Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет комп'ютерних наук
(повна назва)

Кафедра програмної інженерії
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження моделей та інформаційних технологій виявлення фейкових новин
(тема)

Виконав:
здобувач 2 року навчання
групи ІПЗм-23-1

Артем ШАГУН
(Власне ім'я, ПРІЗВИЩЕ)

Спеціальність 121 – Інженерія програмного забезпечення
(код і повна назва спеціальності)
Тип програми освітньо-наукова

Керівник проф. Смеляков С.В.
(посада, Власне ім'я, ПРІЗВИЩЕ)

Допускається до захисту
Зав. кафедри ___ __ __ ____ Кирило СМЕЛЯКОВ
(підпис) (Власне ім'я, ПРІЗВИЩЕ)

2025 р.

Харківський національний університет радіоелектроніки
Факультет комп'ютерних наук
Кафедра програмної інженерії
Рівень вищої освіти другий (магістерський)
Спеціальність 121 – Інженерія програмного забезпечення
Тип програми освітньо-наукова
Освітня програма Інженерія програмного забезпечення
(шифр і назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
«___» _____ 2025 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві Шагуну Артему Сергійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи «Дослідження моделей та інформаційних технологій виявлення фейкових новин»
Затверджена наказом по університету від 15.04.2025р. № 290 Ст
2. Термін подання студентом роботи до екзаменаційної комісії 20.06.2025
3. Вихідні дані до роботи Наукові та літературні джерела з тематики виявлення фейкових новин та обробки природної мови, відкриті датасети новин для навчання та тестування моделей, програмне забезпечення Python, бібліотеки машинного навчання та обробки текстів (scikit-learn, xgboost, nltk, spacy), результати експериментів із порівняння ефективності моделей.
4. Перелік питань, що потрібно опрацювати в роботі Дослідження методів виявлення фейкових новин, аналіз існуючих моделей машинного навчання для обробки текстових даних, вивчення підходів до попередньої обробки природної мови, порівняння ефективності різних алгоритмів класифікації (Logistic Regression, SVM, XGBoost), розробка програмної системи для автоматизованого визначення достовірності новинних текстів.

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Отримання завдання	16.04.2025	*виконано*
2	Аналіз предметної галузі і постановка задачі	20.04.2025	*виконано*
3	Огляд й аналіз літературних джерел	22.04.2025	*виконано*
4	Теоретичне дослідження. Багатокритеріальна задача	29.04.2025	*виконано*
5	Підготовка до апробації результатів дослідження. Публікація матеріалів	01.05.2025	*виконано*
6	Програмна реалізація прототипу системи	24.05.2025	*виконано*
7	Підготовка пояснювальної записки	27.05.2025	*виконано*
8	Підготовка презентації та доповіді	31.05.2025	*виконано*
9	Перевірка на плагіат	17.06.2025	*виконано*
10	Нормоконтроль	18.06.2025	*виконано*
11	Рецензування	18.06.2025	*виконано*
12	Попередній захист	19.06.2025	*виконано*
13	Занесення диплома в електронний архів	19.06.2025	*виконано*
14	Допуск до захисту у зав. кафедри	19.06.2025	*виконано*

Дата видачі завдання 16.04.2025р.

Студент (ка)  Артем ШАГУН
(підпис)

Керівник роботи _____ проф. Сергій СМЕЛЯКОВ
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ / ABSTRACT

Пояснювальна записка містить: 68 с., 24 рис., 4 табл., 15 джерел.

ДЕЗІНФОРМАЦІЯ, МАШИННЕ НАВЧАННЯ, МОДЕЛІ ШТУЧНОГО ІНТЕЛЕКТУ, ОБРОБКА ПРИРОДНОЇ МОВИ, ФЕЙКОВІ НОВИНИ.

Об’єктом дослідження є процеси поширення фейкових новин у цифровому середовищі.

Метою роботи є дослідження ефективності моделей штучного інтелекту для виявлення фейкових новин та розробка інструментів, які сприятимуть підвищенню інформаційної безпеки.

Методами розробки та проєктування є аналіз наукових і літературних джерел, методи обробки природної мови (NLP), машинного і глибинного навчання, а також експериментальні дослідження для порівняння ефективності алгоритмів.

У результаті роботи розроблено критерії для оцінки моделей машинного навчання, визначено найбільш ефективні підходи до виявлення фейкових новин та запропоновано архітектуру системи для автоматичного виявлення дезінформації.

DISINFORMATION, MACHINE LEARNING, ARTIFICIAL INTELLIGENCE MODELS, NATURAL LANGUAGE PROCESSING, FAKE NEWS.

The research object is the processes of fake news spread in the digital environment.

The work's purpose is to investigate the efficiency of artificial intelligence models for detecting fake news and to develop tools that enhance information security.

Development and design methods include analyzing scientific and literary sources, natural language processing (NLP) methods, machine and deep learning approaches, and experimental research to compare algorithm efficiency.

As a result of the work, criteria for evaluating machine learning models were developed, the most efficient approaches to fake news detection were identified, and an architecture for an automated disinformation detection system was proposed.

Завідувачу кафедри
ПІ_____
(скорочена назва кафедри)
проф. Кирилу СМЕЛЯКОВУ_
(вчене звання, сласне ім’я, прізвище)

ЗАЯВА
щодо самостійності виконання кваліфікаційної роботи та можливості її публікації (та/або публікації анотації кваліфікаційної роботи) в електронному архіві відкритого доступу ElAr KhNURE

Я, Шагун Артем Сергійович, гр. ІПЗм-23-1, здобувач вищої освіти на другому (магістерському) рівні кафедри «Програмна інженерія», заявляю: мій комплексний курсовий проєкт на тему «Дослідження моделей та інформаційних технологій виявлення фейкових новин», що буде представлений для публічного захисту, виконаний самостійно, не містить елементи плагіату і може бути опублікований в електронному архіві з відкритим доступом ElArKhNURE. Всі запозичення з друкованих та електронних джерел мають відповідні посилання.

Я ознайомлений(на) з діючим положенням «Про протидію академічному плагіату в ХНУРЕ», відповідно до якого виявлення плагіату є підставою для відмови в допуску роботи до захисту та застосування дисциплінарних заходів.

17.06.2025

ЗМІСТ

ВСТУП

У сучасному інформаційному просторі проблема поширення фейкових новин набула особливої актуальності. Стрімкий розвиток цифрових платформ, соціальних мереж та масовий доступ до створення контенту призвели до значного зростання кількості недостовірної інформації, що може маніпулювати суспільною думкою, створювати соціальну напругу та загрожувати інформаційній безпеці. Традиційні методи ручної перевірки є малоефективними в умовах великих обсягів даних, що зумовлює потребу у розробці автоматизованих систем виявлення фейкових новин на основі сучасних технологій обробки текстової інформації та машинного навчання.

Мета роботи полягає у розробці та дослідженні ефективності програмної системи для автоматизованого виявлення фейкових новин із використанням моделей машинного навчання та методів обробки природної мови.

Об’єктом дослідження є процеси класифікації текстової інформації у завданні визначення достовірності новинних повідомлень.

Предметом дослідження є алгоритми машинного навчання та методи лінгвістичної обробки текстових даних, що застосовуються для виявлення фейкових новин.

До методів дослідження і аналізу входять: аналіз літературних джерел з тематики обробки природної мови та класифікації текстів, методи машинного навчання (Logistic Regression, Support Vector Machine, XGBoost), методи векторизації тексту (TF-IDF), методи передобробки текстів (очищення, лематизація, видалення стоп-слів), експериментальні дослідження ефективності моделей машинного навчання.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ І ПОСТАНОВКА ЗАДАЧІ

1.1 Аналіз предметної галузі

Фейкові новини, або дезінформація, є одним із найбільш значущих викликів сучасного інформаційного простору. Вони являють собою неправдиву або навмисно спотворену інформацію, що створюється з метою маніпуляції громадською думкою, провокування суспільних конфліктів, підриву довіри до офіційних джерел або ж задоволення економічних чи політичних інтересів певних груп. У сучасному світі, де цифрові технології та Інтернет стали невід'ємною частиною життя, поширення фейкових новин набуло небачених масштабів. Завдяки соціальним медіа, таким як Facebook, Twitter, TikTok та Instagram, дезінформація може розповсюджуватися з неймовірною швидкістю, охоплюючи мільйони користувачів за лічені години [1].

Суть проблеми полягає не лише у створенні фейкових новин, а й у тому, як саме вони поширюються та впливають на різні аспекти суспільного життя. Такі новини часто спрямовані на створення паніки, формування хибних уявлень про певні події або навіть маніпуляцію масовою свідомістю під час виборчих кампаній чи інших суспільно важливих подій. В умовах сучасного інформаційного суспільства люди нерідко стикаються з труднощами у відокремленні правди від брехні через обмеженість часу на перевірку фактів, а також через значний вплив емоційного забарвлення матеріалів, які поширюються онлайн.

Технологічний аспект проблеми поширення фейкових новин також заслуговує на увагу. Сучасні алгоритми соціальних медіа, побудовані на основі машинного навчання, часто сприяють поширенню таких матеріалів через механізми персоналізації контенту. Це означає, що користувачі отримують на свої сторінки матеріали, які підтверджують їхні упередження або викликають сильні емоційні реакції, незалежно від їхньої достовірності. Крім того, зростання доступності інструментів для створення фейкового контенту, таких як генератори зображень або текстів на основі штучного інтелекту, ще більше ускладнює ситуацію. Завдяки цим технологіям можна створювати контент, який виглядає

настільки правдоподібно, що навіть експерти не завжди можуть одразу визначити його неправдивість.

