МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ВЕТЕРИНАРНОЇ МЕДИЦИНИ ТА БІОТЕХНОЛОГІЙ ІМЕНІ С.З. ҐЖИЦЬКОГО
ФАКУЛЬТЕТ МЕХАНІКИ, ЕНЕРГЕТИКИ ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

КВАЛІФІКАЦІЙНА РОБОТА

другого (магістерського) рівня вищої освіти

на тему:

«ВИЯВЛЕННЯ ФЕЙКОВИХ НОВИН НА ОСНОВІ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ»

Виконав: студент 6 курсу
спеціальності 126
«Інформаційні системи та технологій»

Курило В.Б.
(прізвище та ініціали)

Керівник: Желєзняк А.М.
(прізвище та ініціали)

ДУБЛЯНИ 2025

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ВЕТЕРИНАРНОЇ МЕДИЦИНИ ТА БІОТЕХНОЛОГІЙ ІМЕНІ С.З. ҐЖИЦЬКОГО
ФАКУЛЬТЕТ МЕХАНІКИ, ЕНЕРГЕТИКИ ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

Рівень вищої освіти другий (магістерський)
Спеціальність 126 «Інформаційні системи та технології»

ЗАТВЕРДЖУЮ
Завідувач кафедри

_____
(підпис)
д.т.н., професор, Тригуба А. М.
(вч. звання, прізвище, ініціали
“____” _____2025
року

З А В Д А Н Н Я
НА КВАЛІФІКАЦІЙНУ РОБОТУ

Курило Валентину Богдановичу
(прізвище, ім’я, по батькові)

1. Тема роботи *«Виявлення фейкових новин на основі моделей машинного навчання»*
керівник роботи *к. е н., доцент., Желєзняк А.М.*
(наук.ступінь, вч. звання, прізвище, ініціали)

затверджені наказом ЛНУВМБТ ім.С.З.Гжицького №140/к-с від 28.02.2025

2. Строк подання студентом роботи 05.12.2025 р.
3. Вихідні дані: *аналітичні дані роботи та характеристика об’єкту дослідження, опис бібліотек мов програмування, науково-технічна і довідкова література.*
4. Зміст кваліфікаційної роботи (перелік питань, які потрібно розробити)
Вступ
1. Аналіз стану питання в теорії та практиці
2. Обґрунтування, вибір та реалізація інструментарію вирішення задачі
3. Результати вирішення задачі
4. Охорона праці та безпека в надзвичайних ситуаціях
5. Визначення ефективності
Висновки
Бібліографічний список
5. Перелік графічного матеріалу
Графічний матеріал подається у вигляді презентації

6. Консультанти розділів

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата		Відмітка про виконання
		завдання видав	завдання прийняв	
1, 2, 3, 5	*Желєзняк А.М., к.е.н., доцент кафедри інформаційних технологій*			
4	*Городецький І.М., к.т.н., доцент кафедри інженерної механіки*			

7. Дата видачі завдання 01.03.2025 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломної роботи	Строк виконання етапів роботи	Відмітка про виконання
1	*Отримання завдання. Вивчення літератури по темі роботи. Написання першого розділу*	*01.03.2025 – 31.05.2025*	
2	*Проектування та опис технічного завдання, обґрунтування та вибір інструментарію реалізації проекту (написання другого розділу).*	*01.06.2025 – 31.08.2025*	
3	*Програмна реалізація поставленого завдання (написання третього розділу)*	*01.09.2025 – 17.10.2025*	
4	*Написання розділу з охорони праці*	*18.10.2025 – 04.11.2025*	
5	*Оцінка ефективності поставленого завдання (виконання п'ятого розділу)*	*05.11.2025 – 18.11.2025*	
6	*Завершення оформлення основної частини, написання висновків та підготовка презентаційного матеріалу*	*19.11.2025 – 01.12.2025*	
7	*Завершення роботи в цілому. Підготовка до захисту кваліфікаційної роботи*	*02.12.2025 – 08.12.2025*	

Студент _____ *Курило В.Б.*
(підпис) (прізвище та ініціали)

Керівник роботи _____ *Желєзняк А.М.*
(підпис) (прізвище та ініціали)

УДК 004.8:004.912:316.774

Виявлення фейкових новин на основі моделей машинного навчання.

Курило Валентин Богданович. Кваліфікаційна робота. Кафедра інформаційних технологій. Дубляни, Львівський національний університет ветеринарної медицини та біотехнологій ім.С.З.Гжицького, 2025 р.

Кваліфікаційна робота: 81 сторінок текстової частини, 7 таблиць, 27 рисунків, 31 джерел літератури, 2 додатки.

Подано характеристику проблеми виявлення фейкових новин та проаналізовано актуальність теми в умовах сучасного інформаційного середовища. Розглянуто предметну область, досліджено існуючі підходи та алгоритми, а також визначено функціональні вимоги до системи автоматизованого виявлення фейкових новин. Здійснено проектування архітектури веб-застосунку, серверної частини та бази даних відповідно до сучасних методик і технологій розробки.

Запропоновано моделі, що включають методи машинного навчання та глибинного навчання для підвищення точності класифікації новин. Реалізовано повний цикл програмної системи від збору та обробки даних до побудови моделі та інтеграції її у веб-інтерфейс.

Здійснено аналіз потенційних травмонебезпечних ситуацій під час роботи з комп’ютерною технікою та наведено основні вимоги з охорони праці й безпеки в надзвичайних ситуаціях.

Ключові слова: фейкові новини, машинне навчання, веб-застосунок, класифікація тексту.

ЗМІСТ

ВСТУП

У наші дні в які активно розвиваються інформаційні технології та обмін даними відбувається миттєво, а соціальні мережі та онлайн-медіа стали основними джерелами новин, проблема поширення фейкових новин набуває дедалі більшої актуальності. Фейкові новини - це навмисно створена або перекручена інформація, яка вводить в оману користувачів з метою маніпулювання громадською думкою, політичного чи економічного впливу, або отримання фінансової вигоди. Масове розповсюдження таких новин ускладнює доступ до достовірної інформації, зменшує довіру до медіа. В умовах інформаційної війни, зростання кількості ботів, а також широкого використання соціальних платформ, виявлення та протидія фейковим новинам стало критично важливим завданням не лише для журналістики, але й для кібербезпеки та державного управління.

Традиційні методи боротьби з фейковими новинами, які базуються на ручному аналізі контенту, не здатні забезпечити необхідну швидкість і масштаб обробки інформації. Тому все більшої популярності набувають автоматизовані системи, що використовують методи машинного навчання та обробки природної мови (NLP) для аналізу текстових повідомлень, виявлення фейкових новин і оцінки їх достовірності. Такі системи дозволяють аналізувати великі обсяги даних, виявляти приховані закономірності та формулювати висновки на основі статистичних моделей.

Вибір теми кваліфікаційної роботи зумовлений зростаючою актуальністю проблеми інформаційної безпеки, а також необхідністю створення інтелектуальних систем для автоматичного аналізу медіа-контенту.

