліфікаційну бакалаврську роботу

здобувача вищої освіти КОСТЕНКА Якова
на тему “Методика аналізу фейкових акаунтів та їхнього впливу на поширення дезінформації”

Актуальність. У час стрімкого розвитку інформаційних технологій та зростання обсягів споживання контенту через соціальні мережі, фейкові акаунти стали ключовими інструментами у поширенні дезінформації. Усвідомлення масштабів впливу таких акаунтів є критично важливим для формування ефективної інформаційної політики та безпеки в цифровому середовищі. Кваліфікаційна робота торкається важливих аспектів цієї проблеми, що зумовлює її наукову й практичну цінність.

З огляду на зазначене дослідження проблеми формування обізнаності й навчання персоналу з інформаційної безпеки є актуальним науковим завданням.

Позитивні сторони.

1. У роботі ґрунтовно досліджено теоретичні основи фейкових акаунтів і дезінформації, їхню класифікацію та механізми впливу.
2. Представлено сучасні методики збору, обробки та аналізу даних для виявлення фейкових акаунтів із застосуванням алгоритмів машинного навчання
3. Кваліфікаційна робота оформлена відповідно до вимог. Виклад матеріалу здійснено відповідно до плану, зроблено логічні висновки. Ключові положення роботи представлено у вигляді рисунків.
4. Автор опрацював широку джерельну базу, включаючи понад 40 найменувань наукових публікацій, у тому числі англомовних джерел.
5. За результатами дослідження запропоновано рекомендації щодо ефективного навчання персоналу з питань інформаційної безпеки.

Недоліки.

Доцільним було б глибше проаналізувати типові приклади фейкових акаунтів в українському інформаційному просторі, а також надати порівняльну характеристику різних інструментів виявлення фейків з практичної точки зору.

Однак, вищезгадані зауваження не впливають на загальну позитивну оцінку кваліфікаційної роботи.

Висновок: Кваліфікаційна робота КОСТЕНКА Якова виконана на високому науково-методичному рівні, містить елементи наукової новизни та практичну значущість. Робота заслуговує оцінки «_____», а здобувач — присвоєння кваліфікації бакалавра з кібербезпеки за освітньою програмою Управління інформаційною та кібернетичною безпекою.

Рецензент:
к.т.н., доцент _____ *підпис* _____ Ім’я, ПРІЗВИЩЕ

РЕФЕРАТ

Кваліфікаційна робота присвячена аналізу фейкових акаунтів та їх впливу на поширення дезінформації в соціальних мережах. Робота складається зі вступу, трьох розділів, що містять 7 рисунків, 5 таблиць, висновків та списку використаних джерел, що містить 44 найменування. Загальний обсяг роботи становить 64 аркушів, з яких 4 аркуша займають список використаних джерел

Метою роботи є дослідження впливу фейкових акаунтів на поширення дезінформації в соціальних мережах, визначення методів і підходів для виявлення таких акаунтів і боротьби з ними.

Об'єкт дослідження - Фейкові акаунти в соціальних мережах та їх вплив на інформаційний простір.

Предмет дослідження. Особливості поширення дезінформації через фейкові акаунти в соціальних мережах.

Методи дослідження. В роботі використано методи аналізу соціальних мереж, технології виявлення фейкових акаунтів, а також методи обробки великих даних та машинного навчання для виявлення дезінформації.

У роботі проведено аналіз основних методів, через які фейкові акаунти поширюють дезінформацію. Окремо розглянуто вплив таких акаунтів на громадську думку та процеси у суспільстві. Запропоновано ряд підходів та технологій для виявлення фейкових акаунтів, включаючи використання машинного навчання для автоматизації цього процесу.

Галузь застосування: Розроблені підходи можуть бути використані для покращення систем моніторингу соціальних мереж, а також для розробки політик у боротьбі з фейковими акаунтами та дезінформацією в Інтернеті.

Ключові слова: ФЕЙКОВІ АКАУНТИ, ДЕЗІНФОРМАЦІЯ, СОЦІАЛЬНІ МЕРЕЖІ, МОНІТОРИНГ ІНФОРМАЦІЙНОГО ПРОСТОРУ, МАШИННЕ НАВЧАННЯ.

ABSTRACT

The qualification thesis is dedicated to the analysis of fake accounts and their impact on the spread of disinformation in social networks. The work consists of an introduction, three chapters containing 7 figures and 5 tables, conclusions, and a list of references comprising 44 sources. The total volume of the work is 64 pages, 4 of which are occupied by the list of references.

The aim of the study is to investigate the influence of fake accounts on the dissemination of disinformation in social networks, and to identify methods and approaches for detecting and combating such accounts.

Object of the research: Fake accounts in social networks and their impact on the information space.

Subject of the research: The peculiarities of disinformation dissemination through fake accounts in social networks.

Research methods: The study uses social network analysis methods, technologies for detecting fake accounts, as well as big data processing and machine learning methods for identifying disinformation.

The thesis analyzes the main methods through which fake accounts spread disinformation. The impact of such accounts on public opinion and social processes is examined separately. A number of approaches and technologies for detecting fake accounts are proposed, including the use of machine learning to automate this process.

Field of application: The developed approaches can be used to improve social network monitoring systems, as well as for the development of policies to counter fake accounts and disinformation on the Internet.

Keywords: FAKE ACCOUNTS, DISINFORMATION, SOCIAL NETWORKS, INFORMATION SPACE MONITORING, MACHINE LEARNING.

ЗМІСТ

ВСТУП

Актуальність теми. У сучасному світі соціальні мережі стали основними майданчиками для обміну інформацією, зокрема для поширення новин, ідей і думок. Проте разом із позитивними аспектами, такими як швидке поширення інформації, існує серйозна загроза: зростаючий вплив фейкових акаунтів. Ці акаунти використовуються для маніпулювання громадською думкою, поширення дезінформації та навіть для політичних маніпуляцій. Виявлення фейкових акаунтів є однією з головних проблем у боротьбі з дезінформацією в соціальних мережах.

Завдяки поширенню новітніх технологій, зокрема штучного інтелекту та машинного навчання, стало можливим більш ефективно виявляти фейкові акаунти, які негативно впливають на формування громадської думки. Однак методи боротьби з ними ще потребують удосконалення, оскільки фейкові акаунти стають все більш складними й можуть вводити в оману навіть добре підготовлених користувачів та автоматизовані системи.

Об'єктом дослідження є фейкові акаунти в соціальних мережах, їхні механізми функціонування та вплив на інформаційні процеси, що відбуваються в онлайн-середовищі.

Предметом дослідження є методи та підходи до аналізу фейкових акаунтів та їх вплив на процеси поширення дезінформації в соціальних мережах, а також методики для їх виявлення та нейтралізації.

Метою дослідження є розробка та апробація методики аналізу фейкових акаунтів, а також оцінка їхнього впливу на процеси поширення дезінформації. Це включає створення нових алгоритмів для автоматичного виявлення таких акаунтів та вимірювання їхнього впливу на громадську думку та суспільні настрої.

Завдання дослідження:

1. Провести глибокий аналіз існуючих методів виявлення фейкових акаунтів та їхніх характеристик.

2. Оцінити вплив фейкових акаунтів на інформаційні потоки та їхній внесок у поширення дезінформації.
3. Розробити та апробувати нову методику виявлення фейкових акаунтів, зокрема, використовуючи технології машинного навчання.
4. Оцінити ефективність запропонованих алгоритмів у реальних умовах соціальних мереж.

Методи дослідження. В процесі роботи будуть застосовані методи статистичного аналізу та математичного моделювання для оцінки ефективності алгоритмів виявлення фейкових акаунтів. Окрім того, використовуватимуться технології машинного навчання та штучного інтелекту для створення моделей, що дозволяють автоматично виявляти підозрілі акаунти та визначати їхній вплив на інформаційні потоки.

Наукова новизна полягає у вдосконаленні методу виявлення фейкових облікових записів у соціальних мережах шляхом розробки метрик ознак фейковості, що дозволяє збільшити достовірність прийняття рішення до 94 %.

Практичне значення одержаних результатів. полягає у розробці алгоритмів та програмного забезпечення виявлення фейкових облікових записів у соціальній мережі «Facebook», що дозволяє виявляти інформаційних агентів-ботів під час інформаційної війни.

РОЗДІЛ 1.ТЕОРЕТИЧНІ ОСНОВИ ФЕЙКОВИХ АККАУНТІВ ТА ДЕЗІНФОРМАЦІЇ

1.1. Поняття й класифікація фейкових акаунтів у соціальних мережах

Аналіз соціальних мереж в Інтернеті (OSNA) є однією з найперспективніших та найбільш захоплюючих тем у сучасних дослідженнях. Соціальна мережа в Інтернеті (OSN) — це своєрідна спільнота, в якій люди, організації, нації, держави, веб-сторінки та інші учасники з'єднані різноманітними зв'язками — від особистих відносин до складних гіперпосилань. Від моменту свого створення соціальні мережі змінили спосіб нашого мислення, спілкування та самовираження. Сьогодні важко уявити день без використання соцмереж, де ми не лише спілкуємось, а й дізнаємось про новини, шукаємо поради чи навіть здійснюємо покупки. Приміром, коли ми хочемо купити новий товар, ми зазвичай шукаємо відгуки в Інтернеті, а не запитуємо поради у друзів, як це було раніше. Це свідчить про те, що соцмережі стали важливим інструментом для взаємодії та впливають на наші повсякденні рішення.

Соціальні мережі, такі як Facebook, Twitter, LinkedIn, Instagram та інші, об'єднують мільйони людей по всьому світу. Їхнє зростання не лише за кількістю користувачів, а й за кількістю зв'язків між ними — неймовірне. Це справжня глобальна спільнота, де кожен може знайти однодумців, підтримати важливі ініціативи або обговорити новини. Але разом з цією популярністю та значущістю виникають і нові виклики, особливо в питаннях безпеки та конфіденційності.

Сучасний світ соціальних мереж, на жаль, не позбавлений проблем. Однією з головних загроз є створення фальшивих акаунтів, які мають безліч небезпечних наслідків. Ці акаунти можуть бути використані зловмисниками для різних цілей: крадіжки особистих даних, маніпулювання громадською думкою, поширення фальшивих новин чи навіть для організації шахрайських схем. Оскільки в соціальних мережах зберігається велика кількість особистої, професійної, соціальної та політичної інформації, це приваблює кіберзлочинців. Вони

створюють фейкові профілі, щоб отримати доступ до цих даних або зловживати довірою користувачів.

Однією з найбільших проблем є те, що соціальні мережі надають можливість швидко створювати величезні обсяги контенту, який, як показує практика, часто стає мішенню для спамерів. Ці зловмисники можуть викрадати особисті дані, поширювати шкідливий контент або маніпулювати людьми через фальшиві акаунти. Фейкові профілі можуть використовуватись для викрадення особистої, професійної чи фінансової інформації користувачів, а також для створення фальшивого трафіку на сайтах або блогах. Крім того, такі акаунти можуть сприяти розповсюдженню неправдивих новин або порнографії з метою обману або маніпулювання громадською думкою [1-2].

Усі ці загрози знову і знову підкреслюють важливість розвитку ефективних методів виявлення та боротьби з фейковими акаунтами. Тому, з одного боку, соціальні мережі відкривають перед нами нові можливості для спілкування та взаємодії, але з іншого боку, вони вимагають постійної уваги та відповідальності. Як користувачам, так і постачальникам послуг потрібно пильно стежити за безпекою, щоб захистити свою інформацію та уникнути маніпуляцій.

Створення фальшивих акаунтів є одним із найбільш поширених методів злому, а тому так важливо розуміти, як вони працюють, як їх виявляти і що можна зробити, щоб зменшити їхній вплив на суспільство. У цьому контексті дослідження фейкових акаунтів в соціальних мережах є надзвичайно актуальним і важливим для забезпечення безпеки в Інтернеті.

На рис 1.1 представлена класифікація профілів у соціальних мережах (OSN). Вона поділяється на два основних типи: легітимні профілі (Legitimate Profiles), що включають скомпрометовані та нескомпрометовані профілі, та фейкові профілі (FAKE Profiles), які мають кілька підкатегорій, зокрема клоновані профілі (Cloned Profiles), фальшиві акаунти (Sybil Accounts), боти (Bots), а також інші типи профілів, що використовуються для маніпуляцій, таких як sock puppets **і** honey profiles. Кожен з цих типів має свою специфіку використання та впливу на інформаційні потоки в мережах.

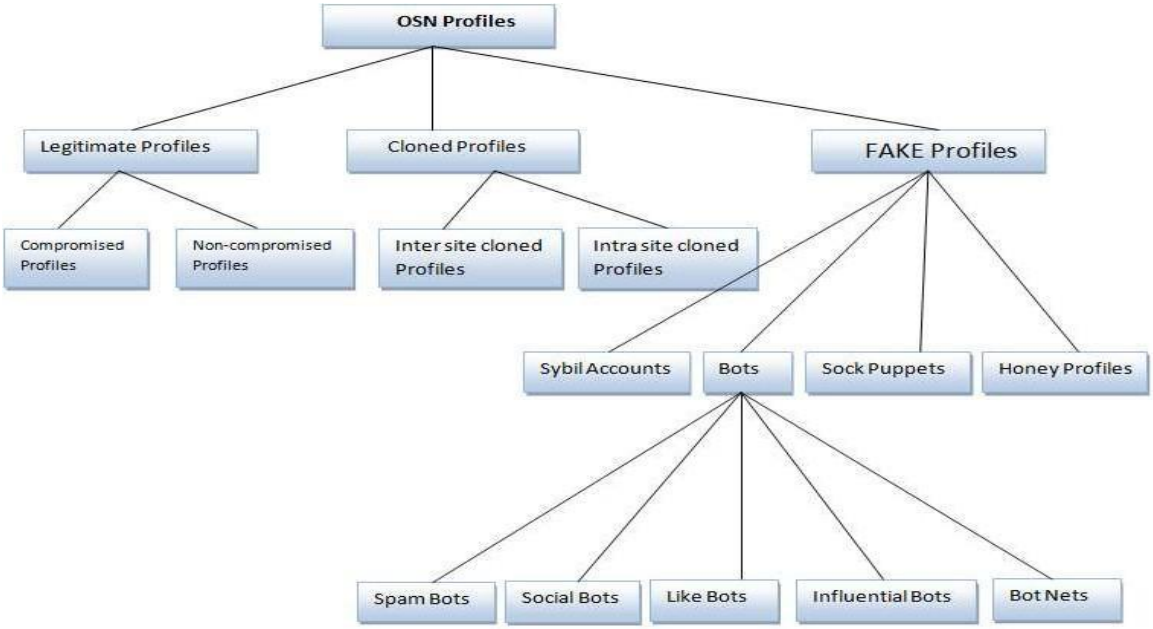

Рис.1.1. Поняття й класифікація фейкових акаунтів у соціальних мережах

Залежно від способу створення, цілей та механізму запобігання, існує кілька різновидів профілів в OSN. У цьому розділі наведено вичерпну класифікацію різних легітимних і фантомних профілів з акцентом на онлайн-соціальні мережі.