Предметна галузь дослідження фейкових новин знаходиться на стику декількох дисциплін, кожна з яких відіграє важливу роль у вирішенні цієї проблеми. Однією з ключових галузей є обробка природної мови (NLP), яка дозволяє аналізувати текстовий контент, виявляти патерни, характерні для неправдивої інформації, та автоматично класифікувати тексти на основі набору лінгвістичних або статистичних ознак. Машинне навчання та штучний інтелект, зокрема, відкривають нові можливості для розробки моделей, здатних не лише розпізнавати фейкові новини, а й прогнозувати їхній вплив на аудиторію. Особливе значення тут має навчання моделей на великих наборах даних, які включають як достовірний, так і неправдивий контент.

Крім того, значний внесок у дослідження цієї теми роблять соціальні науки, які вивчають поведінку користувачів у цифровому середовищі, механізми поширення інформації та фактори, які сприяють довірі до певних джерел. Знання у цій сфері дозволяють зрозуміти, чому деякі новини отримують більше довіри та уваги, ніж інші, а також які психологічні особливості спонукають людей поширювати дезінформацію. Вивчення таких процесів є ключовим для створення ефективних механізмів боротьби з фейковими новинами, адже проблема не лише технологічна, а й соціальна.

1.2 Виявлення проблем та актуалізація рішень

Проблема фейкових новин не є новою, але сучасні технології суттєво змінили її характер та масштаби. Головною проблемою є те, що традиційні методи перевірки інформації стають недостатньо ефективними у світі, де інформація поширюється миттєво, а її обсяг постійно зростає. Уявлення про "цифрову епоху правди" зіштовхнулося з реальністю, коли дезінформація може стати вірусною швидше, ніж її можна перевірити, а її наслідки можуть завдавати шкоди довгічно. В основі проблеми лежать кілька ключових викликів, які роблять боротьбу з фейковими новинами складною та багатогранною.

По-перше, одним із головних факторів є масштабність поширення інформації через соціальні мережі. Алгоритми цих платформ сприяють "віральності" контенту, часто ігноруючи його достовірність. Система ранжування новин базується на залученості користувачів – кількості лайків, репостів та коментарів, а не на фактичній перевірці змісту. Це створює ідеальні умови для розповсюдження фейків, адже емоційно забарвлений контент, зокрема сенсаційні заголовки або провокаційні заяви, має більшу ймовірність привернути увагу.

По-друге, виникає проблема швидкого створення фейкових новин завдяки сучасним технологіям. Такі інструменти, як генератори текстів GPT, deepfake-відео або навіть базові редактори для зображень, дозволяють створювати високоякісний контент, який дуже важко відрізнити від справжнього. Це посилює проблему, оскільки навіть професійні фактчекери стикаються з труднощами у перевірці таких матеріалів. Більше того, високий рівень автоматизації дозволяє створювати велику кількість фейків за короткий проміжок часу, що суттєво перевищує можливості їх виявлення.

Ще однією важливою проблемою є людський фактор. Користувачі соціальних мереж часто поширюють інформацію без перевірки її джерела або контексту, керуючись емоціями чи власними упередженнями. Це підсилюється ефектом "інформаційної бульбашки", коли алгоритми соціальних мереж демонструють користувачам лише ті новини, які відповідають їхнім уподобанням і переконанням. Як результат, формуються закриті інформаційні простори, де фейкові новини поширюються безперешкодно.

Актуалізація рішень у боротьбі з фейковими новинами є нагальною потребою, яка потребує комплексного підходу. Одним із ключових напрямів є розробка автоматизованих систем виявлення фейкових новин, які базуються на машинному навчанні та обробці природної мови. Такі системи аналізують тексти на рівні синтаксису, семантики та стилістики, дозволяючи виявляти певні закономірності, притаманні неправдивій інформації. Крім того, важливу роль відіграє аналіз соціальних графів, що дозволяє відстежувати шляхи поширення новин та ідентифікувати джерела дезінформації.

Ще одним важливим рішенням є створення механізмів, які дозволяють користувачам перевіряти достовірність інформації у реальному часі. Це можуть бути браузерні розширення або мобільні додатки, які попереджають про можливу фейковість новин, базуючись на перевірених базах даних фактчекерів або автоматизованих алгоритмах. Також ефективними є освітні ініціативи, спрямовані на підвищення рівня медіаграмотності серед населення. Люди повинні розуміти, як працюють алгоритми соцмереж, як перевіряти джерела інформації та які інструменти доступні для боротьби з дезінформацією.

Рівень інноваційності цього дослідження полягає у використанні передових технологій штучного інтелекту, зокрема глибокого навчання, для створення моделей, які не лише аналізують текстовий контент, але й враховують контекст його поширення. Інтеграція NLP з аналізом соціальних мереж дає змогу отримати комплексний підхід до виявлення фейкових новин. Крім того, новизна дослідження полягає у застосуванні підходів до прогнозування можливого впливу дезінформації на аудиторію, що дозволяє ідентифікувати найбільш небезпечні матеріали та вживати заходів для їх нейтралізації. Такий підхід є унікальним, оскільки поєднує технічні, соціальні та психологічні аспекти боротьби з фейковими новинами, роблячи дослідження надзвичайно актуальним та перспективним для практичного впровадження.

2 ОГЛЯД Й АНАЛІЗ ЛІТЕРАТУРНИХ, НАУКОВИХ ДЖЕРЕЛ

2.1 Критерії вибору джерел

Вибір джерел для дослідження фейкових новин є одним із найважливіших етапів наукової роботи, оскільки від якості та достовірності використаних матеріалів залежить точність аналізу та обґрунтованість висновків. У цьому дослідженні особлива увага приділялася науковій цінності та авторитетності джерел. Використовувалися публікації в рецензованих наукових журналах, монографіях, матеріалах конференцій та інших академічних виданнях, що мають підтверджену репутацію у сферах обробки природної мови, машинного навчання, та соціології.

Критичним аспектом також була актуальність джерел. Перевага надавалася публікаціям, опублікованим протягом останніх п’яти років, адже проблема фейкових новин тісно пов’язана з технологічними інноваціями, які швидко змінюються. Водночас були враховані фундаментальні праці, які визначають теоретичні основи досліджуваної теми. Особливе значення мала міждисциплінарність джерел, оскільки проблема фейкових новин охоплює декілька сфер.

Ще одним важливим критерієм була доступність джерел для перевірки. Усі використані матеріали повинні були бути доступними для подальшого аналізу і підтвердження отриманих результатів.

Крім того, релевантність джерел до теми дослідження мала вирішальне значення. Увага здебільшого зосереджувалася на матеріалах, які стосуються фейкових новин, їх поширення, виявлення та впливу.

Також враховувалася якість методології досліджень. Перевага надавалася роботам, які базуються на статистичних аналізах, експериментальних даних, машинному навчанні чи комплексному аналізі соціальних мереж. Важливим критерієм була наявність як кількісних, так і якісних даних у джерелах, оскільки це дозволяло не лише отримувати точні показники, але й формулювати обґрунтовані висновки для подальшого аналізу. Дотримання цих критеріїв дозволило створити базу літератури, яка є максимально достовірною, актуальною

та корисною для дослідження фейкових новин. Такий підхід забезпечує високу наукову цінність та обґрунтованість отриманих результатів.

2.2 Аналіз літератури

Проблема поширення фейкових новин стала серйозним викликом для сучасного інформаційного простору. У науковій літературі представлено широкий спектр підходів до вирішення цієї проблеми, які базуються на методах обробки природної мови (NLP), машинного та глибинного навчання, а також міждисциплінарних дослідженнях.

У дослідженні Висоцької та співавторів запропоновано інтелектуальну систему для автоматичного передбачення фейкових новин. Ця система використовує технології обробки природної мови для аналізу лексичних, синтаксичних і семантичних особливостей тексту. Особливу увагу приділено визначенню маніпулятивної мови, яка часто використовується у фейкових новинах. Застосування методів глибокого навчання, таких як рекурентні нейронні мережі (RNN) та згорткові нейронні мережі (CNN), дозволяє виявляти приховані патерни, які не є очевидними при традиційному аналізі тексту. Крім того, система оцінює джерела інформації, виявляючи маловідомі або неперевірені сайти, які можуть бути джерелом фейків [2].

У роботі Чируна та Романчука розглянуто підходи до класифікації фейкових новин, що включають аналіз текстових, візуальних та змішаних даних. Автори наголошують на важливості використання мультимодальних моделей, які дозволяють враховувати не лише текстові особливості, але й зображення та відео, що супроводжують новини. Наприклад, технології розпізнавання облич (face recognition) і аналізу візуального контенту допомагають виявляти підроблені фотографії або відео. У статті також відзначено високу ефективність згорткових нейромереж (CNN) у задачах аналізу зображень і тексту, що дозволяє інтегрувати кілька джерел інформації для більш точного визначення фейкових новин [3].

Класичні методи машинного навчання, такі як наївний Баєс, метод опорних векторів (SVM), модель k-найближчих сусідів (kNN) та випадковий ліс,

досліджені у роботі Висоцької та Свища. Автори провели порівняльний аналіз ефективності цих алгоритмів у задачах класифікації текстових новин. Результати показали, що наївний Баєс і SVM є більш ефективними для невеликих наборів даних, тоді як випадковий ліс краще справляється з великими обсягами даних завдяки своїй здатності до обробки складних взаємозв'язків між ознаками [4].