Метою роботи є розробка та дослідження ефективних моделей машинного навчання для автоматичного виявлення фейкових новин на основі аналізу текстових даних.

Для досягнення мети було поставлено такі завдання:

✔ Провести аналіз предметної області та існуючих методів виявлення фейкових новин;

✔ Дослідити сучасні наукові підходи та моделі обробки природної мови, зокрема трансформерні архітектури;

✔ Зібрати, обробити та підготувати набір даних новин для навчання моделей;

✔ Реалізувати та порівняти ефективність різних моделей машинного навчання (логістична регресія, SVC, LSTM, BERT та ін.);

✔ Проаналізувати отримані результати використовуючи різні метрики (Accuracy, Precision, Recall, F1-score);

✔ Запропонувати рекомендації щодо інтеграції та просту реалізацію системи автоматичного виявлення фейкових новин у практичні інформаційні сервіси.

Об’єктом дослідження є процес автоматизованого виявлення фейкових новин у текстових джерелах за допомогою моделей машинного навчання. Предметом дослідження є методи та моделі машинного навчання, що використовуються для класифікації новин.

Актуальність теми полягає у швидкому зростанні кількості інформаційних маніпуляцій у медіапросторі та необхідності створення надійних інтелектуальних систем, здатних автоматично виявляти фейковий контент. Застосування машинного навчання у цій сфері забезпечує масштабованість, адаптивність до нових джерел даних та підвищення точності і швидкості виявлення фейкових повідомлень.

Наукова новизна магістерської роботи полягає в удосконаленні підходу до виявлення фейкових новин із використанням сучасних моделей машинного навчання та методів обробки природної мови (NLP), який забезпечує підвищення точності класифікації текстового контенту при перенавчанні моделей на основі даних, зібраних із реальних

користувацьких текстів з розробленого інтерфейсу для цих моделей, що попередньо пройшли ручну експертну перевірку.

Результати дослідження апробовано на VII Всеукраїнській науково-практичній конференції «Інформаційна безпека та інформаційні технології» ІБІТ 2025 (27 листопада 2025 р., м.Львів) із публікацією тез.

Практична цінність дослідження полягає у створенні програмної системи, здатної автоматично класифікувати новини на достовірні та фейкові, використовуючи сучасні алгоритми машинного навчання. Результати роботи можуть бути використані у системах моніторингу новин, аналітичних платформах, соціальних мережах та медіа-агенціях для підвищення інформаційної безпеки та боротьби з дезінформацією.

У наступних розділах магістерської роботи розглянуто теоретичні засади проблеми виявлення фейкових новин, здійснено ґрунтовний аналіз наукових джерел та сучасних досліджень у цій сфері, проведено огляд і порівняння методів машинного навчання, що застосовуються для обробки природної мови (NLP). Наведено детальний опис етапів підготовки та попередньої обробки вибраного набору даних, розробки та навчання моделей, представлено результати експериментів із використанням різних алгоритмів машинного навчання. Описано створений веб-застосунок, узагальнено висновки та окреслено перспективи подальших досліджень.

Результати цієї роботи сприятимуть розвитку напрямку інтелектуального аналізу текстових даних та покращенню якості автоматизованої модерації контенту. З наукової точки зору, дослідження сприятиме поглибленню розуміння процесів обробки природної мови та удосконаленню підходів до класифікації текстових даних. З практичної дозволить підвищити довіру користувачів до онлайн-джерел інформації та зменшити вплив фейкових новин на суспільну свідомість.

РОЗДІЛ 1.
АНАЛІЗ СТАНУ ПИТАННЯ В ТЕОРІЇ ТА ПРАКТИЦІ

1.1 Аналіз актуальності виявлення фейкових новин

Проблема поширення фейкових новин у мережі Інтернет набуває все більшої актуальності через її значний вплив на суспільну свідомість, формування громадської думки та інформаційну безпеку. У добу цифровізації та активного розвитку соціальних медіа (Telegram, Facebook, X (Twitter), Instagram тощо) користувачі щодня стикаються з великим обсягом контенту, достовірність якого часто залишається сумнівною. Це явище має глобальний характер, оскільки стосується більшості користувачів Інтернету по всьому світу. Згідно з даними звіту “Social Media Usage and Growth Statistics” [1], кількість користувачів соціальних мереж постійно зростає і у 2025 році сягнула 5.24 мільярдів осіб, що є більше ніж в два рази в порівнянні з 2015 роком, що підкреслює масштабність проблеми дезінформації та актуальність розробки ефективних систем для автоматичного виявлення фейкових новин.

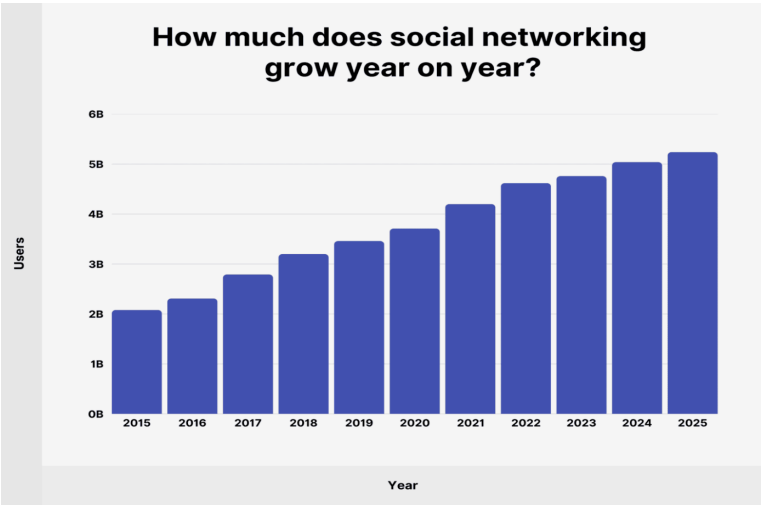
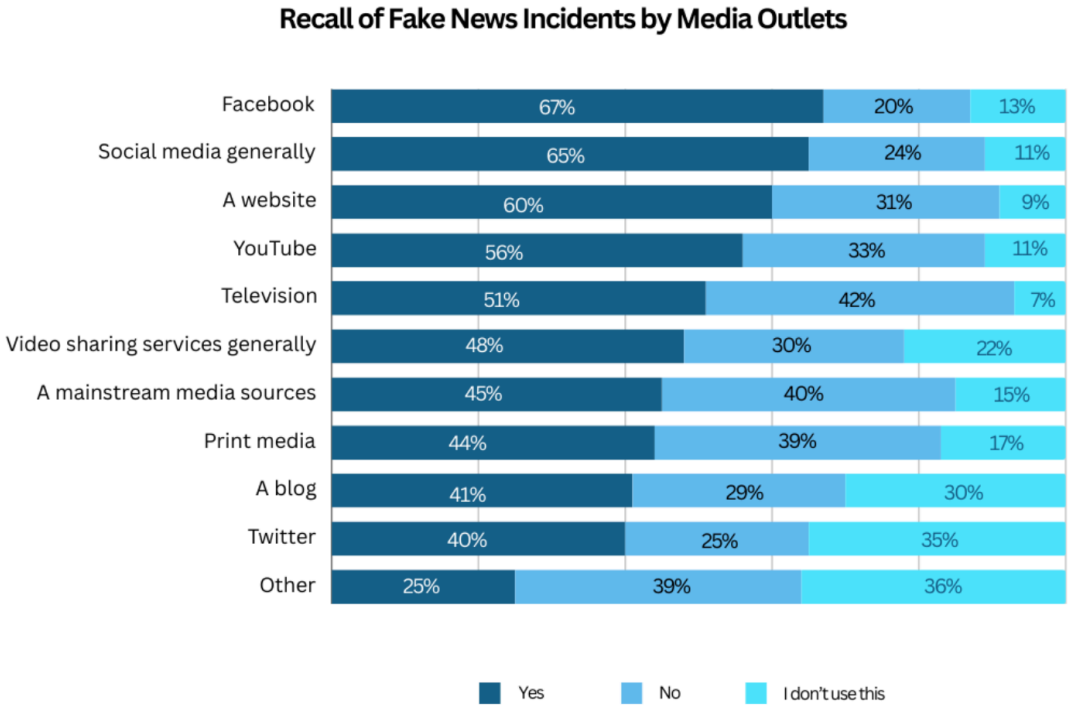