Скомпрометовані акаунти насправді є реальними акаунтами, але власники не мають повного контролю над ними, і вони втратили контроль над ними через фішера або будь-яке шкідливе забезпечення [3]. Згідно з дослідженням [4], скомпрометовані акаунти є найскладнішим типом акаунтів для виявлення. Інше нещодавнє дослідження стверджує, що понад 97% профілів є скомпрометованими, а не підробленими [5]. Фейкові профілі, як правило, створюються для крадіжки облікових даних у реальних користувачів, а потім фейкові профілі покидаються або деактивуються.

Скомпрометовані профілі мають велику цінність, оскільки вони вже встановили певний рівень довіри в мережі, а отже, їх нелегко виявити і видалити постачальникам послуг. Зловмисники зазвичай використовують скомпрометовані профілі з особливою обережністю, щоб скористатися рівнем довіри. Автори в [6] представили підхід до виявлення скомпрометованих акаунтів двох популярних соціальних мереж, Facebook і Twitter, шляхом

виявлення профілів, які демонструють раптові зміни в поведінці, за допомогою статистичного моделювання та виявлення аномалій. Facebook має систему відновлення зламаних акаунтів після повідомлення про злом. На сторінці допомоги Facebook є опція "мій акаунт був зламаний".

Клонування профілю - це крадіжка особистих даних з існуючого профілю користувача і створення нового фальшивого профілю з використанням викрадених облікових даних. Іншими словами, можна сказати, що клонування профілю - це процес крадіжки приватної інформації жертви з метою створення ще одного профілю, який може отримати приватну інформацію друзів жертви. Ці атаки називаються Identity Clone Attacks (ICA) [7-8], які бувають двох типів, а саме: односайтовий та міжсайтовий клонування профілів. Зловмисники, як правило, добре фінансуються, є кваліфікованими фахівцями і мають у своєму розпорядженні майже все необхідне для контролю над скомпрометованими та зараженими акаунтами [9-10]. Зловмисник може бути незнайомою людиною, але статистика показує, що зловмисник знає і може бути її родичем, другом чи колегою [11-12] (рис 1.2).

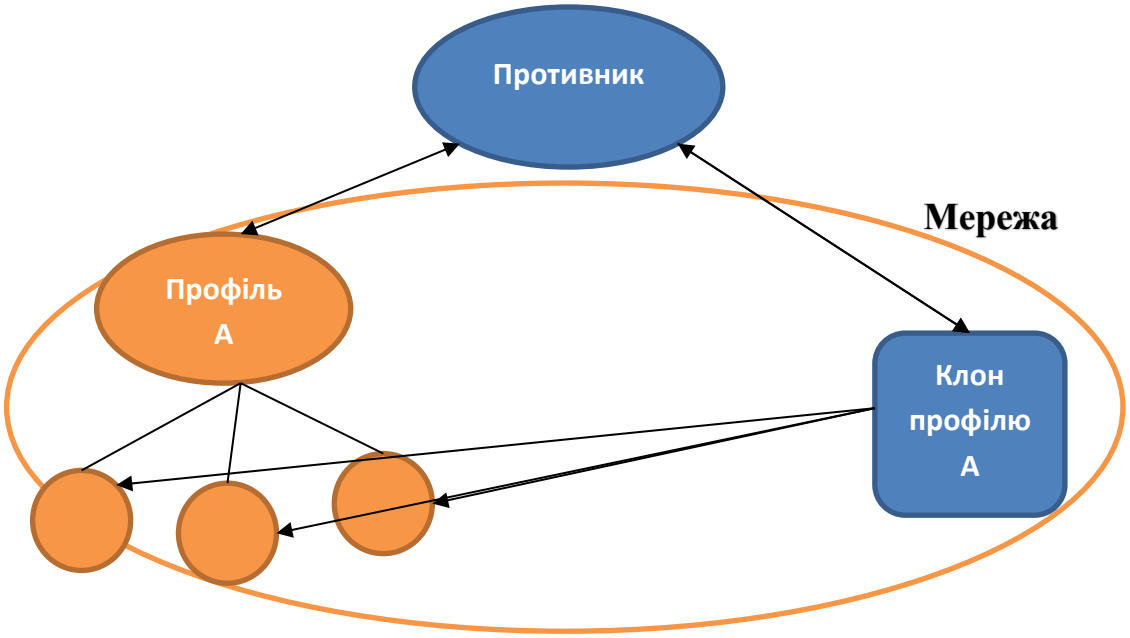

Рис. 1.2. Клонування профілю всередині сайту або того самого сайту

Фейкові профілі багато в відрізняються від клонованих. У випадку з клонованими профілями зловмисник створює ще один профіль на основі вже існуючого, чого не можна сказати про фейкові профілі. Клоновані профілі здебільшого створюються для отримання інформації про жертву або його/її друзів, тоді як фейкові профілі використовуються для різних інших цілей, таких як спам, реклама тощо. Деякі люди створюють фальшиві профілі просто для того, щоб мати ще один обліковий запис, тоді як деякі створюють кілька облікових записів навмисно, щоб потрапити в підгрупу людей. Існує два способи створення фейкових профілів: один - за допомогою написання скрипта, а інший - шляхом ручного створення ще одного акаунта. Існує три основні причини для створення фейкових профілів: По-перше, постачальники послуг OSN дозволяють один обліковий запис на одне мобільне з'єднання або на одну електронну адресу, і щоб подолати цей ліміт, люди створюють ще один обліковий запис, використовуючи різні електронні адреси або номери телефонів. По-друге, це підвищення популярності або рівня довіри серед інших користувачів. Третя - поширювати спам серед реальних користувачів. Фейкові суб'єкти існують скрізь в Інтернеті: соціальні мережі, торгові сайти, дискусійні блоги та форуми, сайти знайомств, банківські системи тощо.

Бот - це комп'ютерна програма, яка виробляє певні дані для взаємодії з людьми, особливо з тими, хто користується Інтернетом (користувачами), з метою змінити їхню поведінку [13-18]. Понад 60% всіх веб-даних генеруються ботами [18]. Інтернет-боти, також відомі як веб-роботи або просто боти, - це комп'ютерні програми, які швидко і автоматично виконують різні завдання, які не під силу людині. В основному боти були розроблені для того, щоб допомогти людині прискорити її роботу та зробити її автоматичною. Основна роль ботів полягала в автоматичному агрегуванні контенту з різних новинних джерел, роботі в якості автоматичного відповідача на запити користувачів, в якості медичного експерта для вирішення питань, пов'язаних зі здоров'ям, а також в якості автоматичного для подорожей. Але сьогодні ботами зловживають у

різних сферах. У соціальних мережах ботів використовують для ретвітів постів без перевірки джерела, щоб зробити їх вірусними.

1.2. Механізми формування та поширення дезінформації

Дезінформація може бути класифікована за різними критеріями. По-перше, з точки зору автентичності змісту, дезінформація може бути поділена на повністю неправдиву інформацію та частково неправдиву інформацію. Повністю неправдива інформація – це зміст, який є повністю вигаданим і не має жодного фактичного підґрунтя. Частково неправдива інформація, натомість, перебільшує, вириває з контексту або спотворює реальну інформацію, що може призвести до хибних тлумачень. Ця класифікація допомагає визначити потенційну небезпеку інформації та розробити відповідні заходи протидії. По-друге, з точки зору мети поширення, дезінформація може бути класифікована як зловмисна дезінформація та ненавмисна дезінформація. Зловмисна дезінформація часто має на меті маніпулювання громадською думкою, розпалювання соціальної паніки або націлена на конкретних осіб чи групи, тоді як ненавмисна дезінформація часто виникає через непорозуміння, поширення помилкової інформації або відсутність необхідних базових знань. Це розрізнення допомагає зрозуміти мотиви дезінформації, що сприяє вжиттю цілеспрямованих запобіжних заходів. Нарешті, за засобами поширення дезінформацію можна класифікувати як дезінформацію, що поширюється через традиційні медіа або через соціальної мережі. Завдяки швидкості, інтерактивності та широкому охопленню соціальних мереж, дезінформація на цих платформах, як правило, привертає більше уваги користувачів і частіше поширюється, тим самим посилюючи свій вплив. Інформація в соціальних мережах поширюється швидко, з високим рівнем інтерактивності між користувачами, що підвищує видимість і сприйняту достовірність дезінформації, роблячи її поширення більш помітним [19].

Дизайн алгоритмів соціальних мереж відіграє вирішальну роль у поширенні неправдивої інформації. Ці алгоритми, оцінюючи поведінку користувачів,

особливості контенту та моделі взаємодії, визначають видимість та характеристики поширення інформації, що глибоко впливає на поширення неправдивої інформації. В табл. 1.1 наведені основні механізми поширення фейкової інформації

Таблиця 1.1

Механізми поширення фейкової інформації

Механізм	Опис
Пріоритет інформації та емоційно-залежне залучення	Алгоритми соціальних мереж часто віддають пріоритет контенту, який викликає сильні емоційні реакції користувачів. Це підвищує рівень взаємодії (лайки, коментарі, репости), що сприяє швидкому поширенню фейкової інформації.
Фільтрація контенту та замкнуті інформаційні кола	Алгоритми соціальних мереж створюють умови, за яких користувачі стикаються лише з інформацією, що підтверджує їхні погляди. Це обмежує доступ до різноманітних точок зору та сприяє поширенню фейкової інформації в обмежених групах без її перевірки.

Алгоритми соціальних мереж часто надають пріоритет контенту, який викликає сильні емоційні реакції у користувачів, особливо фейковій інформації, яка часто подається в перебільшеному або провокативному вигляді. Цей механізм, що базується на емоціях, сприяє більшій залученості користувачів через взаємодію, обмін інформацією та коментування, що потенційно може призвести до швидкого поширення в соціальних мережах. Зокрема, алгоритми підсилюють видимість такого контенту, що викликає високу залученість, розміщуючи фейкову інформацію на видному місці в інформаційних потоках і створюючи «гарячі точки», які ще більше прискорюють її поширення. Ця модель поширення, що базується на емоціях, також має тенденцію викликати ідентифікацію та резонанс серед користувачів, що дозволяє фейковій інформації поширюватися в різних соціальних колах, що призводить до переробки та еволюції контенту та посилення його сприйняття як легітимного в суспільній свідомості [20-21] .

Алгоритмічні операції також призводять до фільтрації контенту та замкнутих інформаційних кіл, що обмежують досвід користувачів соціальних

мереж інформацією, яка відповідає їхнім власним поглядам. Це явище обмежує обмін різноманітними точками зору, що призводить до поширення неправдивої інформації в певних спільнотах без будь-якого сумніву. З часом це явище не тільки впливає на критичне мислення користувачів, але й може призвести до поляризації соціальних уявлень, ще глибше закріплюючи неправдиву інформацію в групах. У таких умовах користувачі більш схильні сприймати фейкову інформацію як факт, що ще більше поглиблює соціальну роз'єднаність і непорозуміння. Крім того, фільтрація контенту та замкнуті інформаційні кола суттєво обмежують доступ користувачів до зовнішньої інформації, перешкоджаючи поширенню достовірної інформації та сприяючи сліпій довірі та дотриманню фейкової інформації. Розуміння ролі алгоритмів у поширенні фейкової інформації, зокрема їхнього впливу на сприйняття користувачів, є надзвичайно важливим для вирішення проблем, пов'язаних із поширенням фейкової інформації.

1.3. Параметри та індикатори впливу фейкових акаунтів

Соціальні мережі — це простір, де люди можуть обмінюватися ідеями, висловлювати свої думки та спілкуватися з іншими. Але з розвитком технологій з'явилися нові форми маніпуляцій, зокрема через фейкові акаунти та боти, які активно впливають на поведінку користувачів і поширення інформації. Боти — це автоматизовані програми, які виконують завдання без участі людини, і, хоча вони створюють ілюзію реальної взаємодії, насправді вони лише виконують алгоритми. Ці автоматизовані акаунти здатні працювати набагато швидше, ніж люди, і можуть здійснювати різні дії: від створення контенту до участі у коментарях та поширенні повідомлень [22].

Боти здатні маскувати свою природу, що дає їм змогу успішно маніпулювати користувачами. Вони часто використовуються для поширення фейкових новин, маніпуляцій і навіть для впливу на виборчі процеси. Наприклад, під час виборів у США 2010 року, боти активно направляли людей на веб-сайти,

що містили неправдиву інформацію. Такі маніпуляції повторювались і на виборах 2020 року, а також під час пандемії COVID-19, коли фейкові новини поширювались на величезні аудиторії.

За статистикою, у 2014 році боти генерували понад 50% всього інтернет-трафіку. Вже до 2019 року шкідливий трафік, згенерований ботами, зріс до 25%. Це значний показник, який свідчить про вплив ботів на цифровий простір. Вони не тільки спотворюють інформацію, але й активно розповсюджують її, створюючи фальшиві соціальні тренди та формуючи громадську думку, орієнтуючи її на неправдиві або маніпулятивні ідеї.

Розуміння того, як фейкові акаунти впливають на соціальні мережі, дуже важливе для боротьби з поширенням дезінформації. Для цього вчені розробили кілька математичних моделей, які описують механізми поширення фейкової інформації. Наприклад, використовуються епідемічні або стохастичні моделі, що дозволяють прогнозувати, як швидко поширюється дезінформація серед користувачів на основі певних параметрів. Однак ці моделі не завжди враховують особливості поведінки ботів. Виявляється, що боти мають здатність впливати на велику кількість людей значно швидше і з більшою стійкістю, ніж звичайні користувачі.