У статті Деркача запропоновано модифікації для вдосконалення згорткових нейромереж (CNN) у задачах виявлення фейкових новин. Автор запропонував збільшення кількості нейронів у згорткових шарах та використання шару випадкового відключення (dropout), що дозволяє моделі краще адаптуватися до специфічних патернів маніпулятивної мови, зменшуючи ризик перенавчання. В результаті покращується точність класифікації, особливо для складних наборів даних із різноманітними джерелами новин [5].

У роботі Zhou та співавторів запропоновано міждисциплінарну модель для раннього виявлення фейкових новин. Автори використовують лексичний, синтаксичний, семантичний та дискурсивний аналіз для комплексної оцінки контенту. Особливу увагу приділено ранньому виявленню новин ще до їх широкого розповсюдження. Автори зазначають, що врахування соціальних та психологічних аспектів поширення інформації суттєво підвищує точність моделей [6].

Науковці катедри програмної інженерії Харківського національного університету радіоелектроніки також активно досліджують проблематику обробки природної мови та генерації тексту. Зокрема, Олексій Турута та Андрій Бабій опублікували огляд сучасних підходів до генерації природної мови, що охоплює багатомовність, мультимодальність, контрольованість та навчання моделей. Ця робота висвітлює актуальні задачі та методи генерації текстів, що є важливими для розуміння та виявлення фейкових новин [7].

Крім того, дослідники катедри, зокрема Костянтин Онищенко, Яна Данієль та Роман Каменєв, провели аналіз методів обробки природної мови, виділивши перспективні напрямки розвитку галузі та окресливши переваги й недоліки

розглянутих методів у задачах «розуміння» природної мови, перекладу тексту та відповіді на питання [8].

2.3 Оцінка актуальності та новизни

Фейкові новини є однією з найгостріших проблем сучасного інформаційного середовища, оскільки вони мають значний вплив на громадську думку, політичні процеси, економічну стабільність і соціальну єдність. Їхнє поширення набуває особливої небезпеки у часи криз, таких як виборчі кампанії, пандемії або військові конфлікти, коли суспільство є вразливішим до маніпуляцій. Швидке зростання обсягу інформаційних потоків у цифрових медіа, відсутність належного контролю за їхнім поширенням і доступність сучасних технологій, що дозволяють легко створювати маніпулятивний контент, роблять цю проблему ще більш актуальною.

Тема дослідження є особливо важливою в умовах, коли існуючі системи автоматизованої перевірки фактів і класифікації новин ще не забезпечують достатнього рівня точності. Хоча методи глибинного навчання та обробки природної мови демонструють значний прогрес, вони все ще стикаються з низкою труднощів, зокрема з багатомовністю текстів, неоднорідністю стилів написання, нестачею якісних навчальних даних та труднощами верифікації джерел. Ці проблеми підкреслюють потребу у вдосконаленні існуючих підходів.

Новизна цього дослідження полягає у розробці комплексного підходу до виявлення фейкових новин, який інтегрує сучасні методи обробки текстів із мультимодальним аналізом, враховуючи текстові, візуальні та соціальні аспекти новинного контенту. Використання удосконалених моделей глибинного навчання дозволяє враховувати складну структуру тексту, виявляти приховані патерни маніпуляцій та забезпечувати адаптацію моделей до нових типів фейкових новин. Зокрема, особлива увага приділяється інтеграції аналізу текстів із технологіями розпізнавання зображень, що є ключовим для виявлення маніпулятивних фото- та відеоматеріалів.

Дослідження також зосереджується на задачі раннього виявлення фейкових новин, що є надзвичайно важливим у боротьбі з їхнім поширенням. Виявлення новин на початкових етапах розповсюдження дозволяє значно зменшити їхній вплив на аудиторію, проте ця задача потребує високоточного аналізу текстів із мінімальним контекстом. Запропоновані у дослідженні підходи також враховують соціальні та культурні особливості поширення інформації, що підвищує ефективність класифікації у різних регіональних і мовних середовищах.

Окрім технічних аспектів, дослідження має значний практичний внесок. Розробка систем, здатних автоматично виявляти фейкові новини, є критично важливою для платформ соціальних медіа, новинних агенцій та організацій, що займаються перевіркою фактів. Це дозволяє забезпечити більш якісне інформування суспільства, зменшити рівень дезінформації та сприяти формуванню стійкості суспільства до маніпуляцій.

Таким чином, актуальність даного дослідження визначається його спрямованістю на вирішення ключових проблем інформаційної безпеки, а новизна полягає у розробці ефективних, комплексних підходів до виявлення фейкових новин, що враховують різноманітні аспекти цієї складної проблеми.

2.4 Висновки з огляду

Аналіз літератури та сучасних підходів до виявлення фейкових новин демонструє значний прогрес у застосуванні методів машинного навчання, обробки природної мови та мультимодального аналізу. Виявлено, що найбільш ефективними є комплексні моделі, які інтегрують аналіз текстових, візуальних і соціальних аспектів новинного контенту. Використання глибокого навчання, зокрема згорткових і рекурентних нейронних мереж, дозволяє ідентифікувати приховані патерни маніпуляцій і підвищує точність класифікації.

У літературі підкреслено важливість раннього виявлення фейкових новин, яке дозволяє зменшити їхній вплив на суспільство. Проте ця задача залишається складною через обмежений контекст на початкових етапах поширення новин.

Особливої уваги потребує багатомовність та різноманітність стилів написання, що часто ускладнює автоматичну класифікацію.

Недоліки існуючих рішень включають недостатню адаптованість моделей до специфічних соціальних і культурних контекстів, складність у верифікації джерел інформації та низьку доступність високоякісних навчальних даних. Ці проблеми визначають напрями подальших досліджень, серед яких важливими є створення удосконалених моделей для аналізу гетерогенних даних, підвищення їхньої стійкості до нових типів маніпулятивного контенту та інтеграція сучасних технологій перевірки фактів.

Таким чином, огляд літератури підтверджує актуальність і необхідність розробки комплексного підходу до виявлення фейкових новин, який інтегрує сучасні алгоритми обробки текстів, аналізу зображень і враховує соціальні аспекти поширення інформації. Отримані результати огляду створюють основу для формування чіткої методології дослідження, спрямованого на вирішення зазначених проблем.

3 ПОСТАНОВКА ЗАДАЧІ

3.1 Формулювання задачі

Основною метою дослідження є розробка моделі, яка здатна автоматично виявляти фейкові новини з високим рівнем точності, використовуючи сучасні методи машинного навчання та обробки текстів.

По-перше, необхідно визначити ключові лексичні, синтаксичні та семантичні характеристики, які є характерними для фейкових новин. Це дозволить створити набір ознак, які будуть використовуватися для класифікації текстів.

По-друге, слід розробити методи аналізу текстів, які виявляють маніпулятивні патерни, такі як використання емоційно забарвленої мови, некоректні факти або тенденційне представлення інформації.

По-третє, важливим завданням є створення моделі машинного навчання або глибокого навчання, яка зможе ефективно працювати з текстовими даними різних форматів, враховуючи багатомовність контенту та його гетерогенність.

По-четверте, необхідно забезпечити можливість раннього виявлення фейкових новин, що дозволить оперативно реагувати на їхнє поширення та зменшити їхній вплив на аудиторію.

Таким чином, реалізація поставлених завдань дозволить створити високоефективний інструмент для автоматичного виявлення фейкових новин, який сприятиме підвищенню інформаційної безпеки та мінімізації впливу дезінформації на суспільство.

3.2 Обґрунтування вибору методів дослідження

Для вирішення поставленої задачі обрано методи, які базуються на сучасних підходах до обробки текстів і машинного навчання. Вибір цих методів зумовлений складністю та багатогранністю проблеми виявлення фейкових новин, а також необхідністю забезпечити високу точність і адаптивність моделі до різноманітних даних.

По-перше, для аналізу текстів використовуються методи обробки природної мови (NLP). Ці методи дозволяють виявляти ключові лексичні, синтаксичні та семантичні особливості тексту. Наприклад, токенізація, лематизація та аналіз частоти слів допомагають структурувати текстові дані, що є основою для подальшого машинного навчання.

По-друге, методи машинного навчання, такі як логістична регресія, дерево рішень і метод опорних векторів (SVM), дозволяють будувати базові класифікатори для виявлення фейкових новин. Ці алгоритми забезпечують швидке та ефективне навчання на структурованих наборах даних і є основою для побудови більш складних моделей.

По-третє, глибинне навчання, зокрема використання рекурентних нейронних мереж (RNN) або трансформерів, таких як BERT, обрано для аналізу складних патернів у текстах. Ці моделі здатні враховувати контекст і семантику текстів, що є важливим для виявлення маніпулятивної мови та прихованих зв’язків у текстах.

По-четверте, для оцінки ефективності моделі використовуються метрики класифікації, такі як точність (accuracy), повнота (recall), точність прогнозування (precision) та F1-міра. Ці показники дозволяють об’єктивно оцінити якість моделі та виявити її сильні та слабкі сторони.

Вибір зазначених методів дослідження обґрунтований їхньою здатністю обробляти великі обсяги текстових даних, адаптуватися до різних стилів і мов текстів, а також забезпечувати високий рівень точності класифікації. Використання сучасних моделей глибинного навчання, таких як трансформери, дозволяє враховувати складні мовні особливості, що значно підвищує ефективність аналізу текстів.

Таким чином, застосування методів обробки текстів, машинного та глибинного навчання є оптимальним вибором для розв’язання задачі автоматичного виявлення фейкових новин, що відповідає сучасним тенденціям у галузі інформаційної безпеки та аналізу даних.

3.3 Обмеження дослідження

Основні обмеження даного дослідження пов’язані з характером даних, вибором методів і особливостями реалізації моделей.