Рисунок 1.1 - Статистика кількості користувачів соціальних мереж 2015-2025 рр. [1]

У роботі “Ключова статистика щодо фейкових новин та дезінформації в ЗМІ у 2024 році” [2] показано, що люди найбільше стикається з неправдивими новинами саму у соціальних мережах, що можна побачити на рисунку 1.2.



Рисунку 1.2 - Статистика по медіа, де користувачі стикалися з неправдивими новинами 2024 р. [2]

Як бачимо з рисунку №1.2 саме більше люди, які брали участь в опитуванні стикалися з неправдивими новинами у соціальній мережі Facebook, а саме 67% користувачів, які хоча б раз бачили дезінформацію там, коли тільки 20% відповіли навпаки, а також 65% опитуваних людей відповіли, що бачили неправдиві новини в загальному у соціальних мережах. Дане опитування показує, що у нас час саме більше неправдивих новин поширюються у соціальних мережах і більшість користувачі розуміють це і відповідно втрачається довіра користувачів до контенту, який шириться соціальними мережами та відповідно з цим зменшуються рейтинги та бажання користуватися певними мережами де ширяться

фейкові новини. Тому потрібно боротися з цією проблемою і ця магістерська робота демонструє приклад, який може допомогти зменшити кількість фейкових новин у соціальних мережах.

Також у статті від “Statista” “Споживачі, які стикаються з неправдивою інформацією з певних тем у всьому світі у 2024 році” [3] показано, що за зиму 2024 року приблизно третина новин була неправдивими, що візуально показано на рисунку №1.3

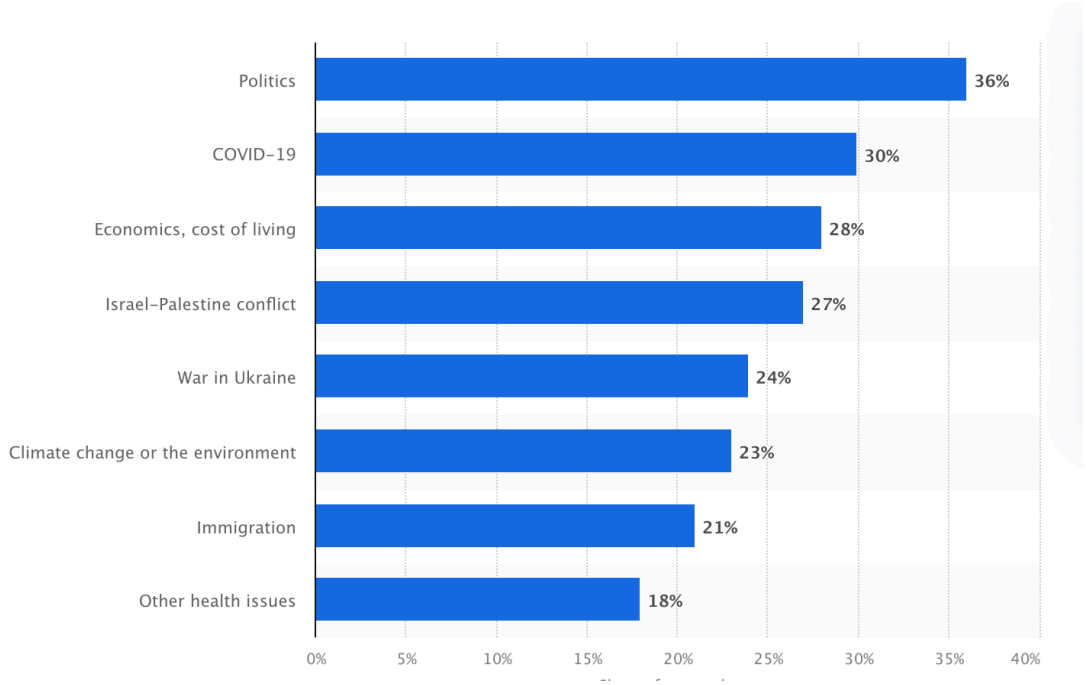


Рисунок 1.3 - Відсоток неправдивих новин, які бачили користувачі за зиму 2024 р. [3]

З цього рисунку, ми можемо побачити, що саме більше неправдивих новин поширювалося про політику, а саме 36%, що є більше ніж третини з усіх новин, також майже третина неправдивих новин поширювалася про Covid-19. Та як можемо побачити, що також, поширювалося багато неправдивих новин про війну в Україні по всьому світі, а саме четвертина з усіх новин. Враховуючи ситуацію в нашій країні для нас є дуже важливим,

щоб світ бачив правду про події в Україні і тому це показує, що ця тема магістерської роботи є дуже актуальною для нас в цей період часу.

Проведений аналіз показав, що проблема поширення фейкових новин у соціальних мережах є однією з найактуальніших загроз інформаційного простору. Стрімке зростання кількості користувачів соціальних платформ сприяє масовому розповсюдженню недостовірної інформації, яка впливає на формування громадської думки, рівень довіри до медіа та загальну інформаційну безпеку суспільства. Результати досліджень, наведених у статистичних звітах, свідчать про те, що більшість користувачів найчастіше стикаються з неправдивими новинами саме у соціальних мережах, таких як Facebook, X (Twitter) та Telegram. Найбільш поширеними темами дезінформації залишаються політичні події та війна в Україні.