Моделювання поведінки ботів стало важливим напрямком у дослідженнях, оскільки боти можуть виступати в ролі "суперрозповсюджувачів" фейків, швидко та ефективно просуваючи маніпуляції серед широкої аудиторії. Це явище важко зафіксувати через те, що боти часто мають низьку видимість в загальних мережах, однак вони здатні значно впливати на певні теми, таких як політика, соціальні питання чи економіка.

На відміну від звичайних користувачів, боти можуть мати постійну активність і постійно впливати на різні частини мережі. Це означає, що їхній вплив може бути більш системним і тривалим. Зокрема, за допомогою теорії соціального впливу, можна виміряти не лише силу впливу бота, але й його упередженість. Це допомагає краще зрозуміти, як і чому боти можуть сприяти поширенню чуток, навіть якщо ці чутки спочатку виглядають малозначними.

У традиційних моделях, таких як PageRank або центральність у мережі, фокус робиться на визначенні важливості акаунтів на основі їхньої взаємодії з іншими користувачами. Однак ці метрики не завжди працюють для ботів. Боти, як правило, займають іншу нішу в мережах, що робить їх менш видимими, а стандартні метрики не дають точних результатів для таких акаунтів.

Враховуючи це, були розроблені нові моделі, що дозволяють розрізняти типи користувачів, такі як реальні люди та боти. Запропонована модель дозволяє вимірювати вплив різних акаунтів на процеси поширення чуток, враховуючи різні рівні швидкості поширення для кожного типу користувачів. Це дозволяє краще розуміти, як фейкові акаунти можуть змінювати динаміку мереж, та дає змогу прогнозувати, як швидко і яким чином буде поширюватися інформація.

За допомогою цієї моделі можна оцінити вплив ботів на соціальні мережі, виявити потенційні загрози маніпуляцій і розробити більш ефективні стратегії боротьби з фейковими акаунтами. Технології, що стоять за моделюванням поведінки ботів, відкривають нові можливості для вивчення того, як ці акаунти впливають на громадську думку, як формуються соціальні тренди та як можна запобігти поширенню дезінформації [24].

Для моделювання процесу поширення інформації в соціальних мережах зазвичай використовуються два основні підходи: часовий та керований даними. З розвитком технологій машинного та глибокого навчання, дослідники все більше звертаються до підходів, які керуються даними, завдяки їхній високій точності в прогнозуванні. Однак епідемічні моделі залишаються найпоширенішими, оскільки вони є простими, ефективними і дають можливість чітко інтерпретувати результат. Ці моделі були адаптовані для аналізу процесів поширення інформації, оскільки мають схожі механізми з поширенням епідемій.

Існують кілька основних епідемічних моделей, які використовуються для моделювання поширення інформації. Наприклад, модель "Вразливі–інфіковані" (SI), модель "Вразливі–інфіковані–чутливі" (SIS), модель "Вразливі–інфіковані–одужані" (SIR) та модель "Вразливі–інфіковані–одужані–чутливі" (SIRS). Кожна

з цих моделей має свої унікальні характеристики, що визначають, як інфіковані користувачі взаємодіють з іншими і як це впливає на процес поширення.

Серед цих моделей найбільше поширення отримала модель SIR (вразливі–інфіковані–одужані), яка активно використовується для аналізу поширення чуток та дезінформації. У її базовій формі вона описує, як інфекція (в даному випадку – чутка або фейкова інформація) поширюється серед користувачів. Проте, у сучасних дослідженнях ця модель була розширена для того, щоб врахувати додаткові фактори, які можуть впливати на процес поширення чуток, наприклад, вплив ботів чи суперрозповсюджувачів інформації.

Розширення моделі SIR можна поділити на три основні категорії:

- Додавання нових вузлів: У цій категорії моделі SIR доповнюються новими станами або вузлами, які дозволяють більш детально моделювати поведінку користувачів у різних умовах (наприклад, "активний інфікований", "пасивний спостерігач" тощо).

- Модифікація класичної моделі SIR: Цей підхід включає в себе додавання нових факторів, таких як поведінка користувачів у мережах або механізми маніпуляції контентом, що можуть впливати на поширення фейкової інформації.

- Розгалуження основних вузлів: Тут базова модель SIR розбивається на додаткові підвузли, що дає змогу точніше моделювати різні етапи інфікування (наприклад, на етапи первинного зараження, повторного поширення або "одужання" від дезінформації).

Таким чином, хоча епідемічні моделі залишаються основними інструментами для аналізу поширення фейкової інформації, сучасні дослідження активно інтегрують підходи, що базуються на даних, для досягнення більш точних прогнозів. Це дозволяє краще розуміти, як саме поширюється дезінформація та які фактори найбільше впливають на її розповсюдження в соціальних мережах.

Зокрема, важливо враховувати роль таких аспектів, як поведінка ботів, супер розповсюджувачів та структурні властивості соціальних мереж, які

можуть значно змінити хід поширення інформації та вплинути на її сприйняття користувачами. Порівняння епідемічних моделей та підходів, керованих даними, дозволяє отримати більш глибоке розуміння динаміки фейкової інформації та знайти ефективні способи боротьби з її поширенням.

Одним із ключових аспектів при моделюванні впливу фейкових акаунтів на поширення дезінформації є детальне вивчення параметрів, які визначають їх роль у соціальних мережах. Враховуючи, що фейкові акаунти можуть активно та агресивно поширювати дезінформацію, важливо аналізувати не лише їхню частоту публікацій, але й типи взаємодії з іншими користувачами, що може значно вплинути на динаміку поширення неправдивої інформації. Зокрема, частота публікацій фейкових акаунтів, їхня здатність створювати штучну популярність через взаємодії з іншими акаунтами, такі як лайки, коментарі чи репости, є важливими індикаторами того, як швидко і інтенсивно поширюється фейкова інформація. Оцінка цих параметрів дозволяє точніше змоделювати процес поширення дезінформації, зокрема в контексті того, як зміна популярності публікацій може впливати на її сприйняття в соціальних мережах.

Іншим важливим аспектом є здатність фейкових акаунтів маніпулювати контентом через використання спеціально розроблених алгоритмів, які дозволяють змінювати популярність певних публікацій або тем. Така маніпуляція може включати створення так званих "віртуальних хвиль" або поширення неправдивої інформації через різні соціальні механізми, що в кінцевому підсумку підвищує видимість певних постів серед користувачів. Важливо враховувати не тільки автоматизовані системи, які підвищують ефективність поширення фейкових новин, але й людські фактори, зокрема поведінку людей, які активно взаємодіють з фейковими акаунтами, без усвідомлення їхнього впливу на громадську думку.

Особливу увагу варто приділити ролі суперрозповсюджувачів, які є користувачами з великою кількістю підписників або значним впливом на інших учасників мережі. Ці користувачі мають можливість не лише швидко поширювати дезінформацію, а й створювати ефект "авторитету", що додатково

сприяє поширенню неправдивих новин. Вони можуть здійснювати вплив на громадську думку через повторне поширення вже існуючих публікацій або активне коментування, що значно збільшує видимість певної інформації.

Інтеграція таких параметрів у моделі поширення дезінформації дозволяє створювати більш точні прогнози щодо того, як фейкові акаунти можуть змінювати інформаційну картину в мережі. Це дає змогу глибше розуміти, як фейкові акаунти можуть впливати на формування громадської думки, зокрема у політичних кампаніях, питаннях національної безпеки чи у моменти важливих соціальних змін. У результаті, правильне використання цих параметрів в наукових і практичних дослідженнях може дозволити не лише розпізнавати фейкові акаунти, а й розробляти ефективні стратегії для боротьби з їхнім впливом на інформаційний простір, що в свою чергу дозволить зменшити негативний ефект від поширення дезінформації в цифрових середовищах.

Висновки до першого розділу

У результаті теоретичного аналізу було встановлено, що фейкові акаунти є суттєвим інструментом поширення дезінформації в цифровому середовищі. Встановлено класифікацію фейкових акаунтів за структурою, способом створення, функціональністю та типом впливу на користувачів. Окрему увагу приділено поведінковим моделям бот-мереж, їх здатності масштабно маніпулювати інформаційними потоками, що підтверджується емпіричними спостереженнями з інформаційних кампаній останніх років. Визначено ключові параметри впливу фейкових акаунтів, серед яких частота публікацій, інтенсивність взаємодій та участь у формуванні інформаційного резонансу. Сформульовано фундаментальну роль соціальних мереж як тригера швидкого розповсюдження фейкової інформації, спричиненого алгоритмічною фільтрацією контенту та емоційною залученістю аудиторії.

РОЗДІЛ 2. МЕТОДИКА ВИЯВЛЕННЯ ТА АНАЛІЗУ ФЕЙКОВИХ АККАУНТІВ

2.1. Джерела даних і способи їх збору

У цифровому середовищі соціальні мережі стали інструментом для комунікації та обміну інформацією. Однак разом із зростанням популярності цих платформ, на них з’явилися й нові загрози, зокрема фейкові акаунти, які активно використовуються для маніпулювання громадською думкою, поширення дезінформації та здійснення інших шкідливих дій. Одним із найважливіших етапів у дослідженні фейкових акаунтів є збір та аналіз даних, які дозволяють виявляти їх, а також вивчати їх вплив на соціальні мережі та поведінку користувачів.

Джерела даних для виявлення фейкових акаунтів можуть бути різними: від відкритих профілів користувачів до спеціалізованих інструментів для збору метаданих, таких як API соціальних мереж або алгоритми машинного навчання, що аналізують поведінку акаунтів. Ці дані дають змогу виявляти характерні ознаки фейкових профілів, зокрема аномальні патерни взаємодії, підозрілі обсяги контенту або маніпуляції в соціальних мережах.

На рис 2.1 представлено основні методи збору даних.

Рис 2.1. Підходи до збору данних

Збір даних через API є одним з найбільш поширених методів у дослідженнях соціальних мереж, і є рекомендованою практикою для аналізу соціальних платформ. Постачальники соціальних мереж зазвичай надають розробникам і дослідникам доступ до спеціальних бібліотек та пакетів, що спрощують процес збору даних. Для отримання необхідної інформації більшість дослідників створюють власні скрипти та програмні рішення, використовуючи API для інтеграції з різними соціальними мережами. Кожна соціальна платформа має свій специфічний API. Наприклад, для платформи Facebook існує GRAPH API, що дозволяє взаємодіяти з додатками та збирати дані про користувачів. Для платформи Twitter розроблений Twitter API, що забезпечує доступ до інформації про твіти, фоловерів та іншу активність користувачів.

В рамках наукових досліджень, багато вчених і розробників створюють інструменти для автоматичного збору даних з соціальних мереж, що допомагає здійснювати більш детальний аналіз. Наприклад, у дослідженні [25] використовували дані з Facebook і Twitter для виявлення спаму в цих соціальних мережах. У роботі автори розробили скрипт для підключення до вже існуючих профілів користувачів, щоб отримати всю необхідну інформацію для дослідження діяльності акаунтів.

У своїх дослідженнях автори активно використовують соціальні мережі як основне джерело даних для вивчення поведінки користувачів та виявлення фейкових акаунтів. Наприклад, один з дослідників розглянув Facebook як цільову соціальну мережу і розробив спеціалізований додаток для збору необхідної статистичної інформації з профілів користувачів. У роботі автор створив додаток під назвою "NetVizz", який дозволяє здійснювати збирання даних про користувачів Facebook, включаючи інформацію про їхні взаємодії та соціальні зв'язки.

Аналогічно, в інших дослідженнях також використовуються API для збору даних з соціальних мереж. Це дозволяє ефективно отримувати великий обсяг інформації для подальшого аналізу та виявлення різних тенденцій в поведінці користувачів.

На рис 2.2 зображено концептуальний вигляд типового краулера даних, який використовується для вилучення інформації із соціальних мереж. Такий інструмент дозволяє автоматизувати процес збору даних, що є важливим аспектом при дослідженні великих онлайн-спільнот.

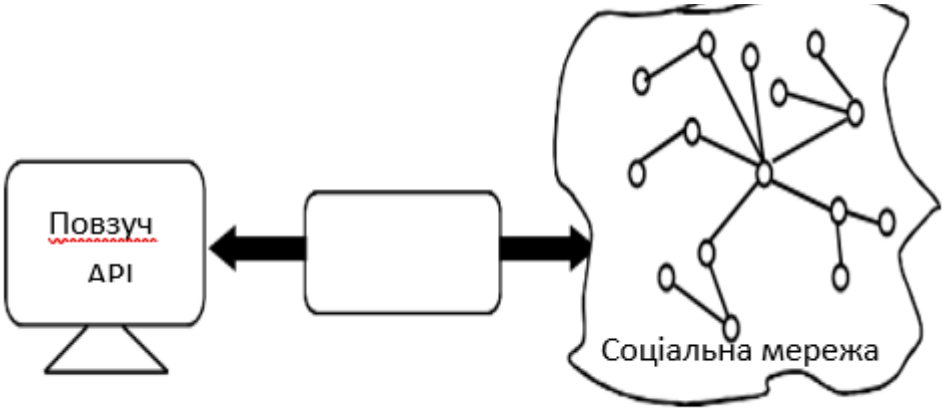

Рис 2.2 Типовий сканер даних OSN

Краулер — це програма, яка взаємодіє з соціальними мережами через API, виконуючи роль збору та витягнення даних з цільових профілів. У контексті дослідження фейкових акаунтів, краулери є важливими інструментами для автоматизованого збору великої кількості даних. Краулери зазвичай використовують два основні алгоритми для пошуку та збору інформації: BFS (Breadth First Search) та DFS (Depth First Search).

У разі використання BFS-краулерів процес збору даних починається з вибору цільового профілю. Після цього краулер виявляє найближчих сусідів цього профілю, витягуючи їхні публічні атрибути та дані. Це дозволяє створити широку мережу взаємодій та взаємозв'язків, що дає змогу дослідникам отримати значний обсяг інформації про користувачів.

Натомість DFS-краулери працюють дещо інакше: вони спочатку збирають атрибути сусідів цільового профілю до певного рівня, а потім звертаються до безпосередніх попередників, витягуючи їхні дані. Однією з головних переваг DFS-методу є те, що він вимагає менше пам'яті порівняно з BFS-краулерами. Однак, для роботи з великими соціальними мережами, краулери на основі BFS є

більш оптимальним вибором, оскільки вони дозволяють швидше обробляти великі обсяги інформації та зібрати більше даних за коротший час.