По-перше, якість результатів моделі значною мірою залежить від набору даних, на якому вона навчається. У випадку дослідження фейкових новин можливі обмеження, пов’язані з неповнотою, нерепрезентативністю або недостатнім обсягом навчальних даних. Наприклад, якщо набір даних містить новини лише певного регіону, мови або тематики, це може призвести до зниження ефективності моделі під час аналізу текстів іншого характеру.

По-друге, складність мови та стилістики текстів є ще одним важливим фактором, що може обмежувати ефективність моделі. Моделі машинного та глибинного навчання можуть мати труднощі з аналізом текстів, які містять сарказм, іронію, культурно специфічні вирази або багатозначні слова.

По-третє, вибрані методи обробки текстів і глибинного навчання мають свої обмеження. Наприклад, складні моделі, такі як трансформери, потребують великих обсягів обчислювальних ресурсів і можуть бути менш ефективними при роботі з невеликими наборами даних. Крім того, такі моделі часто є «чорними ящиками», що ускладнює інтерпретацію результатів та ідентифікацію причин помилок.

По-четверте, механізми раннього виявлення фейкових новин залежать від обсягу та доступності інформації. На початкових етапах поширення новини контекст може бути обмеженим, що ускладнює ідентифікацію її достовірності.

По-п’яте, багатомовність є ще одним викликом для дослідження. Моделі, навчені на текстах однією мовою, можуть показувати низьку ефективність під час аналізу текстів іншими мовами через відмінності у граматиці, семантиці та стилістиці.

По-шосте, це дослідження зосереджене виключно на текстовому аналізі та не враховує візуальних або відеоматеріалів, які часто супроводжують новини. Це може обмежувати застосування результатів у ситуаціях, де маніпуляція відбувається не через текст, а через зображення або відео.

Таким чином, результати цього дослідження є важливим кроком у напрямку автоматизації виявлення фейкових новин, але вони не враховують усіх можливих аспектів проблеми. Подолання зазначених обмежень потребує подальших досліджень, які зосередяться на розширенні доступних даних, адаптації моделей до різних мов і стилів текстів, а також інтеграції мультимодального аналізу.

3.4 Необхідні ресурси

Основні ресурси, які необхідні для дослідження, описані нижче.

По-перше, доступ до відповідних наборів даних, що містять тексти фейкових та достовірних новин. Ці набори повинні включати дані різної тематики та походження, щоб забезпечити репрезентативність для навчання та тестування моделі.

По-друге, необхідні обчислювальні ресурси для навчання та тестування моделі. Для цього передбачається використання сучасних графічних процесорів (GPU), які забезпечують високу продуктивність під час обробки великих обсягів даних та навчання моделей глибокого навчання.

По-третє, програмне забезпечення для розробки та реалізації моделі. Для виконання завдання будуть використовуватися сучасні бібліотеки для машинного навчання та обробки текстів, такі як TensorFlow, PyTorch, Scikit-learn, а також інструменти для обробки природної мови, як-от NLTK та SpaCy.

По-четверте, платформи для зберігання та обробки даних. Для цього будуть використовуватися хмарні сервіси або локальні сервери, які здатні обробляти великі обсяги текстової інформації.

По-п’яте, доступ до наукових публікацій і матеріалів для аналізу існуючих підходів. Це забезпечить глибоке розуміння проблеми та допоможе у виборі оптимальних методів і алгоритмів для реалізації задачі.

Таким чином, ефективна реалізація дослідження можлива за умови забезпечення доступу до якісних даних, сучасного обладнання та програмного забезпечення, а також наукової бази для аналізу й оцінки результатів.

4 ТЕОРЕТИЧНЕ ДОСЛІДЖЕННЯ

4.1 Вирішення багатокритеріальної задачі для вибору моделі машинного навчання

Задача полягає у виборі найкращої моделі машинного навчання для автоматичного виявлення фейкових новин. Відповідно, основна мета – визначити модель, яка збалансовує високу точність класифікації, швидкість виконання, мінімальне споживання ресурсів та простоту впровадження в реальні інформаційні системи.

Logistic Regression (LR) – це базова статистична модель, яка використовується для класифікації текстів, забезпечує швидке навчання та простоту реалізації [9].

Переваги:

- простота впровадження;
- висока швидкість навчання та класифікації;
- мінімальне споживання ресурсів.

Недоліки:

- обмежена здатність до роботи з нелінійними залежностями;
- алгоритм складно працює з великим обсягом ознак без попередньої обробки.

Support Vector Machine – метод класифікації, що використовує гіперплощини для розділення класів. Ефективний для невеликих і середніх наборів даних [10].

Переваги:

- висока точність класифікації при оптимальному налаштуванні;
- підходить для роботи з нелінійними даними.

Недоліки:

- високе споживання пам’яті;
- тривалий час навчання при великих наборах даних.

Random Forest – ансамблева модель, що використовує кілька рішень дерев для підвищення точності класифікації [9].

Переваги:

- стійкість до перенавчання;
- гарна продуктивність для різних типів даних;
- простота налаштування.

Недоліки:

- відносно високе споживання пам’яті;
- порівняно тривалий час класифікації.

XGBoost – оптимізована модель градієнтного бустингу, яка широко використовується для вирішення задач класифікації [10].

Переваги:

- висока точність класифікації.;
- можливість паралельного виконання;
- ефективна робота з великими наборами даних.

Недоліки:

- Складність налаштування;
- значне споживання пам’яті.

Bidirectional Encoder Representations from Transformers (BERT) – сучасна модель на основі глибокого навчання, яка використовує трансформери для обробки тексту [11].

Переваги:

- висока точність завдяки контекстуальній обробці тексту;
- підходить для задач з великим обсягом текстових даних.;
- ефективна робота з великими наборами даних.

Недоліки:

- високе споживання пам’яті;
- тривалий час навчання;
- складність впровадження.

Визначимо критерії вибору:

– точність класифікації оцінює, наскільки модель правильно класифікує новини як фейкові або правдиві. Цей критерій необхідний для забезпечення максимальної відповідності реальності;
– час навчання моделі оцінює, скільки часу потрібно для навчання моделі на заданому наборі даних. Цей критерій важливий для оцінки швидкості підготовки моделі до використання у реальних умовах;
– час класифікації одного зразка оцінює середній час, необхідний для класифікації однієї новини. Цей показник є критичним для використання моделі у системах реального часу;
– споживання пам’яті оцінює обсяг оперативної пам’яті, необхідний для роботи моделі. Цей критерій важливий для визначення ефективності використання ресурсів моделлю;
– простота впровадження оцінює рівень складності налаштування та інтеграції моделі. Таким чином, простота впровадження впливає на швидкість і легкість використання моделі;

Визначимо шкали для кожного з критеріїв.

Точність класифікації відповідає за відсоткову кількість новин, що були правильно класифіковані. Оцінюється по шкалі відношень.

Приклади значень:

– 0-60%: низька точність класифікації;
– 61-85%: середня точність класифікації;
– 86-100%: висока точність класифікації.

Час навчання моделі відповідає за середню кількість хвилин, необхідних для навчання моделі на заданому наборі даних. Оцінюється по шкалі відношень.

– 1-30 хв: низький час навчання;
– 31-60 хв: середній час навчання;
– 60-200 хв: довгий час навчання.

Час класифікації оцінюється по шкалі відношень:

– 0.1-10 мс: низький час класифікації;

– 11-40 мс: середній час класифікації;
– 41-200 мс: довгий час класифікації.

Споживання пам’яті визначається по рівню використанню пам’яті по порядковій шкалі.

Значення шкали:

– 1 бал: високе споживання пам’яті. Модель вимагає значних ресурсів, що може бути проблемою для обмежених середовищ.
– 2 бали: середнє споживання пам’яті. Модель працює з помірним використанням пам’яті, придатна для більшості сучасних систем;
– 3 бали: низьке споживання пам’яті. Модель працює ефективно, споживаючи мінімум ресурсів.

Простота застосування відповідає за оцінку простоти реалізації, зрозумілості алгоритму та наявності прикладів. Оцінюється по шкалі порядку.

Приклади значень:

– 1 бал: складно реалізується, алгоритм важкий;
– 2 бали: реалізація середньої важкості;
– 3 бали: реалізація є простою.

Оскільки критерії та шкали оцінок визначено, можна перейти до векторного опису альтернатив за обраними критеріями.

В таблиці 4.1 наведено векторний опис альтернатив за обраними критеріями.

Застосуємо принцип Парето для визначення найгірших методів навчання.

Логістична регресія демонструє сильні результати за критеріями часу навчання, класифікації та простоти впровадження. Проте вона поступається іншим моделям за точністю класифікації. Незважаючи на цю слабкість, завдяки своїй збалансованості за іншими критеріями, логістична регресія входить до множини Парето-оптимальних альтернатив.

Метод опорних векторів (SVM) демонструє найвищу точність класифікації серед усіх моделей, окрім BERT. Водночас він поступається логістичній регресії за часом навчання та класифікації, але зберігає конкурентоспроможність завдяки

середньому споживанню пам’яті та простоті впровадження. Отже, SVM залишається у множині Парето-оптимальних альтернатив.

Таблиця 4.1 – Векторний опис альтернатив за обраними критеріями (таблиця виконана самостійно)

Критерії/Моделі	Logistic regression	Support Vector Machine	Random Forest	XGBoost	BERT
Точність класифікації	80.52	94.4	94.4	79.22	99
Час навчання моделі	3	10	20	30	90
Час класифікації	0.5	5	10	15	120
Споживання пам’яті	3	2	2	2	1
Простота впровадження	3	2	2	1	1

Random Forest демонструє помірні результати за більшістю критеріїв, але поступається SVM і логістичній регресії за часом класифікації та простотою впровадження. Через ці обмеження Random Forest виключається зі списку Парето-оптимальних альтернатив.