1.2. Огляд існуючих рішень

В сьогоднішні дні, де стрімко поширюється дезінформація та зростає попит на перевірку джерел даних, а особливо на контент, який поширюється в соціальних мережах. Дедалі більше організацій, наукових груп та технологічних компаній розробляють системи та страються їх покращувати для виявлення фейкових новин. Аналіз існуючих рішень дозволяє визначити поточний рівень розвитку цієї галузі, виявити їхні переваги, недоліки та окреслити напрями вдосконалення власної системи.

З метою систематизації інформації та кращого розуміння відмінностей між існуючими інструментами виявлення фейкових новин, у таблиці 1.1 представлено порівняння їхніх основних переваг і недоліків, а також окреслено ключові переваги підходу, запропонованого в рамках даного дослідження.

Таблиця 1.1. - Переваги та недоліки аналогів

Платформа	Переваги	Недоліки	Переваги запропонованого підходу
PolitiFact	1.Висока точність завдяки ручній експертній перевірці 2.Зручний інтерфейс 3.Авторитетність та висока довіра у медіасфері	1.Низька масштабованість через трудомісткі процеси 2.Потребує значного людського втручання 3.Неможливість обробки великих обсягів інформації в реальному часі	Автоматизація процесу виявлення фейків за допомогою ML-моделей забезпечує високу швидкість, масштабованість та роботу в реальному часі без постійної участі експертів
Snopes	1.Велика база перевірених новин 2.Надійність результатів 3.Популярність серед англомовних користувачів	1.Відсутність автоматизації 2.Орієнтація лише на англомовний контент 3.Обмежена глобальна застосовність	Орієнтація на україномовний контент і врахування лінгвістичних особливостей української мови, що робить систему релевантною для українського інформаційного простору
Google Fact Check Tools	1.Часткова автоматизація 2.Інтеграція з екосистемою Google 3.Зручний інтерфейс і наявність API	1.Низька можливість кастомізації 2.Закритий код, обмеження для досліджень 3.Недостатня підтримка україномовних джерел	Гнучка адаптація під конкретні потреби, можливість розширення моделі та орієнтація на українські джерела, що забезпечує високу точність у місцевому інформаційному середовищі

Останніми роками з’являються численні дослідницькі розробки, що базуються на машинному навчанні та глибоких нейронних мережах (наприклад, FakeNewsNet, LIAR dataset, BERT-based Fake News Classifiers). Такі моделі здійснюють автоматичну класифікацію новин на достовірні та фейкові, використовуючи текстовий аналіз, метадані або поведінкові характеристики користувачів. Перевагами таких підходів є масштабованість, можливість роботи в реальному часі, адаптивність до нових даних. Недоліками є потреба у великих наборах розмічених даних,

складність інтерпретації результатів, ризик зниження точності при перенавчання моделей на певних даних, що означатиме те, що моделі добре працюють лише на певних даних на яких проводилося навчання, а от на нових даних модель будуть давати не точні результати.

Аналіз існуючих рішень показав, що більшість платформ для виявлення фейкових новин базуються на ручній перевірці фактів або частковій автоматизації процесу. Незважаючи на високу точність, ці методи не забезпечують необхідної швидкості обробки даних і не придатні для масштабного моніторингу соціальних мереж у режимі реального часу.

Використання моделей машинного навчання та обробки природної мови (NLP) дає змогу створити адаптивну систему, здатну самостійно аналізувати великий обсяг контенту та виявляти потенційно неправдиві публікації. Таким чином, запропонований у цій магістерській роботі підхід відрізняється більшою гнучкістю, масштабованістю та потенціалом для автоматичного аналізу україномовних джерел.

Таким чином, розроблення та впровадження ефективних інтелектуальних систем для автоматичного виявлення фейкових новин на основі моделей машинного навчання є надзвичайно важливим напрямом сучасних наукових досліджень. Це дозволить зменшити вплив дезінформації, підвищити рівень об’єктивності у сприйнятті новинного контенту та сприятиме зміцненню інформаційної безпеки суспільства.

1.3 Аналіз літературних джерел та наукових досліджень

Проведений аналіз показує, що виявлення фейкових новин у соціальних мережах залишається однією з найбільш актуальних проблем сучасної науки і практики. У світі, де інформація поширюється дуже швидко, важливість достовірного контенту значно зростає, особливо для формування об’єктивної громадської думки та підтримки інформаційної безпеки. Тому в цьому підрозділі узагальнено дані щодо тематики

автоматичного виявлення неправдивих новин, представлені в останніх публікаціях та статистичних звітах.

За даними “Stanford HAI / AI Index Report 2025” [4], загальна кількість публікацій з тематики штучного інтелекту зросла з приблизно 100-110 тисяч у середині 2010-х до більше ніж 240-250 тисяч у 2022–2023 роках. Це свідчить про надзвичайно швидке збільшення наукового інтересу до AI, що включає й методи для виявлення дезінформації. Дослідження “A Scientometric Analysis of Deep Learning Approaches for Detecting Fake News” [5] показало, що за період з 2012 до 2022 року опубліковано 569 робіт, які безпосередньо присвячені методам глибинного навчання для виявлення фейкових новин.

Окрім цього, аналіз “Dataset of Fake News Detection and Fact Verification: A Survey” показує, що на сьогодні існує понад 118 публічно доступних наборів даних (datasets), призначених для задач виявлення фейкових новин, перевірки фактів та суміжних задач (наприклад, сатиричні новини) [6]. Це говорить про те, що в дослідницькому середовищі існує значна інфраструктура для роботи з подібним контентом, але великою мірою вона орієнтована на англомовні тексти або міжнародні ресурси.

Щодо україномовного контенту, статистика менш обширна. Наприклад, у дослідженнях USAID-Internews [7], що стосуються України, лише 14 % українців у 2022 році змогли повністю відрізнити правдиві повідомлення від фейкових, коли їм пропонували оцінити три твердження. Це свідчить про низький рівень медіаграмотності у частини населення і підкреслює важливість систем автоматичного виявлення неправдивих новин саме для україномовного середовища.

У процесі аналізу наукових джерел, присвячених проблемі виявлення фейкових новин, були відібрані найактуальніші та найрезультативніші дослідження, що пропонують сучасні методи машинного та глибинного

навчання для підвищення точності класифікації новинного контенту. Розглянуті роботи охоплюють як традиційні алгоритми машинного навчання, так і новітні нейронні архітектури, включно з моделями на основі трансформерів. Під час огляду було проаналізовано основні характеристики кожного джерела: назву наукової праці, мету дослідження, застосовані методи, рік публікації та бібліографічне посилання. Узагальнені результати аналізу представлено у Додатку А.

Дослідження [5] представляє собою бібліометричний аналіз наукових робіт, присвячених виявленню фейкових новин із використанням методів глибинного навчання (Deep Learning) у період з 2012 по середину 2022 року. Автори використали інструменти Bibliometrix (R-package) та VOSviewer для збору, обробки та візуалізації даних, що дало змогу визначити основні наукові тенденції, країни-лідери, популярні джерела публікацій, а також тематичні напрямки досліджень у сфері протидії дезінформації.