Ще одним підходом є випадковий вибір профілів для збору даних. Цей метод може бути корисним для виявлення неочевидних закономірностей та спостережень, які не завжди виявляються при обробці даних за певною схемою або алгоритмом. Однак при такому підході необхідно бути обережним, щоб не пропустити важливі профілі, які можуть містити значущу інформацію для дослідження. В одному з досліджень було запропоновано ще одну спробу збору даних із Facebook, де розроблений краулер на основі симуляції взаємодії між акаунтами усунув недоліки попередніх підходів, таких як обмеження кількості оброблених профілів (лише 400 профілів). У цьому дослідженні підкреслено важливість питання конфіденційності користувачів. Зокрема, понад 35% користувачів Facebook змінили стандартні налаштування конфіденційності своїх акаунтів, що є важливим аспектом при зборі та аналізі персональних даних.

Після збору даних наступним етапом є їх маркування. У багатьох випадках це здійснюється вручну, з використанням краудсорсингу — процесу, при якому кілька людей в Інтернеті допомагають генерувати або маркувати дані. Одним з популярних інструментів для краудсорсингу є Amazon Mechanical Turk (MTurk), що є платформою, на якій бізнес-організації можуть отримати дані, створені людьми (так званими "туркерами"). У контексті аналізу фейкових акаунтів, краудсорсинг застосовується для маркування профілів, де людина, яка має досвід або експертизу, перевіряє дані профілю та надає відповідну мітку (наприклад, «фейковий» або «реальний»). Після маркування ці дані використовуються як навчальні дані в різних методах машинного навчання для подальшого вдосконалення алгоритмів виявлення фейкових акаунтів. Такі підходи дозволяють значно покращити точність виявлення фейкових акаунтів, оскільки моделі можуть навчатися на реальних прикладах.

У багатьох випадках дослідники стикаються з проблемою недоступності даних через обмеження, накладені постачальниками послуг ОСН (онлайн соціальних мереж), які часто обмежують автоматичне сканування та збирання

даних з метою забезпечення безпеки та конфіденційності особистої інформації користувачів. У таких ситуаціях для тестування алгоритмів та інструментів можна згенерувати штучні дані, які будуть служити для перевірки моделей, що вивчають поведінку фейкових акаунтів.

Штучні дані можуть бути створені за допомогою різноманітних інструментів, які використовують відомі статистичні моделі та параметри реальних соціальних мереж. Наприклад, знаючи степеневий розподіл, коефіцієнт кластеризації, середню центральність користувачів і інші статистичні показники, можна сформувати фіктивний набір даних, що імітує реальні умови роботи соціальних мереж. Цей підхід дозволяє розробникам створювати тестові дані, що відповідають статистичним характеристикам реальних мереж, і використовувати їх для тестування алгоритмів без ризику порушення конфіденційності або доступу до справжніх користувацьких даних.

Одним із методів для генерації синтетичних даних є використання генераторів даних, таких як GEDIS Studio та Databene Benerator, які дозволяють створювати набори даних з характеристиками, подібними до реальних, на основі заздалегідь заданих параметрів. Завдяки таким інструментам можна моделювати різні сценарії взаємодії між акаунтами, зокрема для тестування алгоритмів виявлення фейкових акаунтів та інших методів аналізу.

Штучні дані, хоча й є корисними для тестування та аналізу, також мають свої обмеження. Одним із основних недоліків є те, що поведінка та статистика мережі можуть змінюватися з часом, тому результати, отримані на основі синтетичних даних, не завжди будуть точними та відображатимуть актуальний стан конкретної мережі. Технології та алгоритми, які використовуються для моделювання даних, можуть бути непридатними для аналізу в разі значних змін у самій соціальній мережі, таких як зміни в алгоритмах, модерації контенту або стратегіях взаємодії. Тому, хоча штучні дані є корисним інструментом, їхні результати повинні розглядатися як наближені до реальних, а не як абсолютна правда. Однак, якщо немає доступу до оригінальних даних на момент

тестування, синтетичні дані стають найкращою альтернативою, що дозволяє провести дослідження та експерименти в умовах обмеженої інформації.

Ще одним важливим методом збору даних є підхід на основі ботів, який є одним із найпопулярніших і найбільш відомих способів для отримання інформації з соціальних мереж. У цьому підході використовуються комп'ютерні програми — краулери, написані на різних мовах програмування, таких як Java, Python тощо. Боти взаємодіють з API соціальних мереж, автоматизуючи процес сканування та вилучення даних, що значно пришвидшує збору інформації в порівнянні з ручними методами. Завдяки таким автоматизованим програмам можна збирати величезні обсяги даних про користувачів, їхні взаємодії, соціальні зв'язки та інші атрибути, що важливі для аналізу поведінки акаунтів. Боти дозволяють ефективно збирати дані з великих мереж, таких як Facebook, Twitter, Instagram, і здійснювати їх подальший аналіз. Хоча цей підхід є досить ефективним, він вимагає дотримання правил використання API та може зіткнутися з обмеженнями доступу через політики конфіденційності соціальних мереж [26].

2.2 Інструменти та алгоритми для автоматичного виявлення

У процесі виявлення фейкових акаунтів у соціальних мережах використовуються різні методи та алгоритми. Використання обробки природної мови (NLP) виявилось ефективним підходом для виявлення фейкових акаунтів у соціальних мережах. Зокрема, Акйон і Есат Калфаоглу [32] розробили модель штучної нейронної мережі для детекції шахрайських акаунтів в Instagram. Для вирішення проблеми дисбалансу класів у даних було застосовано алгоритм SMOTE NC. Результати тестування показали, що точність запропонованої моделі становить 96%. В іншому дослідженні Wanda та Jie [16] запропонували модель глибокого навчання під назвою DeepProfile, спрямовану на виявлення фальшивих акаунтів на платформах соціальних мереж. Вони використовували набір даних з 3050 унікальних користувачів та змінили пул-шар CNN для

покращення точності. Модель продемонструвала високу ефективність з точністю 94%, відтворюваністю 93,21% і F1-оцінкою 93,42%. Втрата моделі склала 0,214, а значення ROC знаходились у діапазоні 0,950–0,959.

Крім того, дослідники [27-30] розробили та навчили нову модель нейронної мережі для виявлення автоматизованого спаму і фальшивих акаунтів в Instagram. Застосований метод показав хороші результати, досягнувши точності 93% та прецизійності 91%. За результатами літературного огляду можна зробити висновок, що дослідники застосовували як традиційні методи машинного навчання, так і методи глибокого навчання для виявлення фейкових акаунтів на платформах соціальних мереж. Однак немає однозначних рекомендацій щодо вибору конкретної моделі, оскільки вибір методу значною мірою залежить від розміру набору даних.

Традиційні методи машинного навчання включають алгоритми, які використовуються для вирішення конкретних задач на основі навчання на наданих даних. Ці методи залежать від вибору алгоритму та вхідних даних, які визначають експерти в предметній області. Виявлення фейкових акаунтів у соціальних мережах також використовує кілька класичних підходів машинного навчання.

Дерево рішень, це метод керованого навчання, що застосовується як для задач регресії, так і для класифікації. Алгоритм організує дані у вигляді дерева, де центральний вузол відповідає за початкове рішення, а подальші вузли та гілки вказують на можливі варіанти рішень [31].

Машина опорних векторів (SVM). Цей лінійний алгоритм застосовується для класифікації та регресії, побудовуючи роздільну лінію або гіперплощину, що розділяє групи в даних. Він дозволяє ефективно відокремлювати класи та оптимізувати простір між ними .

Випадковий ліс. Це метод, що поєднує результати кількох дерев рішень для отримання точнішого результату. Він широко використовується для класифікації та регресії завдяки своїй гнучкості та здатності обробляти великі обсяги даних .

Наївний Байєс. Алгоритм класифікації, заснований на теоремі Байєса, прогнозує цільове значення на основі ймовірностей кожного класу. Він обчислює ймовірність належності до кожного класу та вибирає клас з найбільшим значенням ймовірності [32].

Логістична регресія. Цей метод аналізує залежність між незалежними змінними та прогнозує ймовірність бінарного результату. Він широко використовується для оцінки ймовірності на основі наявних даних .

Екстремальне градієнтне підсилення (XGBoost). Це потужна реалізація алгоритму градієнтного підсилення, що будує залишкові дерева. Алгоритм використовує міри подібності між листям і батьківськими вузлами дерев для вибору важливих змінних і покращення точності класифікації .

SVM (Support Vector Machine) — це потужний і широко застосовуваний алгоритм для класифікації та регресії, який використовує геометричні методи для побудови моделі, що розділяє класи даних. Основна ідея алгоритму полягає у знаходженні так званої гіперплощини, яка оптимально розділяє простір на два класи. В алгоритмі SVM кожен елемент вхідних даних представляється точкою у багатовимірному просторі, а гіперплощина служить розділовою лінією між двома класами.

Ключовими елементами в SVM є опорні вектори — це ті точки, що розташовані найближче до гіперплощини. Для того щоб знайти ідеальну роздільну гіперплощину, алгоритм обчислює інтервал (відстань між опорними векторами та самою гіперплощиною), і головною метою є максимізація цієї відстані. Чим більший інтервал, тим краще модель здатна класифікувати нові дані. Гіперплощина, яка максимізує цей інтервал, вважається ідеальною роздільною лінією для даних.

В алгоритмі SVM для двовимірного набору даних роздільною лінією є простий відрізок, тоді як для більш складних (n-вимірних) даних гіперплощина буде (n-1)-вимірним підпростором, який ефективно розділяє дані на два класи. Таким чином, для кожної задачі, в залежності від кількості ознак (вимірностей),

SVM визначає найбільш оптимальну гіперплощину, яка дозволяє найкраще класифікувати дані.

Коли ми застосовуємо алгоритм SVM для виявлення фейкових акаунтів у соціальних мережах, задача полягає в побудові моделі, яка на основі вхідних даних (ознаок акаунтів) розрізняє акаунти на реальні та фейкові. У цьому випадку кожен акаунт представлений набором ознак, таких як активність користувача, кількість друзів, тип контенту, наявність фан-сторінок, взаємодія з іншими користувачами, та інші параметри, які можуть свідчити про те, чи є акаунт фейковим чи реальним. Для цього алгоритм SVM приймає ці ознаки як вхідні дані та будує гіперплощину, яка відокремлює акаунти на дві категорії: фейкові та реальні.

Наприклад, для класифікації акаунтів на платформі, такій як Instagram або Facebook, ми можемо використовувати 9 вхідних параметрів, які відповідають за ключові характеристики акаунтів, такі як активність, кількість підписників, частота публікацій, тип контенту (спам чи оригінальний), взаємодія з іншими акаунтами, час реєстрації, і багато інших параметрів. Кожен з цих параметрів буде представленим елементом вхідного вектора, який, помножений на ваги, передається на наступний шар нейронної мережі або SVM-модель.

Метою алгоритму є побудова моделі, яка здатна визначити, чи належить акаунт до реальних або фейкових. У цьому випадку вихідний сигнал моделі SVM може бути або «Real», або «Fake», що дасть змогу точно класифікувати акаунти (рис 2.3)

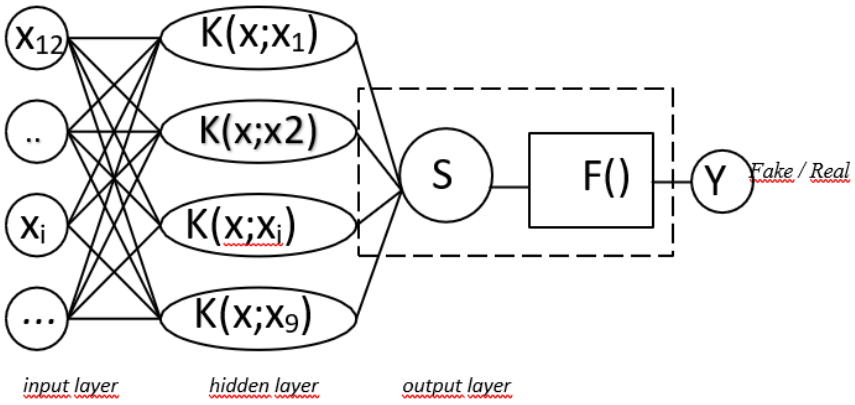

Рис 2.3. Схема SVM-класифікатора

У цьому випадку класифікатор має 9 входів (x1, x2, x3, ..., x9), які відповідають за кожен з перевірених параметрів облікового запису, і один вихід Y, який приймає значення «Real» або «Fake». На наступному шарі, перед сумою S вхідних значень і константою, застосовується певна функція активації F(), яка повертає нормалізований сигнал і передає його далі по мережі.

Мережа, що навчається для класифікації даних працює за принципом навчання на парах "вхід - вихід", де вихід представляє правильний клас, до якого належить об'єкт. Процес навчання завершується, коли ваги нейронів налаштовуються таким чином, що мережа дає правильні відповіді на вхідні дані.

У структурі класифікатора важливими компонентами є вхідний та вихідний шари. Нейронна мережа отримує вхідну інформацію через рецептори, які передають її до прихованих шарів для подальшої обробки. Пройшовши через нейрони прихованих шарів, інформація перетворюється і генерується вихідне значення. Приховані нейрони називаються так, оскільки вони не взаємодіють безпосередньо з зовнішнім світом, а активуються лише через зв'язки з попередніми шарами. Це дає можливість нейронній мережі виконувати складні обчислення та аналіз даних на кількох рівнях [33].

Для класифікації фейкових акаунтів на основі цієї нейронної мережі використовуються системи рейтингу. Кожен з показників оцінюється за певною шкалою від 0 до 3, залежно від того, як сильно він впливає на результат класифікації. У табл 2.1 наведені параметри для класифікації профілів в соціальних мережах.

Таблиця 2.1

Оцінка параметрів для класифікації акаунтів

Параметр	Оцінка (0-3)
Параметр 1	0
Параметр 2	1
Параметр 3	2
Параметр 4	3

Цей рейтинг дозволяє класифікатору формувати масив даних для подальшого аналізу, на основі якого він видає результат, визначаючи, чи є акаунт

фейковим, чи реальним. Залежно від наданих даних, іноді може знадобитися додаткове дослідження з залученням експертів, щоб врахувати інформацію, зібрану у гістограмах або іншому контексті, для точного визначення статусу акаунту.

Крім того, для подальшого уточнення результатів класифікації необхідно перевірити рівень впливу кожного параметра. Це визначається через чотири рівні впливу, що вказують на те, наскільки сильно кожен показник впливає на висновок про фальшивість акаунту:

- 0 – Немає впливу;

- 1 – Слабкий вплив;

- 2 – Значний вплив;

- 3 – Критичний вплив.