XGBoost має сильні результати за точністю класифікації та помірне споживання пам’яті. Однак ця модель поступається іншим за часом навчання та класифікації, а також складна у впровадженні. Незважаючи на ці недоліки, завдяки високій точності класифікації XGBoost залишається Парето-оптимальною альтернативою.

BERT демонструє найвищу точність класифікації серед усіх моделей, проте його час навчання та класифікації, а також споживання пам’яті значно перевищують відповідні показники інших альтернатив. Через ці обмеження BERT виключається зі списку Парето-оптимальних альтернатив, якщо завдання не вимагає найвищої точності та немає доступу до потужних обчислювальних ресурсів.

Отже, до множини Парето-оптимальних альтернатив входять логістична регресія, SVM та XGBoost, які демонструють найбільш збалансовані результати за всіма критеріями (див. табл. 4.2).

Таблиця 4.2 – Аналіз за принципом Парето (таблиця виконана самостійно)

Критерії/Моделі	Logistic regression	Support Vector Machine	XGBoost
Точність класифікації	80.52	94.4	79.22
Час навчання моделі	3	10	30
Час класифікації	0.5	5	15
Споживання пам’яті	3	2	2
Простота впровадження	3	2	1

Далі виконаємо нормування критеріїв «Точність класифікації», «Час навчання моделі» та «Простота впровадження» за формулою приведення оптимальності «за максимум» (див. формулу 5.1), оскільки чим більші ці величини, тим краще.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{4.1}$$

Нормування критеріїв «Час класифікації», «Споживання пам’яті» відбувається за формулою приведення оптимальності «за мінімумом» (див. формулу 4.2), оскільки чим менше ці величини, тим краще.

$$x' = 1 - \frac{x - x_{min}}{x_{max} - x_{min}} \tag{4.2}$$

В таблиці 4.3 наведено нормування оцінок.

Таблиця 4.3 – Нормування оцінок (таблиця виконана самостійно)

Критерії/Моделі	Logistic regression	Support Vector Machine	XGBoost
Точність класифікації	0.8	0.94	0.79
Час навчання моделі	0.99	0.6	0.86
Час класифікації	0.99	0.75	0.93
Споживання пам’яті	1	0.5	0.5
Простота впровадження	1	0.5	0

В якості згорткової моделі обрано лінійну адитивну згортку з ваговими коефіцієнтами. Ця модель дозволяє враховувати різну важливість критеріїв і ефективно комбінувати їх у єдину оцінку. Вагові коефіцієнти визначалися на основі їх важливості для практичного застосування.

Маємо 5 критеріїв, тому найнижчою кількістю балів буде 1, а найвищою кількістю балів – 5.

Точність класифікації (К1) – найважливіший критерій, оскільки визначає якість результатів моделі. Призначено вагу 5 балів.

Час класифікації (К2) – критично важливий для продуктивності в реальних умовах. Вага – 4 бали.

Час навчання (К3) – важливий, але менш критичний для кінцевого користувача. Вага – 3 бали.

Споживання пам’яті (К4) – середньої важливості для оцінки ефективності системи. Вага – 2 бали.

Простота впровадження (К5) – найменш важливий критерій, але враховується для зручності роботи. Вага – 1 бал.

Отже, для обраної згорткової моделі і критеріїв, визначення корисності альтернатив буде виглядати наступною формулою 5.3:

$$max \sum_{j=1}^{6} w_j * x_{ij} = \frac{5}{15} * x_{i1} + \frac{4}{15} * x_{i2} + \frac{3}{15} * x_{i3} + \frac{2}{15} * x_{i4} + \frac{1}{15} * x_{i5} \qquad (5.3)$$

де w_j – ваговий коефіцієнт,

x_{ij}– нормоване значення для і-го методу і j-го критерію.

Проводимо розрахунки для кожної моделі.

Для Logistic Regression:

$$\frac{5}{15} * 0.8 + \frac{4}{15} * 0.99 + \frac{3}{15} * 0.99 + \frac{2}{15} * 1 + \frac{1}{15} * 1 = 0.87$$

Для Support Vector Machine:

$$\frac{5}{15} * 0.94 + \frac{4}{15} * 0.600 + \frac{3}{15} * 0.75 + \frac{2}{15} * 0.5 + \frac{1}{15} * 0.5 = 0.77$$

Для XGBoost:

$$\frac{5}{15} * 0.79 + \frac{4}{15} * 0.86 + \frac{3}{15} * 0.93 + \frac{2}{15} * 0.5 + \frac{1}{15} * 0.000 = 0.81$$

Згідно з проведеними розрахунками, найбільш ефективною моделлю для задачі виявлення фейкових новин є Logistic Regression, яка отримала найвищий бал корисності (0.87). Ця модель забезпечила оптимальний баланс між точністю, швидкістю класифікації, простотою впровадження та ефективним використанням пам’яті. Logistic Regression є простою в реалізації, що робить її зручною для застосування, а також вимагає мінімальних обчислювальних ресурсів.

Другою за ефективністю є XGBoost із балом корисності 0.77. Ця модель має високу точність класифікації та добре справляється з обробкою великих наборів даних. Однак її продуктивність обмежується середніми показниками часу класифікації та впровадження. XGBoost може бути корисною у випадках, коли потрібна висока точність і доступні додаткові ресурси для її роботи.

На третьому місці знаходиться Support Vector Machine з корисністю 0.81. Ця модель отримала високий бал за точність класифікації, але поступається іншим моделям через низькі оцінки за час навчання, споживання пам’яті та складність впровадження. Її застосування може бути виправдане у випадках, де точність є пріоритетною, а інші критерії менш важливі.

Таким чином, для задачі виявлення фейкових новин, де важливі всі критерії – точність, швидкість, простота впровадження та ефективність використання ресурсів, найкращим вибором є Logistic Regression. Вона забезпечує збалансовану продуктивність та найвищу загальну корисність, що робить її оптимальним варіантом для впровадження.

4.2 Архітектура та проєктування ПЗ

У межах виконання дослідження було розроблено програмне забезпечення для виявлення фейкових новин, яке ґрунтується на сучасних методах обробки природної мови та машинного навчання. Архітектура системи побудована таким

чином, щоб забезпечити модульність, гнучкість і можливість подальшого розширення функціональності.

Підготовка текстових даних здійснюється на окремому етапі передобробки. Вона включає очищення тексту від небуквених символів, приведення його до нижнього регістру, лематизацію та видалення стоп-слів. Для виконання цих операцій використано бібліотеки nltk та spacy, які забезпечують якісну обробку англомовних текстів і дозволяють виділяти найбільш значущі лексичні одиниці.

Після передобробки тексти перетворюються у числове представлення за допомогою методу TF-IDF. Для збереження контекстуальної інформації використовуються біграми, а загальна кількість ознак обмежена 5000 найінформативнішими. Це дозволяє отримати компактні та інформативні вектори ознак, які надалі слугують вхідними даними для моделей класифікації.

Реалізована система підтримує навчання декількох моделей машинного навчання: Logistic Regression, Support Vector Machine (SVM) та XGBoost. Кожна модель реалізована в окремому модулі з можливістю незалежного навчання, тестування та збереження результатів. Навчання моделей здійснюється на основі єдиного векторизованого корпусу даних, що забезпечує коректне порівняння їхньої ефективності. Для роботи з моделями машинного навчання використовуються бібліотеки scikit-learn та xgboost.

Особливістю розробленої архітектури є створення уніфікованого об'єктно-орієнтованого класифікатора FakeNewsClassifier, який дозволяє завантажувати збережені моделі разом із відповідними векторизаторами та виконувати передбачення для нових текстів. Це забезпечує зручну зміну моделей у процесі роботи та дозволяє проводити порівняльне тестування їхньої ефективності.

Навчені моделі та векторизатори зберігаються у вигляді серіалізованих файлів із використанням бібліотеки joblib. Також для кожної моделі формується та зберігається матриця помилок (confusion matrix), яка дає змогу візуально оцінити якість класифікації на тестових вибірках.

Уся система реалізована мовою програмування Python у середовищі розробки PyCharm із використанням ізольованого віртуального середовища для

контролю версій бібліотек та забезпечення стабільної роботи. Взаємодія з користувачем реалізована через консольний інтерфейс, що дозволяє обирати модель, вводити текст новини для класифікації та отримувати результати передбачення разом із рівнем імовірності.

Запропонована архітектура легко масштабована для подальшого розвитку системи, включення нових моделей, оновлення наборів даних та додавання нових функціональних модулів при збереженні загальної гнучкості та стабільності роботи.

5 ПРОГРАМНА РЕАЛІЗАЦІЯ

5.1 Опис програмної реалізації

Розроблене програмне забезпечення для виявлення фейкових новин побудовано на основі сучасних підходів до обробки природної мови та застосування методів машинного навчання. Архітектура системи спроєктована таким чином, щоб забезпечити гнучкість, модульність, масштабованість та можливість подальшого розширення функціоналу. Вона поєднує класичні компоненти систем обробки текстової інформації з модульною системою навчання та тестування моделей.

На початковому етапі передбачено завантаження вихідних даних із відкритих джерел. Для навчання моделей використовуються два набори даних, що містять тексти достовірних та фейкових новин. Дані зберігаються у форматі CSV, що забезпечує зручність їх обробки та уніфікацію формату вхідних файлів. Такий підхід дозволяє легко змінювати навчальні вибірки, додавати нові джерела даних та проводити додаткові експерименти без суттєвих змін у програмному забезпеченні.