Результати показали, що активна фаза розвитку досліджень у цій галузі розпочалася після президентських виборів у США 2016 року, коли термін “fake news” набув глобального значення. Особливо інтенсивне зростання кількості наукових робіт спостерігалося після пандемії COVID-19, коли в мережі різко збільшився обсяг інформації та дезінформації. Найбільшу кількість публікацій здійснили науковці з Індії, тоді як серед видань лідирував журнал IEEE Access, а також ACM International Conference Proceedings Series. Аналіз ключових термінів показав, що найчастіше в наукових роботах використовувалися поняття deep learning, fake news detection, social media, natural language processing (NLP), LSTM та deep fake. Це свідчить про міждисциплінарний характер досліджень, що поєднують методи машинного навчання, лінгвістики та аналізу соціальних медіа.Автори відзначили, що галузь дослідження ще перебуває на етапі активного становлення, однак демонструє стабільне

зростання інтересу науковців у всьому світі. Зокрема, за 10 років було проаналізовано 569 публікацій, що охоплюють найрізноманітніші підходи до автоматичного виявлення неправдивої інформації. Серед обмежень дослідження зазначено, що аналіз проводився лише на основі бази даних Scopus, без врахування інших наукометричних ресурсів, таких як Web of Science або Science Citation Index. Також автори підкреслюють, що застосування лише кількісного (бібліометричного) підходу не дає повного розуміння якості робіт, тому подальші дослідження мають включати якісний контент-аналіз публікацій.

Перспективним напрямом, за результатами роботи, є використання технологій пояснюваного штучного інтелекту (Explainable AI, XAI) для підвищення прозорості та довіри до моделей глибинного навчання, які використовуються у виявленні фейкових новин.

Таким чином, дана стаття робить вагомий внесок у систематизацію знань про розвиток наукових досліджень у сфері автоматизованого виявлення фейкових новин, показує глобальні тенденції та вказує на подальші напрями розвитку, а саме інтеграцію XAI, розширення мовного охоплення та залучення міждисциплінарних підходів.

Також розглянемо переваги та недоліку кожного з джерел, де використовувалися різноманітіні методи машинного навчання для виявлення фейкових новин, які розглядалися у даній роботі, що представлено у таблиці 1.2.

Таблиця 1.2. - Переваги та недоліки у обраних джерелах

Джерело	Переваги	Недоліки
[8]	1.Широке порівняння класичних і глибинних моделей 2.Дуже високі показники точності для LSTM/BERT (98%)	1.Немає адаптації під конкретні мовні середовища 2.Дані статичні, без можливості перенавчання
[9]	1.Детальний аналіз за заголовками і повним текстом 2.Bagging дає майже ідеальні показники (0.996)	1.Моделі не адаптуються до нових даних 2.Не використано сучасні трансформери
[10]	1.LLM демонструють найкращі результати (98,6%) 2.Ефективність навіть менших моделей (GPT-4o-mini)	1.Висока обчислювальна вартість 2.Залежність від великих мовних моделей без можливості локального донавчання
[12]	1.Висока точність (0.898) 2.Поєднання текстових і нетекстових даних	1.Складна архітектура 2.Високі вимоги до ресурсів та даних
[13]	1.Дуже висока точність (99.88%) 2.Проста архітектура для розгортання	1.Ризик перенавчання 2.Результати залежать від якості ембеддінгів GloVe
[14]	1.Комплексне порівняння багатьох моделей та технік векторизації 2.Враховано метрики MCC та ROC-AUC	1.Відсутність веб- або системної реалізації 2.Обмеження мовою датасету

У роботі [8] розглядається проблема автоматичного виявлення фейкових новин із використанням методів машинного навчання (ML) та обробки природної мови (NLP). Автори наголошують, що в сучасному інформаційному середовищі цифрові платформи активно сприяють поширенню дезінформації, що має значний вплив на політику, економіку та соціальну стабільність. У зв’язку з цим дослідження спрямоване на створення автоматизованої системи ідентифікації неправдивих повідомлень, яка поєднує алгоритми класичного та глибинного навчання.

Для класифікації фейкових новин у дослідженні було застосовано як традиційні алгоритми машинного навчання, так і сучасні моделі

глибинного навчання. Зокрема, модель Logistic Regression продемонструвала точність 73.75%, що свідчить про її базову ефективність при розв’язанні задачі класифікації текстів. Алгоритм Decision Tree показав дещо кращий результат із точністю 89.66%, що пояснюється його здатністю моделювати складніші нелінійні залежності між ознаками. Модель Naive Bayes досягла точності 74.19%, підтверджуючи свою ефективність на невеликих наборах текстових даних, однак з обмеженням у врахуванні контексту. Алгоритм Support Vector Machine (SVM) показав точність 76.65%, забезпечуючи добрий баланс між узагальненням і точністю класифікації. Серед моделей глибинного навчання найкращі результати продемонстрували нейронні архітектури. Модель LSTM (Long Short-Term Memory) показала значно вищу точність порівняно з класичними підходами, а саме 96%, завдяки здатності ефективно враховувати послідовні залежності в тексті. Найвищі показники точності отримано за допомогою моделі BERT (Bidirectional Encoder Representations from Transformers), що рівна приблизно 98%, яка перевершила всі інші методи завдяки глибокому контекстному розумінню тексту та здатності аналізувати зв’язки між словами в обох напрямках. Отримані результати свідчать, що глибинні нейронні мережі (особливо LSTM та BERT) значно перевершують класичні ML-моделі за точністю класифікації та здатністю розуміти контекст повідомлення. У майбутньому автори пропонують удосконалити підхід шляхом використання attention-based моделей, word2vec для аналізу мультимодальних даних (зображення, відео) та розширення системи на різні тематичні категорії, такі як політичні, релігійні, медичні тощо.

Загалом, ця публікація має високу практичну цінність, оскільки демонструє порівняльний аналіз різних ML і DL-підходів, визначає оптимальні алгоритми для виявлення фейкових новин і закладає основу

для подальших досліджень у напрямі мультимовних та мультимодальних систем перевірки достовірності контенту.

У роботі [9] “Rapid Detection of Fake News Based on Machine Learning Methods” автори Barbara Probierz, P. Stefański та J. Kozak досліджують проблему автоматизованого виявлення фейкових новин із використанням алгоритмів машинного навчання. Основною метою дослідження є розробка швидкої та ефективної моделі класифікації новин, яка здатна визначати достовірність інформації лише за заголовком або за повним текстом новини. Автори порівняли ефективність кількох популярних алгоритмів машинного навчання, таких як CART (Classification and Regression Tree), SVM (Support Vector Machine), Random Forest, AdaBoost та Bagging. В експерименті застосовувалися NLP-техніки для попередньої обробки тексту (токенізація, видалення стоп-слів, лематизація, векторизація). Дослідження проводилося у двох сценаріях - класифікація лише за заголовками новин і класифікація за повним текстом.