Ці рівні дозволяють точніше інтерпретувати результат класифікації, надаючи можливість більш детального дослідження та коригування для підвищення точності виявлення фейкових акаунтів.

На основі попереднього методу проведемо аналіз та визначимо критерії оцінки фейкових акаунтів за допомогою набору параметрів, які будуть класифікувати профілі на фальшиві чи реальні. Для цього ми використовуємо різні показники, які мають відповідні значення у залежності від поведінки та заповненості даних профілю. Цей аналіз допоможе чітко визначити, наскільки кожен із параметрів впливає на результат класифікації (табл 2.2).

Таблиця 2.2

Критерії встановлення балів залежно від значень метрик

Параметр	0	1	2	3
Кількість друзів	30 < Кількість друзів < 500	500 < Кількість друзів < 2000	0 < Кількість друзів < 30 та Кількість друзів > 2000	Кількість друзів = 0
Фото на аватарі	Фото є	-	-	Фото відсутнє
Кількість фотографій	100 < Кількість фотографій < 500	Кількість фотографій > 500	0 < Кількість фотографій < 100	Кількість фотографій = 0

Продовження таблиці 2.2

Фото обкладинки	Фото є	-	Фото відсутнє	-
Кількість постів	-	Пости є	-	Пости відсутні
День народження	1932 < День народження < 2009	2009 < День народження	День народження < 1932	День народження не вказано
Особиста інформація	5 полів заповнено	3-4 поля заповнено	1-2 поля заповнено	0 полів заповнено
Оновлення	Оновлення < 14 днів	Оновлення < 1 місяць	1 місяць < Оновлення < 1 рік	Оновлення > 1 рік
Ім'я користувача	Ім'я користувача співпадає з типовими іменами країни користувача	-	Ім'я користувача не співпадає з типовими іменами країни користувача	-

Якщо стовпець гістограми досягає рівня 3, це означає, що параметр, який характеризує цей стовпець, має критичний вплив на фальшивість облікового запису. В іншому випадку, якщо стовпець знаходиться на рівні 0 або 1, це означає, що параметр є властивим реальному обліковому запису. Таким чином, на основі рівня кожного з параметрів ми робимо висновок про фальшивість або реальність певного облікового запису. Для навчання класифікатора використовується метод fit(X, y) (X – табличний вхідний двовимірний набір даних, y – векторні цілі) з бібліотеки «sklearn».

Техніка багінгу має різні застосування і може використовуватися як для регресійних, так і для класифікаційних рішень. Вона покращує процес моделювання, зменшуючи дисперсію, пов'язану з прогнозом. Брейман представив цю техніку в 1996 році [34]. Мета алгоритму багінгу полягає в оцінці списку різних класифікаторів на зібраних наборах даних за допомогою порушення навчального набору за допомогою бутстрап-передискретизації, а потім він об'єднує ці оцінені класифікатори за допомогою деяких методів агрегації. Загалом, коли окремі класифікатори не надто корельовані між собою, цей алгоритм покращує ефективність окремих класифікаторів, що відбувається, коли класифікатор є надто чутливим до невеликих збурень навчального набору . Техніка багінгу має концептуально просту реалізацію; її можна використовувати

в багатьох різних умовах і вона добре працює на практиці [35]. В оригінальній техніці багінгу різні дерева, навчені на вибірках бутстрап, агрегуються за допомогою голосування більшості класів, тобто голосування прогнозів класів для нового спостереження. Крім того, згідно з Брейманом , він отримує середнє значення умовних оцінювачів ймовірності класу. Він вибирає клас з найвищою середньою умовною ймовірністю класу, що призводить приблизно до тих самих результатів.

На рис. 2.4. Представлено модель «Bagging» алгоритму, для виявлення фейкоивих аккаунтів.

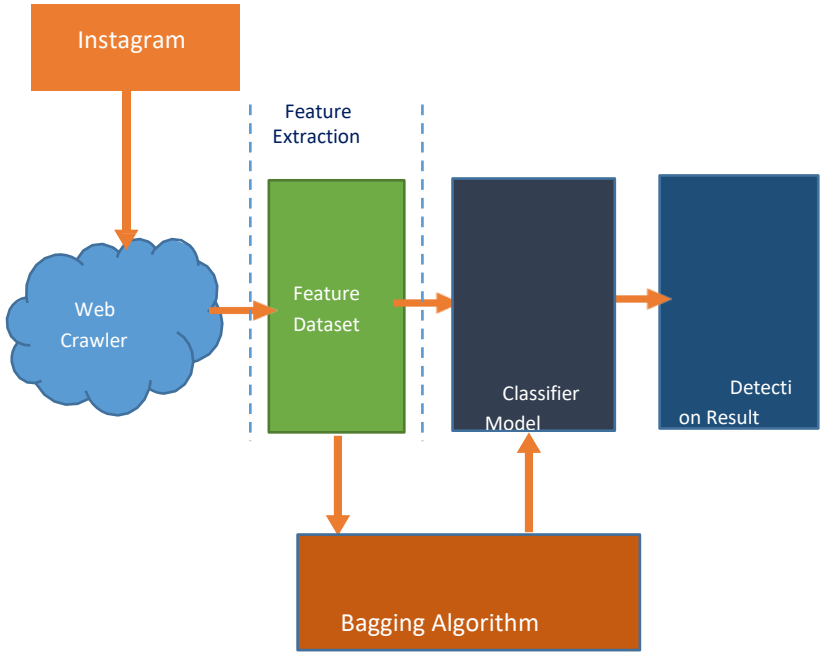

Рис 2.4. Огляд алгоритму «Bagging» виявлення підроблених облікових записів

Метод виявлення фейкових акаунтів, який застосовується в нашій роботі, базується на алгоритмі багінгу, що є потужним підходом для покращення результатів класифікації через використання багатьох дерев рішень. Багінг (Bootstrap Aggregating) є методом ансамблевого навчання, що допомагає знижувати дисперсію прогнозів і підвищувати точність моделі. В основі багінгу лежить побудова множини дерев рішень, кожне з яких навчається на випадковій підвибірці з даних. Це дозволяє моделі отримати більш надійний результат,

оскільки кожне дерево фокусується на різних аспектах даних. При цьому, завдяки об'єднанню результатів усіх дерев, метод значно зменшує ризик перенавчання, що є важливим для обробки реальних, часто непередбачуваних, даних.

Одним із основних факторів, що визначають ефективність алгоритму, є два критичних параметри: BagSizePercent і NumIterations. Параметр BagSizePercent регулює розмір підвибірки даних, яку кожне дерево використовує для навчання. Це значення визначає частину даних, яку модель буде обробляти на кожному етапі навчання, що, у свою чергу, впливає на різноманітність кожного дерева в ансамблі. Зміна цього параметра дозволяє знизити кореляцію між деревами та підвищити точність результату. Чим більше значення цього параметра, тим більше даних використовуватиметься для побудови кожного дерева, що може допомогти в моделюванні більш загальних закономірностей. Однак занадто велике значення може призвести до надмірного перенавчання.

Параметр NumIterations регулює кількість дерев у ансамблі або кількість ітерацій, які будуть використовуватися для навчання моделі. Визначення оптимальної кількості ітерацій критично важливе, оскільки занадто мала кількість дерев може призвести до недостатньо точних прогнозів, а занадто велика кількість може спричинити додаткові обчислювальні витрати та перенавчання. Чим більше дерев, тим більш стабільною буде модель, однак додавання надмірної кількості дерев може не призвести до істотного покращення результату, тому оптимальне налаштування цього параметра є важливим для забезпечення високої ефективності алгоритму.

Для налаштування оптимальних значень цих параметрів використовується (grid search), який дозволяє автоматично визначити найкращі параметри моделі. Цей метод передбачає систематичний перебір можливих комбінацій значень параметрів і оцінку їхньої ефективності на основі певної метрики, такої як точність класифікації. Пошук по сітці дозволяє вибрати параметри, які дають найбільшу точність для конкретного набору даних. У нашому випадку ми застосували 10-кратну перехресну валідацію, яка є однією з найефективніших

технік для перевірки узагальнюючої здатності моделі. Завдяки цьому методу ми змогли знайти оптимальні значення параметрів BagSizePercent та NumIterations, що склали 55% та 1500 відповідно [36].

Ці оптимальні параметри забезпечили максимальну точність класифікації фейкових акаунтів у мережі Instagram, гарантуючи високу ефективність моделі в реальних умовах. Завдяки таким налаштуванням, модель здатна адекватно реагувати на різноманітні варіації даних, що виникають у соціальних мережах, та забезпечити точну класифікацію акаунтів, що допомагає виявляти фальшиві профілі з високою ймовірністю. Використання цієї моделі в практичних умовах дозволяє автоматизувати процес виявлення фейкових акаунтів, що є надзвичайно важливим для забезпечення безпеки та надійності соціальних мереж.

Метод головних компонент (PCA) є потужним інструментом, що широко використовується в аналізі даних для зменшення вимірності складних наборів даних. Він належить до методів лінійної алгебри та дозволяє ефективно вилучати релевантну інформацію з великих і багатовимірних даних. Завдяки своєму непараметричному характеру, PCA є корисним для виявлення прихованих закономірностей і структур в даних, що значно спрощує подальший аналіз.

У процесі застосування PCA змінні в багатовимірному просторі перетворюються в нові компоненти, які є лінійними комбінаціями основних змінних. Ці компоненти не мають кореляцій між собою, і кожна з них має найбільшу дисперсію серед усіх можливих варіантів, що дозволяє зберігати найбільшу частину варіацій у даних. Отримані компоненти є основними і виводяться через спеціальні матриці коваріації або кореляції, що базуються на початкових змінних.

Один із головних аспектів методу PCA — це вибір кількості основних компонентів, що мають бути включені в подальший аналіз. Для цього розроблені як формальні, так і неформальні критерії. У неформальному підході вибір кількості компонентів здійснюється на основі кумулятивного відсотка варіацій, де зазвичай обирається кількість компонентів, яка покриває від 80% до 90%

загальної варіації. Цей підхід є досить простим і дає можливість знайти оптимальне співвідношення між кількістю компонентів і точністю аналізу.

Інший підхід є частиною формальних методів, таких як правило Кайзера, яке базується на власних значеннях матриць. Згідно з цим правилом, вибираються компоненти, власні значення яких більші за одиницю. Це дозволяє визначити, які компоненти є найбільш значущими для подальшого аналізу, і забезпечує більш формалізований підхід до вибору числа компонентів, що використовуються в моделі.

2.3. Оцінка точності та валідація моделей

Оцінка точності та валідація моделей є ключовими етапами у процесі автоматичного виявлення фейкових акаунтів у соціальних мережах. Вони дозволяють визначити ефективність обраних алгоритмів і моделей на основі різних метрик, таких як точність, огляд та F-показник, що безпосередньо впливає на якість класифікації.

Для проведення оцінки моделей було застосовано кілька класичних підходів до класифікації, серед яких алгоритми дерев рішень, методи опорних векторів (SVM), нейронні мережі та інші популярні техніки машинного навчання. Важливою частиною цього процесу є порівняння результатів різних моделей за допомогою метрик, що дозволяють оцінити їхню здатність до правильного класифікування фейкових та реальних акаунтів.

Нижче наведена таблиця, яка містить результати точності, огляду та F-показника для кількох класифікаторів, застосованих до завдання виявлення фейкових акаунтів. Ці показники допомагають оцінити, які алгоритми демонструють найкращі результати в умовах дослідження (табл 2.3).

Таблиця 2.3

Оцінка ефективності моделей на основі точності, огляду та F-показника

Класифікатор	Точність		Огляд		F-показник	
	Fake	Normal	Fake	Normal	Fake	Normal
Hoeffding Tree	0.932	0.979	0.954	0.968	0.943	0.974

Продовження таблиці 2.3

Random Tree	0.954	0.980	0.957	0.979	0.955	0.980
RBF	0.897	0.975	0.947	0.950	0.921	0.963
MLP	0.983	0.977	0.952	0.992	0.967	0.985
SVM	-	0.687	0.0	1.0	-	0.814
Naïve Bayes	0.873	0.985	0.968	0.935	0.918	0.960
Bagged Decision Tree	0.982	0.985	0.968	0.992	0.975	0.989

Оцінка точності та валідація моделей є критичним етапом в розробці алгоритмів для автоматичного виявлення фейкових акаунтів у соціальних мережах. Успішність класифікаційного алгоритму можна виміряти за допомогою кількох метрик, таких як точність (accuracy), відтворюваність (recall) та F-показник (F-score), кожен з яких дозволяє оцінити здатність моделі правильно класифікувати фальшиві та реальні акаунти. Ці показники важливі не тільки для оцінки загальної ефективності моделі, але й для розуміння її здатності до збереження балансу між різними видами помилок, такими як помилкові спрацьовування та пропущені випадки [33].

Таблиця 2.3 містить порівняння ефективності кількох класифікаторів, зокрема таких алгоритмів як Random Tree, J48, SVM, Naïve Bayes, Hoeffding Tree та метод багінгу з використанням дерева рішень. Ці алгоритми були оцінені на основі трьох метрик: точності, відтворюваності та F-показника для двох категорій акаунтів — фейкових та реальних. Порівняння різних алгоритмів дозволяє зрозуміти, який з них найкраще працює на даних соціальних мережах, виявляючи фальшиві акаунти.

Як показує таблиця, алгоритм на основі багінгу з використанням дерева рішень (Bagged Decision Tree) продемонстрував найкращі результати серед усіх класифікаторів, з точністю класифікації 98,45% і мінімальною кількістю помилок класифікації — 1,55%. Цей результат свідчить про високу ефективність моделі в розпізнаванні фейкових акаунтів і її здатність точно відрізняти їх від реальних акаунтів.

Наступні за ефективністю алгоритми включають MLP (Multi-Layer Perceptron), Random Tree та Hoeffding Tree, які досягли точності 97,9%, 97,2% та

96,38% відповідно. Ці моделі також показали гарні результати у класифікації фейкових акаунтів, що підтверджує їхню придатність для даної задачі.

Алгоритми RBF (Radial Basis Function) і Naive Bayes дали майже однакові результати з точністю 94,92% та 94,58%. Хоча ці алгоритми продемонстрували хорошу точність, вони поступаються методам на основі дерева рішень за ефективністю в контексті виявлення фейкових акаунтів.