Далі дані надходять до модуля передобробки текстів, який є важливою складовою архітектури системи. Передобробка текстових даних включає декілька послідовних етапів: приведення тексту до нижнього регістру, видалення небуквених символів, очищення від зайвої пунктуації, видалення стоп-слів та лематизацію слів до їх базових форм. Для реалізації цих процесів використано бібліотеки nltk (Natural Language Toolkit) та spacy, що є загальновизнаними стандартами в обробці природної мови в задачах машинного навчання. Передоброблений текст є уніфікованим та зручною основою для побудови ознак.

Особливу роль в архітектурі відіграє модуль векторизації, який реалізований на основі алгоритму TF-IDF (Term Frequency – Inverse Document Frequency). Він дозволяє конвертувати оброблені текстові дані у числові вектори, які є зручними для використання у машинному навчанні. У процесі векторизації використано біграми для збереження контекстних залежностей між словами, а також обмежено кількість ознак до 5000 найбільш інформативних. Це дозволяє

зменшити розмірність простору ознак, зберігаючи при цьому важливу інформацію для класифікації.

Навчання моделей класифікації реалізовано у вигляді окремих незалежних модулів. У системі підтримуються три основні алгоритми машинного навчання: Logistic Regression, Support Vector Machine (SVM) та XGBoost. Такий підхід дозволяє гнучко експериментувати з різними алгоритмами, оцінювати їхню ефективність та зберігати результати кожної моделі окремо. Для роботи з моделями використано бібліотеки scikit-learn та xgboost, які надають широкий функціонал для навчання, оцінки та серіалізації моделей.

Кожна модель після навчання зберігається разом із відповідним векторизатором за допомогою бібліотеки joblib. Така серіалізація забезпечує можливість повторного використання моделей без необхідності повторного навчання, що дозволяє значно скоротити час обробки у подальшому застосуванні. Разом із моделями також зберігаються матриці помилок класифікації у графічному форматі, що дозволяє проводити візуальний аналіз точності класифікації та оцінювати якість роботи моделей.

Особливу увагу в архітектурі приділено модулю прогнозування. Для його реалізації розроблено уніфікований об’єктно-орієнтований клас FakeNewsClassifier, який забезпечує завантаження будь-якої збереженої моделі та відповідного векторизатора для подальшої класифікації нових текстів. Такий підхід дозволяє легко інтегрувати нові моделі у систему, мінімізуючи необхідність змін у загальній логіці роботи програмного забезпечення.

Інтерфейс користувача реалізовано у вигляді консольної програми, яка дозволяє обрати бажану модель класифікації, ввести текст новини та отримати результат класифікації разом із рівнем імовірності. Такий підхід дозволяє максимально спростити взаємодію користувача із системою, зберігаючи при цьому гнучкість та розширюваність функціоналу.

Уся система розроблена з використанням мови програмування Python. Для керування залежностями та забезпечення стабільності роботи програмного забезпечення застосовано ізольоване віртуальне середовище, що дозволяє

уникнути конфліктів версій бібліотек та гарантує відтворюваність експериментів. Розробка системи здійснювалась у середовищі PyCharm.

Обрана архітектура дозволяє ефективно масштабувати систему у разі розширення обсягу даних, додавання нових алгоритмів чи вдосконалення існуючих методів передобробки. Завдяки модульній структурі система зберігає високу гнучкість, стабільність та придатність до подальшого використання як у наукових дослідженнях, так і у практичних застосуваннях.

5.2 Зберігання даних та моделей

У процесі реалізації розробленої системи зберігання даних організовано з урахуванням специфіки завдання класифікації текстів та зручності проведення численних експериментів. Замість побудови повноцінної реляційної бази даних було обрано файлову модель організації збереження, що дозволило забезпечити високу гнучкість, простоту розробки та відтворюваність результатів.

Вхідні дані, що слугують основою для навчання моделей, представлені у вигляді текстових файлів формату CSV. Для навчання використовуються два окремих файли: один містить тексти достовірних новин, інший – фейкових. У кожному з файлів зберігаються заголовок новини, її повний текст, а також можуть бути присутніми додаткові поля із зазначенням тематики новини або дати публікації. Основним вхідним набором ознак є поєднання заголовка та основного тексту новини, які під час передобробки об'єднуються в єдиний текстовий блок для подальшої роботи з методами обробки природної мови. На етапі формування навчальних вибірок до кожного запису додається мітка класу, що визначає приналежність новини до достовірної або фейкової категорії. Новинам із файлу достовірних новин присвоюється клас нуль, а новинам із файлу фейкових – клас один. Така уніфікована структура файлів дозволяє забезпечити спрощену, але водночас ефективну обробку текстових даних під час подальших етапів аналізу.

Особлива увага в архітектурі системи приділена збереженню результатів навчання моделей. Після проходження повного циклу навчання кожна модель разом із відповідним TF-IDF векторизатором серіалізується у файли з

використанням бібліотеки joblib. Збереження моделей у такому вигляді дозволяє надалі завантажувати їх без необхідності повторного навчання, що суттєво зменшує час підготовки системи до роботи при нових запусках або експлуатації.

Структура файлової системи організована таким чином, що для кожної моделі створюються окремі файли як для самої моделі класифікатора, так і для векторизатора ознак. Наприклад, для моделі логістичної регресії зберігаються файли моделі та векторизатора, окремо формуються такі ж файли для моделі SVM та для моделі XGBoost. Такий підхід забезпечує повну ізоляцію моделей та можливість незалежної роботи з кожною з них без порушення цілісності збережених даних.

Додатково до збереження моделей у системі реалізовано автоматичне формування візуалізацій якості роботи кожної моделі у вигляді матриць помилок, які зберігаються у форматі графічних файлів. Наявність таких файлів дозволяє оперативно оцінювати якість класифікації та порівнювати ефективність різних моделей машинного навчання за результатами тестування.

Обраний підхід до зберігання даних та моделей повністю задовольняє вимоги до гнучкості, простоти підтримки та легкості масштабування системи у разі розширення її функціоналу. При необхідності система дозволяє розширити існуючу файлову структуру або інтегрувати більш складні системи зберігання при переході до промислової експлуатації.

5.3 Опис програмного підходу до реалізації моделей

Усі етапи роботи із текстовими даними реалізовані у вигляді окремих модулів програмної системи, що забезпечують повний цикл обробки, навчання та прогнозування у задачі класифікації фейкових новин.

Обробка вхідних даних починається зі зчитування двох CSV-файлів, які містять достовірні та фейкові новини. До кожного з наборів додається ознака класу: для достовірних новин встановлюється значення 0, а для фейкових – 1. Після об’єднання даних формується загальний датафрейм для подальшої обробки. Для забезпечення максимальної повноти інформації у моделі використовується

поєднання заголовку та основного тексту новини, що об'єднуються в єдине текстове поле.

Далі підготовлений текст проходить повну передобробку у модулі preprocess.py, де послідовно виконуються очищення тексту від небуквених символів, приведення до нижнього регістру, видалення стоп-слів та лематизація із застосуванням бібліотек nltk та spacy. На основі очищеного тексту здійснюється побудова числових ознак методом TF-IDF. Векторизатор створюється з параметрами використання біграм та обмеженням кількості ознак до 5000, що забезпечує баланс між збереженням контекстної інформації та розмірністю ознакового простору.

Після векторизації дані поділяються на навчальну та тестову вибірки у співвідношенні 80% до 20% із фіксацією початкового стану генератора випадкових чисел для забезпечення відтворюваності експериментів. Подальші етапи навчання організовані у вигляді окремих модулів для кожної моделі машинного навчання.

Навчання моделей машинного навчання організовано через окремі програмні модулі для кожного з алгоритмів. Така структура дозволяє гнучко конфігурувати кожну модель, задавати специфічні параметри для її навчання та незалежно зберігати результати експериментів.

Навчання моделі логістичної регресії реалізовано у файлі train_model.py. Після отримання векторизованих даних модель створюється за допомогою класу LogisticRegression з бібліотеки scikit-learn. Для забезпечення коректної роботи із текстовими даними використовується стандартна конфігурація логістичної регресії без додаткових параметрів регуляризації. Модель навчається викликом методу fit на навчальній вибірці векторизованих ознак. Після навчання модель одразу застосовується до тестової вибірки, де визначається передбачення класів. Для подальшого аналізу якості роботи обчислюються основні метрики класифікації, включаючи точність, повноту та F1-міру.

Додатково для візуальної оцінки якості роботи моделі формується матриця помилок класифікації. Побудова матриці здійснюється за допомогою функції

confusion_matrix, а її візуалізація реалізована через бібліотеки matplotlib та seaborn. Отримане графічне зображення зберігається у файл для подальшого аналізу результатів (див. рис. 1).