Результати експериментів показали, що аналіз лише заголовків дозволяє досягти прийнятної точності класифікації, однак використання повного тексту новини суттєво підвищує якість виявлення фейків. Найвищу точність показав алгоритм Bagging, який досяг середнього значення accuracy, який рівний 0.9964 та f1-score 0.9965, демонструючи найкращий баланс між точністю і стабільністю результатів. AdaBoost і CART показали подібно високі результати, тоді як SVM продемонстрував нижчу точність (0.9889) і мав найбільший час навчання, понад 1380 секунд, що у 40–45 разів довше, ніж у інших моделей. Це свідчить про обмежену ефективність SVM у завданнях, які потребують швидкої обробки великої кількості текстів. Авторами зроблено висновок, що для досягнення оптимального балансу між швидкістю та якістю класифікації доцільно використовувати гібридну (багатокритеріальну) модель, у якій перший етап включає швидку перевірку достовірності за заголовком, а

другий глибший аналіз повного тексту лише для підозрілих матеріалів. Такий підхід дозволяє скоротити час обробки, водночас зберігаючи високу точність класифікації.

Таким чином, хоча це дослідження робить важливий внесок у напрям швидкого виявлення фейкових новин, запропонований у магістерській роботі підхід є більш сучасним, контекстно глибоким та орієнтованим на україномовні дані, що робить його значно релевантнішим для практичного застосування в умовах українського медіапростору.

У дослідженні [10] автори здійснили порівняльний аналіз ефективності різних архітектур штучного інтелекту, зокрема Convolutional Neural Networks (CNNs), моделей на основі обробки природної мови (NLP) та великих мовних моделей (LLMs) - у задачі класифікації фейкових новин. Основна увага приділяється моделям GPT-4 Omni (GPT-4o) та GPT-4o-mini, які були спеціально доопрацьовані (fine-tuned) для роботи з корпусом новинного контенту. Результати експериментів показали, що ці моделі досягають вражаючої точності - 98,6%, що суттєво перевищує показники традиційних моделей на основі CNN, які продемонстрували лише 58,6% точності. Водночас дослідження засвідчило, що менша за розміром модель GPT-4o-mini показала майже ідентичні результати при значно нижчих обчислювальних витратах, що підтверджує її економічну доцільність у практичному використанні.

Окрім технічного внеску, дослідження має практичну спрямованість, оскільки підкреслює потенціал використання оптимізованих LLM-моделей у роботі журналістських редакцій, фактчекінгових платформ та новинних агентств. Автори пропонують використовувати менші, але точно налаштовані моделі, щоб досягати балансу між продуктивністю, точністю та обчислювальними витратами, що є критично важливим для медіаіндустрії.

В результаті, дане дослідження робить вагомий внесок у розвиток застосування LLM-моделей для виявлення фейкових новин, демонструючи їхню значну перевагу над класичними підходами. Проте запропонована в магістерській роботі система розвиває ці ідеї далі, поєднуючи високу точність трансформерних моделей із локальною лінгвістичною адаптацією та оптимізацією для україномовних даних, що робить її більш релевантною для реалій українського інформаційного простору.

У статті [11] представлено глибокий огляд сучасних підходів до виявлення фейкових новин із використанням методів глибинного навчання (Deep Learning) та обробки природної мови (NLP). Автори узагальнили результати численних досліджень, що охоплюють розвиток систем автоматичного розпізнавання неправдивої інформації, класифікацію існуючих підходів, а також детально проаналізували сильні й слабкі сторони різних архітектур. Основна ідея роботи полягає у створенні таксономії методів виявлення фейкових новин, яка охоплює традиційні моделі машинного навчання, рекурентні та згорткові нейронні мережі (RNN, CNN), а також сучасні трансформерні архітектури (наприклад, BERT, RoBERTa, XLNet). Автори наголошують, що саме інтеграція NLP-технік із глибинними моделями забезпечує найвищу точність у класифікації контенту завдяки здатності моделей аналізувати семантичний контекст, тональність та стилістичні особливості тексту.

У дослідженні проведено порівняльний аналіз ефективності різних DL-архітектур у контексті завдань виявлення дезінформації. Зокрема, автори зазначають, що моделі на основі LSTM і GRU добре справляються з послідовними текстами, однак поступаються трансформерам, які здатні обробляти довгі текстові залежності та виявляти глибинні зв’язки між словами. Також у роботі розглянуто низку метрик оцінки якості моделей (accuracy, precision, recall, F1-score), що дозволяє узагальнити порівняння результатів різних досліджень. Огляд робить висновок, що глибинне

навчання значно знижує ризик похибок при виявленні фейкових новин, особливо коли моделі навчаються на великих і різноманітних наборах даних. Водночас автори визнають, що галузь все ще перебуває на етапі активного розвитку: з'являються нові архітектури, комбіновані моделі та мультимодальні підходи, які враховують не лише текст, а й зображення, відео чи соціальні взаємодії.

В загальному стаття надає вагомий внесок у систематизацію наукових знань у сфері боротьби з фейковими новинами. Водночас, розроблена у межах магістерського дослідження модель розвиває ці ідеї далі - поєднуючи теоретичні висновки з практичним застосуванням та адаптуючи сучасні методи глибинного навчання для українського інформаційного середовища.

У [12] дослідженні представлено вдосконалений підхід до класифікації фейкових новин із використанням ансамблевої глибинної нейронної архітектури, що поєднує кілька моделей для досягнення високої точності. Основна мета роботи - зменшити негативний вплив дезінформації в соціальних мережах шляхом автоматичного виявлення неправдивих новин, які швидко поширюються в онлайн-просторі. Автори у своєму дослідженні застосували попередню обробку даних, що включала очищення текстів, токенізацію та нормалізацію, а також використали методи обробки природної мови (NLP) для роботи з текстовими атрибутами новин. Для побудови моделі було реалізовано гібридну архітектуру, яка поєднувала два основні компоненти: Deep Learning Dense Model, призначену для обробки дев'яти нетекстових атрибутів (таких як джерело новини, дата публікації, тематика тощо), та Bi-LSTM-GRU-Dense Model, яка використовувалася для аналізу основного тексту твердження (statement feature).

Отримані результати експериментів показали, що запропонована авторами модель перевершила всі попередні підходи, протестовані на

наборі даних LIAR, досягнувши точності 0.898, що є найвищим результатом серед відомих досліджень у цій галузі. Для порівняння, моделі, побудовані на основі CNN та NLP, забезпечували точність у межах 0.27–0.60, модель CapsNet показала результат 0.649, алгоритми SVM та Bi-LSTM - 0.61, а Random Forest - лише 0.442. Це показує, що використання ансамблевого підходу, що поєднує різні глибинні моделі, дозволило суттєво підвищити якість класифікації фейкових новин без істотного збільшення часу обчислень, що свідчить про ефективність і перспективність даного методу для автоматичного виявлення неправдивого контенту.