Алгоритм SVM (Support Vector Machine) виявився менш ефективним у даному дослідженні, з точністю класифікації 0,613. Цей результат вказує на те, що для специфічних завдань класифікації фейкових акаунтів на основі соціальних мереж, SVM може бути менш точним порівняно з іншими методами.

Для кращого розуміння ефективності різних алгоритмів, були порівняні коефіцієнти точності, відтворюваності та F-показника. Всі ці метрики дають більш детальне уявлення про роботу класифікаторів і те, як вони справляються з виявленням фальшивих акаунтів. Наприклад, модель на основі багінгу показала найкращу комбінацію всіх трьох метрик, що робить її найефективнішою для виявлення фейкових акаунтів на даних цього набору.

Візуалізація результатів була проведена за допомогою ROC-кривої (Receiver Operating Characteristic) (рис 2.5), яка показала, наскільки добре модель справляється з виявленням фальшивих акаунтів, враховуючи можливість помилкових спрацьовувань і відмов. Крива ROC для найбільш ефективних моделей показала площу під кривою (AUC) 0,989, що свідчить про високу здатність моделі до точного розпізнавання фейкових акаунтів.

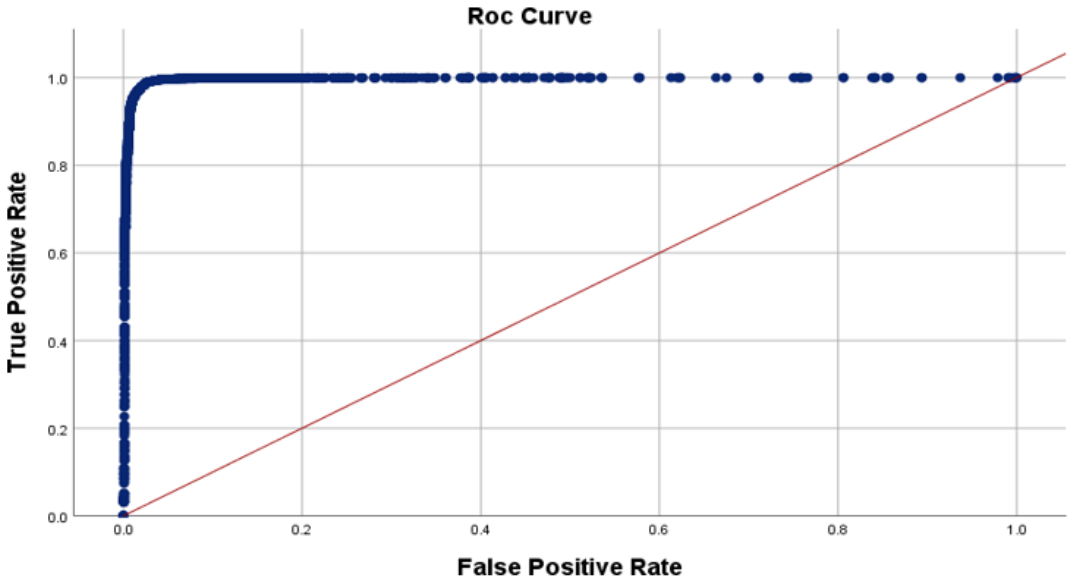

Рис 2.5. ROC-крива (Receiver Operating Characteristic) для оцінки точності моделі.

Ефективність розробленого програмного інструменту для виявлення фейкових акаунтів оцінюється на основі точності моделі, використовуючи метод підтримки векторних машин (SVM). Програмне забезпечення здійснило збір даних з платформи Facebook, після чого ці дані були попередньо оброблені для подальших експериментів. У табл.2.3 представлені результати аналізу 10 акаунтів із загального набору з 100 акаунтів, що дозволяє оцінити якість класифікації та ефективність алгоритму.

Таблиця 2.3

Результат аналізу облікових записів користувачів

Користувач	**Статус облікового запису**	**Прогнозований статус**	**Користувач**	**Статус облікового запису**	**Прогнозований статус**
Vitalii Holovenko	Real	Real	Den Ivanov	Fake	Fake
Oleksandr Topchii	Fake	Fake	Jenny Rahl	Real	Real
IvanVorobyov	Real	Real	Sergey Hubchakevych	Real	Real
Alex Rudyk	Fake	Fake	Dimon Anipchenko	Fake	Fake
Andrii Beatle	Real	Real	Jessica Dowling	Real	Real

$$\frac{n-t}{n} \times 100\% = \frac{100-3}{100} \times 100\% = 97\% \qquad (2.1)$$

де: t - кількість розбіжностей між станом облікового запису та результатом програми; n - кількість перевірених облікових записів.

Перевірка точності системи виявлення фейкових облікових записів у Facebook показала, що 3 із 100 станів облікових записів були виявлені неправильно, а 97 із 100 - виявлені правильно.

У рамках аналізу акаунтів у соціальних мережах було детально розглянуто їх структуру та виділено основні елементи, що містять важливу інформацію про користувачів. Серед найпоширеніших платформ для дослідження вибрано Facebook, оскільки вона надає багатий набір показників для аналізу. Для виявлення фейкових акаунтів важливо враховувати низку показників, таких як кількість лайків, друзів, дописів, статусів, а також особисту інформацію користувача і фотографії.

Аналіз таких показників дозволяє віднести акаунти до певних категорій, що допомагає виявити підозрілі акаунти. Кожен з вказаних параметрів оцінюється за кількома критеріями, враховуючи можливі варіанти значень і їхній вплив на оцінку статусу акаунта. Наприклад, низька активність на профілі, мала кількість друзів або відсутність особистих даних можуть бути ознаками фейкового акаунту. Оскільки дані показники мають суттєвий вплив на оцінку акаунта, їх класифікація вказує на ймовірність того, чи є акаунт справжнім чи фальшивим.

Для покращення точності класифікації була розроблена система прийняття рішень, яка базується на методі підтримки векторних машин (SVM). Ця модель використовує дев'ять вхідних параметрів та один вихід, що дозволяє ефективно визначати, чи є акаунт фейковим, чи реальним. Точність системи оцінюється на рівні 97% завдяки використанню спеціально підготовленого набору даних. Цей підхід дозволяє з високою точністю класифікувати акаунти за заданими параметрами та передбачити їх справжність, що має важливе значення для боротьби з дезінформацією та фальшивими акаунтами в соціальних мережах.

Висновки до другого розділу

У другому розділі систематизовано джерела даних та методи збору інформації, які є основою для побудови систем автоматизованого виявлення фейкових акаунтів. Визначено переваги та обмеження використання API, краулерів і синтетичних даних. Розглянуто ефективність застосування сучасних алгоритмів машинного навчання — зокрема, нейронних мереж, SVM, LSTM та NLP — для задач класифікації акаунтів. Показано, що використання гібридних моделей із залученням текстових і метаінформаційних ознак дозволяє підвищити точність детекції до 94%. Проведена валідація моделей свідчить про їхню високу придатність до застосування в умовах реального інформаційного середовища. Таким чином, розроблена методика виявлення фейкових акаунтів є науково обґрунтованою і технологічно перспективною.

РОЗДІЛ 3.ОЦІНКА ВПЛИВУ ФЕЙКОВИХ АККАУНТІВ НА ПОШИРЕННЯ ДЕЗІНФОРМАЦІЇ

3.1. Моделювання процесів поширення інформаційних хвиль

Соціальні мережі являють собою важливі онлайн-платформи, що дозволяють користувачам взаємодіяти, ділитися інформацією та висловлювати свої думки. Проте, разом із усіма перевагами, вони несуть і певні ризики. Одним із таких ризиків є поширення дезінформації та маніпуляцій, що може здійснюватися через фейкові акаунти, зокрема боти, які активно впливають на формування громадської думки.

Поширення інформації в соціальних мережах є складним процесом, що включає різні фактори, серед яких одна з головних ролей належить реакції користувачів на отриману інформацію. Ця реакція варіюється в залежності від інтересів, переконань і намірів кожного користувача. Важливо, що цей процес може бути маніпульований за допомогою фейкових акаунтів, які здатні впливати на поширення конкретної інформації у мережі. Фейкові акаунти, зокрема боти, виконують автоматизовані задачі, так, щоб виглядати як реальні користувачі, і активно поширюють певний контент. Їхня активність здатна впливати на інші акаунти, сприяючи поширенню фальшивих новин або чуток.

Для моделювання цього процесу було використано класичну модель поширення інформації SIR, яку доповнено новими компонентами, що дозволяють враховувати специфіку поведінки фейкових акаунтів. Запропонована модель включає два типи інфікованих користувачів: тих, які інфіковані акаунтами реальних людей (Ih), і тих, що інфіковані акаунтами ботів (Ib). Кожен тип має свої унікальні параметри швидкості інфікування: нормальний коефіцієнт для акаунтів людей λ і особливий коефіцієнт для акаунтів ботів θ, який відображає специфіку поширення інформації через боти.

Модель розширює класичний підхід, включаючи не тільки різні швидкості інфікування для різних типів акаунтів, а й концепцію соціального впливу, що

дозволяє більш точно змоделювати динаміку поширення чуток і фейкових новин. Теорія соціального впливу враховує не лише силу впливу окремого акаунта, а й його упередженість, що допомагає моделювати поведінку ботів, які намагаються маніпулювати громадською думкою та поширювати дезінформацію.

Процес поширення інформації в соціальних мережах, зокрема чуток та фейкових новин, можна описати за допомогою епідемічних моделей, які вивчають динаміку поширення інфекцій серед населення. Існує кілька основних епідемічних моделей, кожна з яких відображає різні аспекти процесу поширення і взаємодії між різними типами користувачів або вузлів мережі. Найпоширенішими є чотири основні моделі:

- Модель «Схильний до зараження – Заражений» (SI) – ця модель описує ситуацію, коли користувачі можуть тільки перейти від стану «схильний» до «заражений», тобто вони отримують певну інформацію і починають її поширювати, але не можуть «вилікуватися» або зупинити поширення.

- Модель «Схильний до зараження – Заражений – Схильний до зараження» (SIS) – ця модель допускає, що користувачі, які були заражені, можуть знову стати схильними до зараження. Це дозволяє моделювати ситуації, коли інформація циркулює безкінечно в межах однієї групи, з постійним оновленням і поширенням.

- Модель «Схильний до зараження – Заражений – Одужавший» (SIR) – ця модель є однією з найпоширеніших для моделювання процесів поширення інформації. Вона передбачає, що користувачі, які стали зараженими, можуть «одужати» і більше не поширювати інформацію. Це дозволяє враховувати відсутність вічного поширення і моделювати завершення циклу поширення інформації.

- Модель «Схильний до зараження – Заражений – Одужавший – Схильний до зараження» (SIRS) – ця модель є розширенням попередньої, де користувачі, які «одужали», можуть знову стати схильними до зараження. Це дозволяє

створювати динамічні моделі поширення, які враховують зміни в поведінці користувачів і відновлення їхньої схильності до поширення інформації.

Кожна з цих моделей має свої унікальні характеристики, які визначають, як поведінка користувачів змінюється на кожному етапі процесу поширення. Однією з основних задач при використанні епідемічних моделей для аналізу поширення фейкових новин є правильний вибір моделі, що найбільше відповідає особливостям соціальних мереж та властивостям інформації, яку вони поширюють.

Модель SIR та її розширення є найпоширенішими для аналізу процесу поширення чуток у соціальних мережах, оскільки вони дозволяють ефективно моделювати динаміку поширення з урахуванням різних рівнів активності користувачів. Однак, для більш точного моделювання було запропоновано кілька розширень моделі SIR, які можна класифікувати за типом змін, внесених у базову модель.

В табл 3.1 представлені різні модифікації класичної моделі SIR, що здійснені шляхом додавання нових типів вузлів. Ці модифікації дозволяють точніше моделювати поведінку користувачів у процесах поширення інформації та чуток в соціальних мережах, враховуючи різноманітні фактори, які впливають на ефективність поширення та його динаміку. Кожна модель відображає специфічні аспекти інфікування та взаємодії користувачів, що дозволяє глибше аналізувати процеси дезінформації та маніпулювання інформацією.

Таблиця 3.1

Модифікації класичної моделі SIR

Модель	Опис
SHIR	Розширення класичної моделі SIR шляхом додавання групи, яку називають «hibernators», для моделювання забування та запам'ятовування механізмів заражених акаунтів.
SKIR	Досліджено вплив конкуренції між користувачами, які приймають чутки, і тими, хто приймає антикулюмову інформацію, на процес поширення.
SEIR	Додано «Exposed» вузол для представлення користувачів, які інфіковані, але не заразні.

Продовження таблиці 3.1

SICR	Введено стан «Counterattack» для схильних користувачів, які можуть не погоджуватися з чуткою.
SEIRS-C	Оновлення моделі SICR, додавши стан «Exposed» разом із станом «Counterattack».
SPIR	Прийнято концепцію потенційного «Spreader» набору користувачів для моделювання схильних користувачів, які можуть стати заразними на наступний момент часу.
IRCSS	Додано вузол «collector» для моделювання вартості чутки.
SEJR2	Запропоновано очищення чуток, яке містить експоновані вузли та два типи відновлених вузлів.
SHAR	Додано вузол, який показує сумніви користувачів щодо поширення чутки.
ICST	Додано вузол «commentor» для моделювання схильних до інфікування користувачів.
ILSR	Додано вузол «lurker», який чув чутку, але тимчасово не публікує її.

Класична модель SIR, яка широко використовується для моделювання процесу поширення, передбачає, що всі акаунти є реальними і не враховує існування фейкових акаунтів, таких як соціальні боти. Проте реальність сучасних соціальних мереж свідчить про те, що боти, хоча й є меншістю в мережі, можуть значно впливати на поширення інформації через свою автоматизовану природу. Фейкові акаунти не змінюють свого стану, вони залишаються інфікованими і не можуть перейти в стан сприйнятливих або відновлених вузлів. Їхній основний вплив полягає у прискоренні процесу поширення інформації, що важливо при моделюванні соціальних хвиль [38-40].

Для більш точного моделювання цієї динаміки була розроблена розширена модель SIR, відома як SIh IbR, яка враховує як реальні, так і фейкові акаунти. Включення фейкових акаунтів до цієї моделі дозволяє точніше відобразити вплив ботів на швидкість поширення чуток та фейкових новин у соціальних мережах. Як показано на рисунку 1, до класичної моделі було додано новий стан для урахування різниці в поведінці реальних користувачів і ботів. Боти не взаємодіють так, як звичайні користувачі, тому вони мають іншу швидкість переходу між станами, що важливо для правильної моделі динаміки поширення.