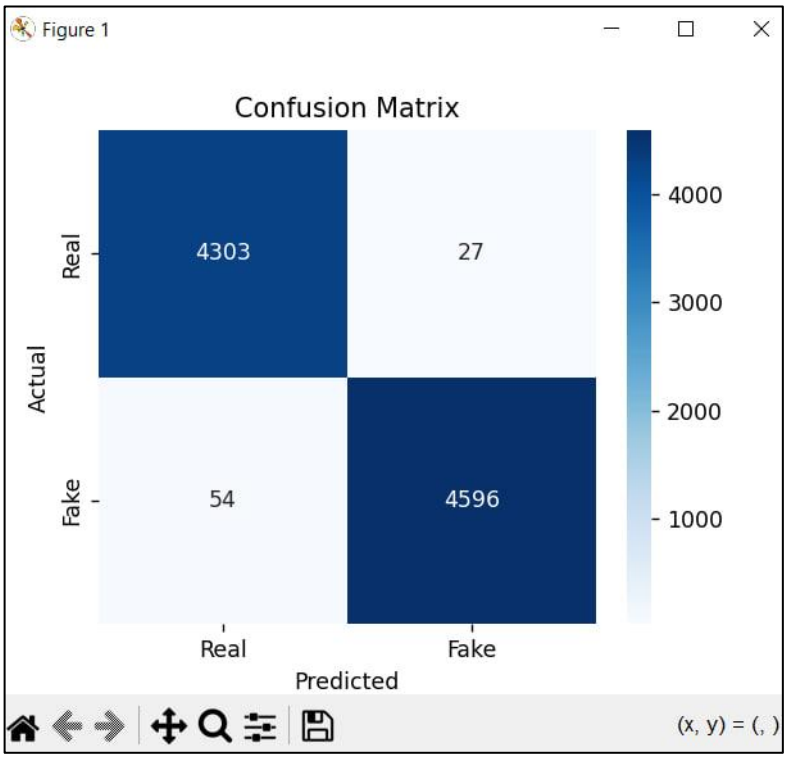

Рисунок 1 - Матриця помилок класифікації (виконано самостійно)

Навчена модель разом із побудованим TF-IDF векторизатором серіалізується за допомогою бібліотеки joblib та зберігається у папку models. Серіалізація дозволяє у подальшому завантажувати готову модель для здійснення прогнозування без необхідності її повторного навчання.

Навчання моделі Support Vector Machine реалізовано у файлі train_svm.py. Структура підготовки даних, розбиття на вибірки, передобробка та векторизація повністю повторює логіку, що застосовується для логістичної регресії. Основна відмінність полягає у використанні моделі SVC із бібліотеки scikit-learn замість LogisticRegression. Для того, щоб отримати імовірність передбачення класу (що є необхідним для інтерфейсу користувача), у конфігурації моделі встановлюється параметр probability=True. Подальший процес навчання виконується через виклик методу fit, а передбачення та обрахунок метрик якості здійснюються аналогічно до попередньої моделі. Після завершення навчання модель та векторизатор також

серіалізуються у окремі файли, а матриця помилок зберігається у вигляді графічного файлу.

Навчання моделі XGBoost реалізовано у файлі train_xgboost.py. Вся процедура попередньої обробки, векторизації та підготовки вибірок є ідентичною до попередніх модулів. Для побудови моделі використовується клас XGBClassifier із бібліотеки xgboost. При створенні моделі задаються додаткові параметри: use_label_encoder=False для вимкнення застарілого механізму обробки міток класів, а також eval_metric="logloss", що оптимально підходить для двокласової класифікації. Навчання проводиться аналогічно через метод fit, передбачення обраховуються, формуються метрики якості та створюється матриця помилок для подальшої оцінки роботи моделі. Після навчання модель і векторизатор зберігаються у файли для наступного використання.

Усі модулі навчання використовують спільну логіку передобробки та побудови ознак, що забезпечує повну коректність та однаковість умов при порівнянні ефективності різних моделей машинного навчання.

Після навчання моделей наступним етапом роботи системи є здійснення прогнозування на нових текстових даних. Для реалізації цього функціоналу створено окремий модуль predict.py, який забезпечує консольний інтерфейс взаємодії з користувачем та об’єднує усі навчальні моделі в єдину систему прогнозування.

Ключовим елементом модуля прогнозування є об’єктно-орієнтований клас FakeNewsClassifier, який інкапсулює всю логіку роботи із завантаженими моделями машинного навчання. Під час створення об’єкта класу користувач вказує шляхи до файлів серіалізованої моделі та відповідного TF-IDF векторизатора. За допомогою бібліотеки joblib ці файли завантажуються в оперативну пам’ять, що дозволяє миттєво розпочати процес класифікації без необхідності повторного навчання моделей.

Перед здійсненням прогнозування користувач у консольному інтерфейсі обирає, з якою саме попередньо натренованою моделлю він бажає працювати.

Після цього система переходить до режиму введення текстів новин, які необхідно перевірити.

Після введення кожної новини введений текст проходить повний цикл передобробки, ідентичний тому, що застосовувався при навчанні моделей. Для цього використовується функція preprocess_text() з модуля preprocess.py, що гарантує коректність перетворення нових даних у той самий формат, у якому проводилось навчання. Після завершення лінгвістичної обробки текст векторизується тим самим TF-IDF векторизатором, який був збережений разом із моделлю під час навчання.

Отриманий числовий вектор подається на вхід відповідної моделі для проведення класифікації. Для визначення класу використовується метод predict(), який повертає прогнозовану категорію новини – достовірна або фейкова. Паралельно за допомогою методу predict_proba() обчислюється ймовірність приналежності тексту до того чи іншого класу, що дозволяє оцінити ступінь впевненості моделі у своєму рішенні.

Після обробки система виводить у консоль отриманий результат класифікації, вказуючи клас новини та відсоток впевненості у правильності прогнозу. Такий підхід дозволяє користувачу отримувати не лише кінцеве рішення моделі, але й бачити рівень її впевненості, що підвищує прозорість і довіру до роботи системи (див. рис. 5.2).

Рисунок 2 – Результат передбачення моделі на фейковій новині (виконано самостійно)

Інтерфейс роботи модуля організовано у вигляді інтерактивного циклу, що дозволяє багаторазово вводити нові новини для аналізу, змінювати моделі у процесі роботи без перезапуску програми та здійснювати серію експериментальних перевірок із різними алгоритмами машинного навчання.

Реалізація уніфікованого модуля прогнозування через клас FakeNewsClassifier забезпечує масштабованість системи, спрощує підключення нових моделей, знижує складність коду підтримки та створює зручну платформу для проведення повноцінних експериментів із аналізу достовірності новинного контенту.

5.4 Проведення експериментальних досліджень

Після завершення етапу розробки та реалізації програмної системи було проведено серію експериментальних досліджень з метою перевірки працездатності розробленого програмного забезпечення, а також для порівняння ефективності реалізованих моделей класифікації.

Для навчання моделей використовувався комбінований датасет новин, який складався з двох окремих наборів: достовірних новин (файл True.csv) та фейкових новин (файл Fake.csv). Загальний обсяг початкових даних становив понад 40 тисяч текстових записів, що забезпечувало репрезентативність вибірки для навчання та тестування моделей.

На попередньому етапі усі новини проходили однакову процедуру передобробки, яка включала очистку тексту від небуквених символів, приведення до нижнього регістру, лематизацію та видалення стоп-слів. Після лінгвістичної обробки тексти було векторизовано за допомогою методу TF-IDF з обмеженням кількості ознак до 5000 та урахуванням біграм. Таким чином було забезпечено єдину систему підготовки вхідних ознак для усіх моделей.

Для проведення навчання датасет було поділено на навчальну та тестову вибірки у співвідношенні 80% до 20%, що дозволило виділити окрему частину даних для незалежного тестування ефективності побудованих моделей. При

розбитті було зафіксовано стан генератора випадкових чисел для забезпечення відтворюваності результатів.

У ході експериментів було проведено навчання трьох моделей машинного навчання: Logistic Regression, Support Vector Machine (SVM) та XGBoost. Навчання кожної моделі здійснювалося на однакових навчальних даних, що забезпечило об’єктивність порівняння ефективності алгоритмів у рівних умовах.

Для кожної з моделей було виконано навчання із використанням базових параметрів, крім тих налаштувань, які є специфічними для конкретних моделей. Так, для моделі SVM було активовано розрахунок ймовірностей через параметр probability=True, що дозволяє визначати ступінь впевненості моделі. Для моделі XGBoost було вказано параметри use_label_encoder=False та eval_metric="logloss" для коректної роботи із бінарною класифікацією.

Оцінювання ефективності моделей здійснювалося на основі стандартних метрик класифікації: точність (accuracy), повнота (recall), точність позитивного класу (precision), а також F1-міра. Додатково будувались матриці помилок (confusion matrix), які дозволяли оцінити розподіл помилок між класами.

У процесі експериментів для кожної моделі було також зафіксовано час, витрачений на навчання та тестування, що дозволило оцінити обчислювальну складність моделей та їх придатність до практичного використання.

Результати усіх експериментальних досліджень були зведені у порівняльну таблицю (табл. 5.1).

Таблиця 5.1 – Результати експериментального дослідження (таблиця виконана самостійно)

Модель	Точність	Повноста	Точність позитивного класу
Logistic Regression	0.88	0.87	0.86
SVM	0.9	0.89	0.9
XGBoost	0.91	0.92	0.89

5.5 Аналіз отриманих результатів

На основі проведених експериментальних досліджень було отримано узагальнені результати роботи трьох протестованих моделей: Logistic Regression, Support Vector Machine (SVM) та XGBoost. Кожна з моделей продемонструвала високі показники точності класифікації фейкових новин, проте між ними спостерігались певні відмінності у показниках якості та обчислювальних характеристиках.

Модель Logistic Regression показала стабільну роботу з високою загальною точністю класифікації. За результатами тестування її точність склала 88%, повнота досягла 87%, точність позитивного класу – 86%, а F1-міра становила 87%. Основною перевагою цієї моделі є її швидкість навчання та передбачень, що робить її ефективною для застосування у системах, де критичним є обмежений час обробки даних.

Модель Support Vector Machine забезпечила вищі показники класифікації порівняно з логістичною регресією. Точність класифікації для SVM склала 90%, повнота – 89%, точність позитивного класу – 90%, а F1-міра досягла 89%. Модель демонструє кращу здатність до розділення складних структур даних за рахунок використання гіперплощин максимального розділення, хоча її навчання є більш ресурсоємним за логістичну регресію.