У контексті магістерського дослідження дана робота є важливою, адже вона демонструє ефективність ансамблевих глибинних моделей, що можуть бути використані як основа для розробки україномовних систем виявлення дезінформації. Проте, на відміну від англомовного корпусу LIAR, у цій магістерській роботі зроблено акцент на україномовних новинах, а також реалізовано підхід із повторним навчанням моделей на локальних даних, що підвищує адаптивність системи до специфіки національного інформаційного простору.

Отже, дослідження “Fake Detect” робить суттєвий внесок у розвиток технологій боротьби з фейковими новинами, а результати, отримані в межах цієї магістерської роботи, продовжують розвивати цей напрям у контексті україномовного цифрового середовища, поєднуючи точність ансамблевих методів із контекстною глибиною трансформерних моделей.

У даному дослідженні під назвою “Optimization and improvement of fake news detection using deep learning approaches for societal benefit” [13] автори зосередилися на розпізнаванні та класифікації фейкових і достовірних новин із метою мінімізації їхнього негативного впливу на суспільство та державні процеси. Мета роботи полягає в підвищенні обізнаності користувачів щодо поширення неправдивої інформації та

впливу її пропагандистів у сучасному медійному середовищі. Для досягнення цієї мети була використана глибинна нейронна мережа на основі архітектури LSTM (Long Short-Term Memory). Перед навчанням моделі застосовувалися методи попередньої обробки даних, включаючи видалення стоп-слів та токенізацію для виділення ознак і векторизації тексту. Для подання слів у векторному форматі використовувався метод GloVe, який дозволяє зберігати семантичне значення кожного слова в багатовимірному просторі.

Інформація з тексту новини проходить через кілька рівнів LSTM-архітектури, що дозволяє ефективно враховувати контекст і послідовність слів у реченнях. Модель навчається на класифікації даних із метою прогнозування, чи є новина достовірною, чи фейковою. В якості метрики оцінки ефективності використана точність (accuracy), яка для запропонованої моделі склала 99.88%, що є надзвичайно високим показником. Автори також пропонують перспективу розвитку дослідження шляхом розширення моделі для автоматизованого виявлення фейкових новин на платформах електронної комерції, де питання достовірності інформації стає критично важливим для користувачів і бізнесу. Таким чином, робота демонструє значний внесок у підвищення якості автоматичного виявлення неправдивих новин і має практичне значення для захисту інформаційного середовища та підвищення соціальної безпеки.

У дослідженні [14] автори запропонували комплексну методологію виявлення фейкових новин, що поєднує сучасні підходи машинного навчання та обробки природної мови (NLP) для підвищення точності і ефективності класифікації неправдивої інформації. Було оцінено широкий спектр моделей - від класичних алгоритмів, таких як Naïve Bayes, SVM та Random Forest, до глибинних мереж, включаючи CNN, LSTM та BERT, на різноманітних наборах даних, що охоплюють соціальні мережі, новинні статті та контент користувачів. Експериментальні результати показали, що

моделі глибинного навчання, зокрема BERT, суттєво перевершують традиційні методи, демонструючи вищу точність та стійкість до складних мовних структур. Використання контекстно-залежних ембеддінгів, таких як BERT і Word2Vec, значно покращує ефективність виявлення порівняно з класичними методами векторизації на кшталт TF-IDF чи Bag of Words. Додатково, оцінка моделей за допомогою MCC та ROC-AUC дозволила отримати глибше розуміння надійності класифікаторів, особливо для дисбалансних наборів даних, де показник точності сам по собі може вводити в оману.

У підсумку, робота значно сприяє розвитку методів виявлення фейкових новин, визначаючи ефективні стратегії класифікації, оцінюючи сучасні NLP-техніки та окреслюючи перспективні напрями досліджень. Інтердисциплінарна співпраця між AI-дослідниками, політиками, медіа організаціями та інститутами цієї галузі є ключовою для створення масштабованих, надійних і етично обґрунтованих рішень у боротьбі з дезінформацією.

У цьому дослідженні [15] запропоновано метод виявлення фейкових новин на основі розподіленого машинного навчання з використанням технології Spark. Автори застосували Headline Stance Checker, який аналізує відповідність заголовка та основного тексту новини, класифікуючи її як agree, disagree, discuss або unrelated. Для обробки тексту використовували N-grams, HashingTF-IDF та Count Vectorizer. Основною моделлю став stacked ensemble, який був протестований на кластері Spark із великими обсягами даних (Fake News Challenge FNC-1). Результати показали високу якість класифікації з F1-score 92,45%, що перевищує показники базових методів на 9,35%.

Проведений аналіз показує, що сучасні дослідження у сфері виявлення фейкових новин охоплюють широкий спектр моделей починаючи від класичних ML-алгоритмів до трансформерних архітектур

та великих мовних моделей. Однак більшість розглянутих робіт зосереджені на англомовних даних, не враховують специфіку локальних інформаційних середовищ, обмежуються статичними моделями або не передбачають механізмів подальшого перенавчання. На цьому тлі запропонована у магістерській роботі система має низку суттєвих переваг, які комплексно поєднують сучасні технологічні підходи з практичною спрямованістю:

- орієнтація на україномовний інформаційний простір, що робить систему релевантною для локальних медіа та соціальних мереж;
- використання сучасних трансформерних моделей, адаптованих до української мови та специфіки новинного контенту;
- застосування додаткової ручної перевірки та структурування датасету, що підвищує якість навчання й точність прогнозів;
- реалізація адаптивного перенавчання моделей, що забезпечує стійкість системи до нових типів дезінформації;
- створення повноцінного веб-застосунку, який дозволяє як користувачам, так і адміністраторам взаємодіяти з моделями, керувати даними та здійснювати повторне навчання.

Таким чином, ця класифікаційна робота не лише узагальнює напрацювання попередніх досліджень, але й розвиває їх, долаючи їхні ключові обмеження та адаптуючи сучасні методи до потреб україномовного медіапростору. Система поєднує точність, гнучкість, адаптивність і масштабованість, що робить її вагомим внеском у галузь автоматичного виявлення фейкових новин та відкриває широкий простір для її подальшого розвитку.

РОЗДІЛ 2.
ОБҐРУНТУВАННЯ, ВИБІР ТА РЕАЛІЗАЦІЯ ІНСТРУМЕНТАРІЮ ВИРІШЕННЯ ЗАДАЧІ

2.1. Аналіз та обробка вхідних даних

Задача виявлення фейкових новин належить до напрямку обробки природної мови (NLP) і передбачає автоматичне визначення достовірності текстового контенту. У межах цієї роботи аналізується тільки текстова компонента новинних повідомлень, оскільки саме текст є основним носієм інформації, через який відбувається маніпулювання фактами.