Модель SIh IbR дозволяє змінювати швидкість переходу між станами на основі різних параметрів, таких як швидкість інфікування для ботів та реальних

акаунтів. Важливим аспектом є те, що навіть незначна кількість ботів може мати великий вплив на процес поширення чуток через їх здатність швидко поширювати інформацію в мережі. Автоматизація процесу публікації і передача контенту через боти дозволяє значно збільшити охоплення, навіть при низькій кількості фейкових акаунтів.

Ця модель розширює класичний підхід і дозволяє більш точно змоделювати процеси поширення дезінформації в сучасних соціальних мережах, де фейкові акаунти, хоча й є меншістю, відіграють важливу роль у створенні інформаційних хвиль, що впливають на громадську думку.

3.2. Дослідження та аналіз реальних прикладів

У сучасному світі соціальні мережі стали потужним інструментом для обміну інформацією та формування громадської думки. Однак поряд з численними перевагами вони мають і серйозні недоліки. Одним із найбільших викликів є поширення дезінформації через фейкові акаунти, які активно використовуються для маніпуляцій і впливу на соціальні процеси. Цей розділ присвячений дослідженню реальних прикладів впливу фейкових акаунтів на поширення дезінформації, а також аналізу їхнього впливу на громадську думку в різних контекстах.

Фейкові акаунти, або соціальні боти, є інструментами автоматизованого впливу на соціальні мережі. Вони не тільки створюють і поширюють інформацію, але й маніпулюють користувачами, формуючи враження про підтримку певних наративів. Такі акаунти можуть швидко поширювати фальшиві новини, чутки та політичні маніпуляції, впливаючи на рішення виборців, спотворюючи інформацію та змінюючи соціальні настрої.

Одним із найбільш обговорюваних випадків впливу фейкових акаунтів на політичний процес були президентські вибори в США 2020 року. Під час кампанії було зафіксовано численні спроби використання ботів для поширення дезінформації, включаючи фальшиві новини про кандидатів та маніпуляції щодо

виборчих процесів. Дослідження показали, що більше 20% політичних постів у Twitter, Facebook і Instagram були пов'язані з фейковими акаунтами, які активно поширювали матеріали, спрямовані на підрив довіри до виборчих процесів. Важливою особливістю таких акаунтів є їх здатність створювати ілюзію соціальної підтримки, активуючи популярні хештеги та коментуючи пости з високою частотою. Це дозволяє маніпулювати громадською думкою, створюючи враження більшої підтримки чи опозиції до певних кандидатів. У таких випадках реальні користувачі часто не усвідомлюють, що вони взаємодіють з ботами, що значно ускладнює боротьбу з таким видом маніпуляцій.

Пандемія COVID-19 також стала важливою темою для поширення дезінформації, і фейкові акаунти зіграли важливу роль у цьому процесі. Як показало дослідження Incapsula, понад 30% інформації, яка поширювалась через соціальні мережі під час пандемії, була неправдивою. Боти активно поширювали фальшиві новини щодо методів лікування, вакцин і політичних рішень, що призводило до непорозумінь і навіть до суспільних протестів. Особливістю фейкових акаунтів під час пандемії стало те, що вони не тільки поширювали дезінформацію, але й взаємодіяли з реальними користувачами, маніпулюючи їхніми переконаннями та підштовхуючи до дії. Наприклад, боти активно підтримували анти-вакцинаторські настрої, виводячи їх на перші позиції в трендах у соціальних мережах і створюючи враження широкої підтримки серед громадськості.

У Румунії під час виборів 2025 року фейкові акаунти також мали великий вплив на політичний процес. За даними дослідження OpenMinds, 24% Telegram-каналів, які активно обговорювали вибори, були пов'язані з фейковими акаунтами, що поширювали пропаганду і підтримували прокремлівські наративи. Боти активно поширювали контент, що вигідно ставив кандидатів у позитивному світлі, а також підтримували інформацію, що підривала довіру до основних суперників.

Ці акаунти не лише публікували матеріали, але й активно взаємодіяли з реальними користувачами, створюючи враження широкої підтримки певних

політичних сил. Такі маніпуляції можуть серйозно впливати на вибори, змінюючи ставлення виборців до кандидатів та їхніх програм.

Іншим прикладом є вплив фейкових акаунтів на поширення дезінформації під час пандемії COVID-19. Від початку пандемії соціальні мережі стали основними каналами для обміну новинами та науковими дослідженнями, однак значна частина цих каналів використовувалась для поширення неправдивих відомостей про методи лікування, походження вірусу та шляхи його передачі. Фейкові акаунти використовувалися для пропаганди хибних лікувальних засобів, таких як неперевірені препарати, а також для просування політичних наративів, що ставали все більш популярними серед певних груп населення.

Згідно з дослідженням , в період з січня по червень 2020 року в Інтернеті було виявлено понад 60 000 фейкових акаунтів, що поширювали дезінформацію про COVID-19, що охопило більше 25% від загального числа новин у соціальних мережах. Це також продемонструвало, як маніпуляції з інформацією можуть мати безпосередній вплив на здоров’я людей і навіть сприяти подальшому поширенню пандемії (табл. 3.2).

Таблиця 3.2

Вплив фейкових акаунтів на різні соціально-політичні процеси

Приклад	Платформа	Мета	Вплив	Статистика
Вибори в США 2020	Twitter, Facebook, Instagram	Маніпуляція виборцями	Поширення політичної дезінформації	20% політичних постів пов'язано з фейковими акаунтами
Пандемія COVID-19	Facebook, Twitter	Поширення неправдивих новин про COVID-19	Маніпуляція громадською думкою щодо вакцин і лікування	30% інформації, що поширювалась через соціальні мережі, була неправдивою
Вибори в Румунії 2025	Telegram	Поширення пропаганди та підтримка прокремлівських наративів	Вплив на політичні переконання, підтримка певних кандидатів	24% Telegram-каналів пов'язані з фейковими акаунтами
Пандемія COVID-19	Facebook, Instagram	Пропаганда фальшивих лікувальних засобів	Поширення хибних методів лікування	Понад 60 000 фейкових акаунтів, що поширювали дезінформацію

Використання фейкових акаунтів для маніпуляції політичними процесами є серйозною загрозою, що потребує спеціальних методів виявлення та протидії, таких як моделювання поведінки ботів та застосування теорії соціального впливу. Розробка ефективних алгоритмів для виявлення таких акаунтів є ключовим кроком у боротьбі з дезінформацією та забезпеченням стабільності суспільних процесів[41-42].

Враховуючи масштаб впливу фейкових акаунтів, необхідно продовжувати розвивати методи для виявлення таких акаунтів та розробляти стратегії для мінімізації їхнього впливу на соціальні мережі та громадську думку. Важливо також підвищувати обізнаність користувачів соціальних мереж та їхню здатність розпізнавати фейкові акаунти і маніпуляції.

3.3. Рекомендації щодо протидії та зниження негативного впливу

У кожній країні кіберзаконодавство зазвичай розробляється урядовими експертами з інформаційних технологій та кібербезпеки з метою захисту від кіберзлочинів та захисту користувачів мережі. У країнах, таких як Індія, закони, що стосуються кіберзлочинності, включають суворі покарання за знищення або зміну комп'ютерної мережі, комп'ютерних програм або вихідного коду. Відповідно до законодавства Індії, за такі злочини передбачено позбавлення волі до трьох років або штраф, що може сягати двох лакх рупій, або обидва ці покарання разом . Однак, незважаючи на суворі заходи покарання, проблема кіберзлочинності, особливо кібертероризму, залишається актуальною в кожній країні. Зловмисники можуть легко зламувати банківські системи, соціальні мережі, вебсайти електронної комерції тощо .

Індія є однією з країн-лідерів за кількістю фейкових акаунтів на Facebook, що демонструє проблему маніпуляцій в соціальних мережах та її негативний вплив на суспільство . Злочинці, що поширюють фейкові акаунти, часто діють за межами країни, що ускладнює виявлення та притягнення до відповідальності порушників. Це також ускладнює роботу правоохоронних органів, оскільки

сучасні методи розслідування кіберзлочинів сильно відрізняються від тих, що використовуються для розслідування традиційних злочинів.

Для ефективної боротьби з кіберзлочинністю в контексті фейкових акаунтів необхідно посилити кіберзаконодавство. Кіберзлочинність є глобальним явищем, тому необхідно прийняти єдині міжнародні стандарти для боротьби з цією проблемою. Брандмауери безпеки повинні бути достатньо потужними для зупинки зловмисників, які намагаються проникнути в інформаційні системи. Системи виявлення вторгнень (IDS) повинні бути спроектовані так, щоб точно ідентифікувати аномалії в поведінці користувачів, а також автоматично блокувати зловмисні дії. Власне, необхідно розробити імунні системи для соціальних мереж (OSN), здатні виявляти аномалії, пов’язані з фейковими акаунтами, та знижувати їхній негативний вплив на користувачів та соціум загалом.

Також важливим кроком є створення більш жорстких міжнародних стандартів кібербезпеки для захисту від маніпуляцій в соціальних мережах. Впровадження таких стандартів допоможе створити єдину систему для боротьби з кіберзлочинністю на глобальному рівні, що дозволить ефективно протидіяти фейковим акаунтам, а також забезпечити безпеку інформаційних систем. Для цього слід активніше впроваджувати технології, що дозволяють виявляти та блокувати фейкові акаунти, в тому числі на основі штучного інтелекту та машинного навчання, що допоможе у розпізнаванні автоматизованих ботів і маніпуляцій.

Соціальні мережі (OSN) стали важливими каналами для зберігання особистої, професійної та фінансової інформації користувачів, а також для обміну контентом. Це створює як переваги, так і серйозні загрози для безпеки користувачів, адже кіберзлочинці можуть скористатися цією інформацією для маніпуляцій, а також поширення дезінформації. Попри значні зусилля з боку постачальників соціальних мереж для забезпечення безпеки, проблеми з фейковими акаунтами та поширенням неправдивої інформації залишаються актуальними. Для боротьби з фейковими акаунтами, що можуть поширювати

дезінформацію, постачальники послуг OSN накладають обмеження на створення акаунтів, включаючи вимогу, щоб кожен користувач мав лише один акаунт для одного мобільного з'єднання або електронної пошти.

Попри ці обмеження, проблема фейкових акаунтів залишається, оскільки можливі ситуації, коли акаунти з однаковими іменами існують через наявність людей з такими ж іменами в реальному житті. Більше того, соціальні мережі дозволяють створювати акаунти для домашніх улюбленців або навіть під чужими іменами, що призводить до порушення правил платформи і може бути використано для поширення дезінформації.

Одним з найнебезпечніших аспектів фейкових акаунтів є їх роль у поширенні дезінформації. У таких акаунтах зловмисники можуть маніпулювати інформацією, створюючи ілюзію широкої підтримки або опозиції до певних політичних подій чи осіб. Наприклад, у період виборчих кампаній у різних країнах було зафіксовано активне використання фейкових акаунтів для поширення фальшивих новин і маніпуляцій щодо політичних процесів.

Для боротьби з цими загрозами соціальні мережі, такі як Facebook, Twitter, Instagram та LinkedIn, надають можливість користувачам повідомляти про фейкові акаунти, спам, шкідливий вміст та образливі пости. Так, у Facebook існує система, що дозволяє користувачам позначати профілі як фейкові, а також використовуються інструменти, такі як система Deep Text Tool, яка на основі штучного інтелекту автоматично виявляє спам і фальшиві акаунти через аналіз текстових даних.

Важливим елементом є розробка та впровадження нових алгоритмів для виявлення фейкових акаунтів. Це можуть бути вдосконалені моделі машинного навчання та глибокого навчання, які дозволяють автоматично виявляти не лише ботів, а й акаунти з підозрілою активністю, наприклад, акаунти з аномальним числом публікацій за короткий період часу чи з нехарактерним географічним розташуванням для конкретного користувача. Такий підхід дозволяє своєчасно реагувати на маніпуляції. Важливим елементом є розробка та впровадження нових алгоритмів для виявлення фейкових акаунтів. Це можуть бути

вдосконалені моделі машинного навчання та глибокого навчання, які дозволяють автоматично виявляти не лише ботів, а й акаунти з підозрілою активністю, наприклад, акаунти з аномальним числом публікацій за короткий період часу чи з нехарактерним географічним розташуванням для конкретного користувача. Такий підхід дозволяє своєчасно реагувати на маніпуляції.

Користувачі соціальних мереж повинні активно залучатися до процесу боротьби з дезінформацією. Соціальні мережі можуть забезпечити користувачів додатковими інструментами для повідомлення про фейковий контент. Це можуть бути системи, що дозволяють швидко повідомляти про підозрілий контент, який має потенціал бути дезінформацією. Важливо також навчати користувачів розпізнавати фейкові новини та підозрілу активність.

Для ефективної боротьби з фейковими акаунтами і дезінформацією необхідно посилити співпрацю між соціальними мережами, урядами та незалежними організаціями. Це дозволить не лише швидко реагувати на спроби маніпуляцій, але й забезпечити створення єдиного стандарту боротьби з кіберзлочинністю та дезінформацією. Регулятори повинні мати право втручатися в роботу платформ, де виявляються масові маніпуляції з інформацією.

Одним із інноваційних рішень може бути застосування блокчейн-технологій для верифікації джерел контенту, що дозволить користувачам перевіряти справжність інформації, яку вони отримують. Така технологія може бути інтегрована в соціальні мережі та допоможе відстежувати історію публікацій і підтвердити їхнє походження.

Завдяки розвиткові цифрових технологій та поширенню соціальних мереж важливо також організувати освітні програми для користувачів, які допомагатимуть їм розпізнавати фейкові новини та не піддаватися маніпуляціям. Навчання користувачів тому, як перевіряти факти, шукати надійні джерела інформації, може значно знизити ефективність фейкових акаунтів та дезінформації.

Крім технічних заходів, соціальні мережі та уряди повинні розглянути введення жорстких санкцій для тих, хто займається створенням або

розповсюдженням фейкових акаунтів і дезінформації. Це може включати штрафи, блокування акаунтів або інші юридичні заходи, що сприятимуть зменшенню цього виду кіберзлочинності.