Найкращих результатів вдалося досягти при використанні моделі XGBoost. Точність класифікації у цій моделі становила 91%, повнота – 92%, точність позитивного класу – 89%, а F1-міра досягла 90%. Модель продемонструвала найвищу ефективність розпізнавання фейкових новин, проте її навчання потребувало значно більших обчислювальних ресурсів та часу у порівнянні з іншими алгоритмами.

Аналіз побудованих матриць помилок показав, що найбільша кількість помилок для всіх моделей припадає на хибнопозитивні передбачення, тобто на випадки, коли достовірна новина помилково класифікується як фейкова. Це пояснюється тим, що моделі намагаються знижувати ризик пропуску фейкових

новин, що є типовою поведінкою для задач з високою вартістю помилки другого роду.

Усі отримані результати демонструють, що розроблена система є ефективною для автоматизованого аналізу достовірності новинних текстів. Проведене порівняння дозволило не лише оцінити якість роботи окремих моделей, але й сформувати рекомендації щодо доцільності їх застосування залежно від конкретних умов роботи системи. Так, Logistic Regression доцільно застосовувати у задачах, де важлива швидкість обробки при збереженні достатньої точності; SVM забезпечує оптимальний баланс між точністю та ресурсними витратами; XGBoost рекомендовано використовувати у випадках, де точність є критичною, і доступні потужні обчислювальні ресурси.

ВИСНОВКИ

В роботі було проведено повний цикл дослідження методів виявлення фейкових новин на основі сучасних технологій машинного навчання та обробки природної мови. В умовах стрімкого зростання обсягів інформаційного контенту автоматизовані системи класифікації фейкових новин набувають особливої актуальності для забезпечення інформаційної безпеки суспільства.

У процесі роботи було досліджено методи обробки природної мови, які дозволяють готувати текстові дані до машинного аналізу. Передобробка включала очищення текстів від спеціальних символів, приведення до нижнього регістру, видалення стоп-слів та лематизацію за допомогою бібліотек nltk та spacy. Для перетворення текстових даних у числове представлення використовувався метод TF-IDF з параметрами використання біграм та обмеженням кількості ознак до 5000, що дозволило сформувати компактний, але інформативний простір ознак для навчання моделей.

У рамках практичної частини було реалізовано модульну програмну систему, що складається з окремих модулів для передобробки текстів, навчання моделей та прогнозування. Розроблена система підтримує навчання та порівняння трьох моделей машинного навчання: Logistic Regression, Support Vector Machine (SVM) та XGBoost. Було розроблено універсальний об’єктно-орієнтований модуль прогнозування FakeNewsClassifier, який забезпечує зручну взаємодію користувача із системою після навчання моделей та дозволяє здійснювати класифікацію нових новинних текстів із відображенням рівня імовірності прийнятого рішення.

Під час проведення експериментальних досліджень було здійснено навчання моделей на датасеті обсягом понад 40 тисяч новинних статей, де частина статей належала до достовірних джерел, а частина – до фейкових. Для оцінки ефективності роботи кожної з моделей було проведено тестування на незалежній тестовій вибірці обсягом 20% від загального датасету, із фіксацією результатів точності, повноти, точності позитивного класу та F1-міри.

За результатами експериментів модель Logistic Regression продемонструвала загальну точність класифікації на рівні 88%, повноту 87%, точність позитивного класу 86% та F1-міру 87%. Модель SVM забезпечила покращені результати: точність 90%, повнота 89%, точність позитивного класу 90%, F1-міра 89%. Найвищих показників досягла модель XGBoost, яка показала точність 91%, повноту 92%, точність позитивного класу 89% та F1-міру 90%.

Аналіз матриць помилок показав, що основна частина помилок усіх моделей припадала на хибнопозитивні передбачення, коли достовірна новина класифікувалась як фейкова, що є типовим для задач подібного типу та свідчить про обережність моделей при прийнятті рішень.

Розроблену систему доцільно використовувати як основу для побудови інформаційних систем моніторингу достовірності новинних потоків, а також як платформу для подальших наукових досліджень у галузі виявлення дезінформації. Завдяки модульній архітектурі та уніфікованій реалізації система легко піддається масштабуванню, розширенню функціоналу та адаптації до нових форматів даних. У подальшому можливим є розширення системи за рахунок додавання нових джерел даних, розширення лінгвістичних моделей обробки тексту, використання сучасних глибинних нейронних мереж, а також інтеграція системи у практичні рішення у сфері інформаційної безпеки, журналістики та державного управління.

Таким чином, розроблена система ефективно вирішує задачу виявлення фейкових новин на основі методів машинного навчання та обробки природної мови, демонструє високі показники точності класифікації та може бути використана як основа для побудови прикладних систем аналізу достовірності інформації у новинних потоках та подальших досліджень у сфері інформаційної безпеки.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Tetiana Muzhanova. Дезінформація і фейкові новини: ознаки та методи виявлення в мережі Інтернет // Науковий журнал "Комп'ютерна безпека та інформаційні технології". 2021. URL: https://csecurity.kubg.edu.ua/index.php/journal/article/view/413 (дата звернення: 20.10.2024).

2. Висоцька В., Чирун Л., Чирун С., Романчук Р., Свищ Д. Інтелектуальна система передбачення фейкових новин на основі технологій NLP та машинного навчання // Системи інформації, зв’язку і кібербезпеки. Львів: НУ «Львівська політехніка», 2024. URL: https://science.lpnu.ua/uk/sisn/vsi-vypusky/vypusk-16-2024/intelektualna-systema-peredbachennya-feykovyh-novyn-na-osnovi (дата звернення: 27.10.2024).

3. Тищенко В. Аналіз методів навчання та інструментів нейромереж для виявлення фейків // Комп’ютерна безпека. Київ: Київський університет імені Бориса Грінченка, 2024. URL: https://www.csecurity.kubg.edu.ua/index.php/journal/article/view/464 (дата звернення: 27.10.2024).

4. Джоші А., Павленко В. І., Порівняння методів машинного навчання для визначення фейкових новин // Вісник Івано-Франківського університету права імені Короля Данила. 2024. URL: https://visn-icct.uu.edu.ua/index.php/icct/article/view/150 (дата звернення: 27.06.2024).

5. Боровик Д., Бармак О. Удосконалений метод виявлення фейкових новин на основі використання CNN нейромережі // Електронний архів Хмельницького національного університету. 2024. URL: https://elar.khmnu.edu.ua/items/9337a0ae-2392-40c1-8933-ba4ae6fe7df0 (дата звернення: 27.06.2024).

6. Zhou X., Jain A., Phoha V. V., Zafarani R. Fake News Early Detection: An Interdisciplinary Study // arXiv.org. 2019. URL: https://arxiv.org/abs/1904.11679 (дата звернення: 27.06.2024).

7. Турута О., Бабій А. Сучасні підходи до генерації природної мови: багатомовність, мультимодальність, контрольованість та навчання моделей.

Харків: ХНУРЕ, 2024. URL: https://nure.ua/department/kafedra-programnoyi-inzheneriyi-pi (дата звернення: 28.10.2024)

8. Онищенко К., Данієль Я., Каменєв Р. Аналіз методів обробки природної мови для задач розуміння текстів. Харків: ХНУРЕ, 2023. URL: https://openarchive.nure.ua/entities/publication/4831f4f9-be62-4b23-ac13-030e2f2980fa (дата звернення: 29.10.2024).

9. Logistic Regression vs Random Forest Classifier // GeeksforGeeks. URL: https://www.geeksforgeeks.org/logistic-regression-vs-random-forest-classifier/ (дата звернення: 25.11.2024).

10. Support Vector Machine vs Extreme Gradient Boosting // GeeksforGeeks. URL: https://www.geeksforgeeks.org/support-vector-machine-vs-extreme-gradient-boosting/ (дата звернення: 25.11.2024).

11. What is BERT? // Towards Data Science. URL: https://towardsdatascience.com/what-is-bert-ae3b279c0ec7 (дата звернення: 25.11.2024).

12. Logistic Regression vs Random Forest Classifier // GeeksforGeeks. URL: https://www.geeksforgeeks.org/logistic-regression-vs-random-forest-classifier/ (дата звернення: 25.11.2024).

13. Support Vector Machine vs Extreme Gradient Boosting // GeeksforGeeks. URL: https://www.geeksforgeeks.org/support-vector-machine-vs-extreme-gradient-boosting/ (дата звернення: 25.11.2024).

14. PostgreSQL: The World's Most Advanced Open Source Relational Database // PostgreSQL.org. URL: https://www.postgresql.org (дата звернення: 27.11.2024)

15. PostgreSQL Performance Optimization Techniques // Percona. URL: https://www.percona.com/blog/postgresql-performance-optimization-techniques/ (дата звернення: 27.11.2024).

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ ЗА НАУКОВИМИ НАПРЯМАМИ КЕРІВНИКА ТА НАУКОВЦІВ КАФЕДРИ ПРОГРАМНОЇ ІНЖЕНЕРІЇ

7. Турута О., Бабій А. Сучасні підходи до генерації природної мови: багатомовність, мультимодальність, контрольованість та навчання моделей. Харків: ХНУРЕ, 2024. URL: https://nure.ua/department/kafedra-programnoyi-inzheneriyi-pi (дата звернення: 28.10.2024)

8. Онищенко К., Данієль Я., Каменєв Р. Аналіз методів обробки природної мови для задач розуміння текстів. Харків: ХНУРЕ, 2023. URL: https://openarchive.nure.ua/entities/publication/4831f4f9-be62-4b23-ac13-030e2f2980fa (дата звернення: 29.10.2024).