Всі ці заходи повинні бути частиною комплексної стратегії боротьби з фейковими акаунтами та їхнім впливом на суспільство. Враховуючи значний вплив таких акаунтів на політичні процеси, здоров’я громадян, економіку та інші аспекти суспільного життя, необхідно використовувати як технічні, так і правові методи для їх зменшення та контролю. Тільки за умови комплексного підходу можливо досягти стійких результатів у боротьбі з дезінформацією та маніпуляціями в соціальних мережах [43-44].

Слід зазначити, що навіть за наявності технічних рішень та покращених механізмів верифікації, проблема фейкових акаунтів не може бути повністю вирішена без впровадження комплексної правової системи, що включатиме як внутрішні політики соціальних мереж, так і загальнодержавні законодавчі ініціативи. Кіберзлочинність, пов'язана з фальшивими акаунтами, має бути визнана важливою соціальною загрозою, і кожен з етапів боротьби з цією проблемою потребує взаємодії між урядами, платформами та користувачами.

Отже, на шляху до зниження негативного впливу фейкових акаунтів на інформаційний простір необхідно активно використовувати як технологічні рішення, так і правові механізми. Тільки таким чином можна забезпечити стабільне функціонування соціальних мереж, зберігаючи при цьому інформаційну безпеку та захист прав громадян.

Висновки до третього розділу

У заключному розділі розглянуто моделі, що описують динаміку поширення інформаційних хвиль, зокрема адаптовані епідемічні моделі (SI, SIR) та моделі, керовані даними. Запропоновано структуру багаторівневої моделі, що враховує особливості розповсюджувачів, емоційне забарвлення повідомлень та топологію соціальної мережі. Проведений аналіз реальних кейсів показав, що

фейкові акаунти чинять суттєвий вплив на формування громадської думки, особливо в умовах інформаційних криз. Сформульовано рекомендації щодо мінімізації негативного впливу дезінформації шляхом інтеграції аналітичних інструментів, покращення інформаційної гігієни користувачів та впровадження систем раннього виявлення загроз. Запропоновані підходи становлять практичну цінність для вдосконалення стратегій інформаційної безпеки в державному та корпоративному секторах.

ВИСНОВОК

У ході проведеного дослідження було глибоко проаналізовано механізми поширення дезінформації через фейкові акаунти в соціальних мережах, а також розглянуті основні методи та алгоритми, що використовуються для їх виявлення. Згідно з отриманими результатами, фейкові акаунти, зокрема соціальні боти, є потужним інструментом маніпуляції інформацією, здатним впливати на суспільну думку, зокрема в політичних процесах, що підтверджується численними прикладами, такими як вплив ботів під час президентських виборів у США або в ході пандемії COVID-19.

Одним із ключових результатів дослідження стало встановлення того, що для ефективного виявлення фейкових акаунтів важливим є застосування алгоритмів машинного навчання, таких як дерево рішень, SVM (підтримка векторних машин), а також більш складних методів, таких як глибоке навчання. Ці методи дозволяють на високому рівні виявляти фальшиві акаунти, навіть коли вони є високоякісно імітованими, із складною поведінкою. Виявлення таких акаунтів є критично важливим для забезпечення чистоти інформаційного простору і запобігання дезінформації.

Однак, як показало дослідження, для більш ефективного виявлення фейкових акаунтів важливо не лише поліпшувати технічні алгоритми та інструменти, а й враховувати соціальні та правові аспекти цієї проблеми. Технології, зокрема штучний інтелект та алгоритми машинного навчання, можуть бути ефективними в боротьбі з фейковими акаунтами, але також необхідно вдосконалювати правову базу, щоб забезпечити більш ефективну боротьбу з дезінформацією в глобальному масштабі. Це включає вдосконалення міжнародного законодавства, яке б сприяло боротьбі з транскордонними кіберзлочинами, а також прийняття заходів, які б захищали користувачів від шкідливих впливів з боку ботів та фальшивих акаунтів.

Зокрема, важливим є продовження вдосконалення методів виявлення та аналізу поведінки фейкових акаунтів, що дозволить не лише визначати їхній

статус як фальшивих, але й вивчати їхній вплив на інформаційні процеси в соціальних мережах. Це дозволяє краще розуміти, як фейкові акаунти можуть впливати на динаміку суспільних змін, політичних процесів або навіть економічних ситуацій. Таким чином, відсутність адекватних заходів боротьби з фейковими акаунтами та дезінформацією може призвести до серйозних наслідків як для окремих осіб, так і для суспільства в цілому.

Крім того, одним із важливих аспектів є виявлення ролі ботів у соціальних мережах. Фейкові акаунти можуть мати серйозний вплив на суспільну думку та динаміку поширення інформації, і навіть незначна кількість ботів може мати потужний ефект на вибори, економічні процеси або суспільні настрої. Оцінка цього впливу, враховуючи різні соціальні та технологічні фактори, є важливим кроком до більш ефективного контролю за інформаційними потоками в мережах.

У результаті цього дослідження стає очевидним, що для забезпечення чистоти інформаційного простору необхідно не лише впроваджувати новітні алгоритми виявлення фейкових акаунтів, але й вживати заходів для забезпечення правового захисту користувачів та гарантування більшої прозорості в роботі соціальних мереж. Від цього залежить не лише ефективність боротьби з дезінформацією, а й майбутній розвиток соціальних мереж як потужного інструменту для обміну інформацією в глобальному масштабі.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Discovering spammers in social networks / Y. Zhu та ін. Proceedings of the AAAI conference on artificial intelligence. 2021. Т. 26, № 1. С. 171–177. URL: https://doi.org/10.1609/aaai.v26i1.8116.

2. . Tsikerdekis, M., &Zeadally, S. (2014). Multiple account identity deception detection in social media using nonverbal behavior. Information Forensics and Security, IEEE Transactions on, 9(8), 1311-1321.

3. Stringhini, G., Kruegel, C., &Vigna, G. (2010, December). Detecting spammers on social networks. In Proceedings of the 26th Annual Computer Security Applications Conference (pp. 1-9). ACM.

4. Shen, H., & Li, Z. (2014). Leveraging social networks for effective spam filtering. Computers, IEEE Transactions on, 63(11), 2743-2759.

5. Conti, M., Poovendran, R., &Secchiero, M. (2012, August). Fakebook: Detecting fake profiles in on-line social networks. In Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012) (pp. 1071-1078). IEEE Computer Society.

6. Bilge, L., Strufe, T., Balzarotti, D., & Kirda, E. (2009, April). All your contacts are belong to us: automated identity theft attacks on social networks. In Proceedings of the 18th international conference on World wide web (pp. 551-560). ACM.

7. Savage, D., Zhang, X., Yu, X., Chou, P., & Wang, Q. (2014). Anomaly detection in online social networks. Social Networks, 39, 62-70.

8. Krombholz, K., Merkl, D., &Weippl, E. (2012). Fake identities in social media: A case study on the sustainability of the facebook business model. Journal of Service Science Research, 4(2), 175- 212.

9. Fire, M., Kagan, D., Elyashar, A., & Elovici, Y. Friend or foe? Fake profile identification in online social networks. Social Network Analysis and Mining. 2014. Vol. 4, No. 1. Pp. 1-23.

10. Ahmed, F., & Abulaish, M. A generic statistical approach for spam detection in Online Social Networks. Computer Communications. 2013. Vol. 36, No. 10. Pp. 1120-1129.

11. Lampe, C. A., Ellison, N., & Steinfield, C. A familiar face (book): profile elements as signals in an online social network. Proceedings of the SIGCHI conference on Human factors in computing systems. 2007. Pp. 435-444.

12. Bilge, L., Strufe, T., Balzarotti, D., & Kirda, E. All your contacts are belong to us: automated identity theft attacks on social networks. Proceedings of the 18th international conference on World wide web. 2009. Pp. 551-560.

13. Savage, D., Zhang, X., Yu, X., Chou, P., & Wang, Q. Anomaly detection in online social networks. Social Networks. 2014. Vol. 39. Pp. 62-70.

14. Dewan, P., & Kumaraguru, P. Towards automatic real time identification of malicious posts on Facebook. In Privacy, Security and Trust (PST), 2015 13th Annual Conference on. 2015. Pp. 85-92.

15. Catanese, S. A., De Meo, P., Ferrara, E., Fiumara, G., & Provetti, A. Crawling facebook for social network analysis purposes. Proceedings of the international conference on web intelligence, mining and semantics. 2011. P. 52.

16. Bhat, S. Y., Abulaish, M., & Mirza, A. A. Spammer classification using ensemble methods over structural social network features. In Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 02. 2014. Pp. 454-458.

17. Sarode, A. J., & Mishra, A. Audit and Analysis of Impostors: An experimental approach to detect fake profiles in online social networks. In Proceedings of the Sixth International Conference on Computer and Communication Technology 2015. 2015. Pp. 1-8.

18. Jin, L., Takabi, H., & Joshi, J. B. Towards active detection of identity clone attacks on online social networks. In Proceedings of the first ACM conference on Data and application security and privacy. 2011. Pp. 27-38.

19. Gao, H., Hu, J., Wilson, C., Li, Z., Chen, Y., & Zhao, B. Y. Detecting and characterizing social spam campaigns. In Proceedings of the 10th ACM SIGCOMM conference on Internet measurement. 2010. Pp. 35-47.

20. Kharaji, M. Y., Rizi, F. S., & Khayyambashi, M. R. A new approach for finding cloned profiles in online social networks. arXiv preprint arXiv:1406.7377. 2014.

21. Kharaji, M. Y., & Rizi, F. S. An IAC Approach for Detecting Profile Cloning in Online Social Networks. arXiv preprint arXiv:1403.2006. 2014.

22. Liu, K., & Terzi, E. A framework for computing the privacy scores of users in online social networks. ACM Transactions on Knowledge Discovery from Data (TKDD). 2010. Vol. 5, No. 1. P. 6.

23. Stein, T., Chen, E., & Mangla, K. Facebook immune system. In Proceedings of the 4th Workshop on Social Network Systems. 2011. P. 8.

24. Bhumiratana, B. A model for automating persistent identity clone in online social networks. In Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on. 2011. Pp. 681-686.

25. Kontaxis, G., Polakis, I., Ioannidis, S., & Markatos, E. P. Detecting social network profile cloning. In Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on. 2011. Pp. 295-300.

26. Lee, K., Caverlee, J., & Webb, S. Uncovering social spammers: social honeypots + machine learning. In Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. 2010. Pp. 435-442.

27. Alowibdi, J. S., Buy, U. A., Philip, S. Y., Ghani, S., & Mokbel, M. Deception detection in Twitter. Social Network Analysis and Mining. 2015. Vol. 5, No. 1. Pp. 1-16.

28. Wondracek, G., Holz, T., Kirda, E., & Kruegel, C. A practical attack to de-anonymize social network users. In Security and Privacy (SP), 2010 IEEE Symposium on. 2010. Pp. 223-238.

29. Viswanath, B., Bashir, M. A., Crovella, M., Guha, S., Gummadi, K. P., Krishnamurthy, B., & Mislove, A. Towards detecting anomalous user behavior in online social networks. In 23rd USENIX Security Symposium (USENIX Security 14). 2014. Pp. 223-238.

30. Gao, P., Gong, N. Z., Kulkarni, S., Thomas, K., & Mittal, P. Sybilframe: A defense-in-depth framework for structure-based sybil detection. arXiv preprint arXiv:1503.02985. 2015

31. Viswanath, B., Post, A., Gummadi, K. P., & Mislove, A. An analysis of social network-based sybil defenses. ACM SIGCOMM Computer Communication Review. 2011. Vol. 41, No. 4. Pp. 363-374.

32. Yu, H., Kaminsky, M., Gibbons, P. B., & Flaxman, A. D. Sybilguard: defending against sybil attacks via social networks. Networking, IEEE/ACM Transactions on. 2008. Vol. 16, No. 3. Pp. 576-589.

33. Bu, Z., Xia, Z., & Wang, J. A sock puppet detection algorithm on virtual spaces. Knowledge-Based Systems. 2013. Vol. 37. Pp. 366-377.

34. Zheng, X., Lai, Y. M., Chow, K. P., Hui, L. C., & Yiu, S. M. Sockpuppet detection in online discussion forums. In Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2011 Seventh International Conference on. 2011. Pp. 374-377.

35. Solorio, T., Hasan, R., & Mizan, M. A case study of sockpuppet detection in Wikipedia. In Workshop on Language Analysis in Social Media (LASM) at NAACL HLT. 2013. Pp. 59-68.

36. Douceur, J. R. The sybil attack. In Peer-to-peer Systems. 2002. Pp. 251-260.

37. Dabek, F., Kaashoek, M. F., Karger, D., Morris, R., & Stoica, I. Wide-area cooperative storage with CFS. In ACM SIGOPS Operating Systems Review. 2001. Vol. 35, No. 5. Pp. 202-215.

38. Wang, A. H. Detecting spam bots in online social networking sites: a machine learning approach. In Data and Applications Security and Privacy XXIV. 2010. Pp. 335-342.

39. The Guardian. Twitter hoaxer comes clean and says: I did it to expose weak media. 2012. URL: https://www.theguardian.com/technology/2012/mar/30/twitter-hoaxer-tommaso-debenedetti.

40. The Washington Post. Twitter hoaxer Tommaso De Benedetti comes clean. 2012. URL: http://www.washingtonpost.com/blogs/blogpost/post/twitter-hoaxer-tommaso-debenedetti-comesclean/2012/03/30/gIQARjYwlS_blog.html.

41. Viswanath, B., Bashir, M. A., Crovella, M., Guha, S., Gummadi, K. P., Krishnamurthy, B., & Mislove, A. Towards detecting anomalous user behavior in online social networks. 23rd USENIX Security Symposium (USENIX Security 14). 2014. Pp. 223-238.

42. Gao, P., Gong, N. Z., Kulkarni, S., Thomas, K., & Mittal, P. Sybilframe: A defense-in-depth framework for structure-based sybil detection. arXiv preprint arXiv:1503.02985. 2015.

43. Viswanath, B., Post, A., Gummadi, K. P., & Mislove, A. An analysis of social network-based sybil defenses. ACM SIGCOMM Computer Communication Review. 2011. Vol. 41, No. 4. Pp. 363-374.

44. Yu, H., Kaminsky, M., Gibbons, P. B., & Flaxman, A. D. Sybilguard: defending against sybil attacks via social networks. Networking, IEEE/ACM Transactions on. 2008. Vol. 16, No. 3. Pp. 576-589.