Вхідними даними для моделі є набір текстів новин, кожен з яких має мітку класу - “фейк” або “реальна новина”. Перед етапом навчання проводиться попередня обробка даних, яка включає очищення текстів від шуму (спецсимволів, пунктуації, стоп-слів), нормалізацію, токенізацію та лематизацію. Після цього текстові дані перетворюються у числове подання за допомогою методів векторизації, таких як TF-IDF, Word2Vec або сучасні контекстуальні ембедінги.

Для класифікації новин використовуються різні підходи машинного та глибинного навчання - від класичних моделей (Logistic Regression, Decision Tree, SVM, Gradient boosting) до більш складних нейронних архітектур, таких як RNN та LSTM, або трансформерні моделі (BERT). Основна мета полягає у визначенні оптимального алгоритму, здатного ефективно відрізняти фейкові тексти від достовірних з високою точністю.

Вихідними даними є результати класифікації, отримані під час тестування моделі. Для оцінки ефективності використовуються стандартні метрики, такі як accuracy, precision, recall та F1-score. Також проводиться порівняння результатів різних моделей з метою визначення найефективнішого підходу для поставленої задачі.

Оскільки україномовний сегмент фейкових новин є недостатньо дослідженим, додаткову складність становить робота з україномовним набором даних, що потребує адаптації алгоритмів NLP до мовних особливостей. Крім того, спостерігається швидка втрата актуальності даних, адже тематика фейкових новин постійно змінюється, а моделі мають тенденцію до перенавчання на застарілих або ведених користувачами даних.

Отже, створення ефективної системи для автоматичного виявлення фейкових новин на основі текстових даних є важливою науковою та практичною задачею, що сприятиме підвищенню інформаційної безпеки, зменшенню впливу дезінформації та покращенню якості контенту у цифровому просторі.

Математична постановка задачі буде мати наступний вигляд. Нехай дано набір новин N, кожна з яких описується вектором ознак x_i, де $x_i = (k_i)$, а k_i - це текстова компонента новини (заголовок, опис або повний текст). Також маємо множину класів Y = {0, 1}, де 0 означає, що новина є достовірною, а 1 те що новина є фейковою.

Кожний запис набору N відповідає певному класу з множини Y, і нехай кожне значення x_i має своє числове представлення, отримане після обробки текстових даних методами TF-IDF, Word2Vec або BERT.

Для вирішення задачі необхідно побудувати модель **f**, яка буде класифікувати новини на дві категорії - “реальна” або “фейкова”. Цю модель можна подати у вигляді (2.1):

$$f(x_i) = P(y_i = 1 \mid x_i) \quad (2.1)$$

де $P(y_i = 1 \mid x_i)$ - ймовірність того, що новина i є фейковою за умови, що її вектор ознак дорівнює x_i. Після обчислення цієї ймовірності отримане значення порівнюється з певним пороговим рівнем (наприклад,

0.5) для прийняття рішення про належність новини до одного з двох класів. Згідно з наведеною математичною моделлю, для вирішення задачі будуть використані алгоритми машинного та глибинного навчання, спрямовані на класифікацію новин за їхнім текстовим вмістом.

У процесі побудови моделей ключове значення має якісна попередня обробка даних, вибір ефективного способу текстового векторизування та забезпечення збалансованості вибірки, оскільки україномовні новини характеризуються власними мовними й граматичними особливостями, що впливають на точність класифікації. Крім того, слід враховувати, що актуальність даних швидко змінюється, а моделі можуть перенавчатися на користувацьких даних, тому необхідним є періодичне оновлення набору даних та повторне навчання моделей для збереження високої точності класифікації.

Основним джерелом інформації у задачі виявлення фейкових новин є текстова компонента, адже саме зміст новини визначає її достовірність або неправдивість. Текст містить ключові ознаки, за якими можна виявити маніпулятивні висловлювання, суб’єктивність, емоційне забарвлення чи використання оціночних суджень тому усе це часто є маркерами фейковості.

Математично текст новини можна подати як послідовність символів або слів, кожне з яких має певне числове подання. Для ефективної роботи моделей машинного навчання текст представляється у векторному вигляді, де кожен елемент вектора відповідає певному слову, фразі або ознаці. Наприклад, речення *“*Україна переможе !*”* може бути перетворене у вектор числових значень, що відповідають позиціям слів у словнику, або у вагові коефіцієнти частоти появи слів за допомогою TF-IDF. Такий підхід дозволяє моделі навчатися на текстових закономірностях і розрізняти правдиві та фейкові новини. Для кращого розуміння наведу приклад, як можна подати текст для навчання моделі машинного навчання чи

нейронної мережі, для прикладу візьму текст *“*Україна переможе!*”*, в що ми дуже віримо. Якщо переводити цей текст у набір даних, де кожен елемент відповідає певному символу, то він буде мати наступний вигляд [56, 78, 98, 8, 106, 77, 8, 163, 94, 58, 98, 58, 83, 49, 72, 58, 163, 202] або ж якщо переводити у набір даних, де кожен елемент відповідає певному слову чи виразу, то цей набір буде мати наступний вигляд [569, 778, 948], де відповідно 569 відповідає за слово “Україна”, 778 означає “переможе” та 948 відповідає за символ “!”. Також хотів би наголосити, що це тільки приклад вигляду даних для навчання моделей і у реальному перетворенню тексту у набір чисел, вони будуть мати інший вигляд, але схожий на наведений вище.

У даній роботі використовувався набір даних українською мовою, що робить задачу складнішою через наявність мовних, морфологічних та синтаксичних особливостей української мови, а також меншу кількість великих корпусів порівняно з англомовними наборами даних.

Попередня обробка текстових даних є одним із ключових етапів у задачі виявлення фейкових новин, оскільки якість вхідних даних безпосередньо впливає на точність і стабільність роботи моделей машинного навчання. На даному етапі текстові новини проходять низку перетворень, що дозволяють підготувати їх до подальшої векторизації та навчання моделей.

Основні етапи попередньої обробки тексту:

1. Зчитування та підготовка даних. На вхід моделі подаються тексти новин разом із відповідними класами (1 - правдива новина, 0 - фейкова). Для зручності та ефективності роботи з набором даних використовується бібліотека pandas, яка дозволяє швидко зчитувати та опрацьовувати великі обсяги табличних даних у форматі CSV.

2. Перетворення тексту до нижнього регістру. Всі символи тексту зводяться до нижнього регістру. Це дозволяє уникнути дублювання токенів, коли одне й те саме слово записане з великої або малої літери.

3. Очищення тексту від небажаних символів. За допомогою регулярних виразів із текстів видаляються небуквені символи, пунктуація, цифри, емодзі та інші елементи, що не несуть семантичного значення. У результаті залишаються лише українські слова, літери та пробіли, що забезпечує чистоту даних і покращує якість подальшої токенізації.

4. Видалення стоп-слів. З тексту вилучаються часто вживані службові слова, які не мають суттєвого смислового навантаження (наприклад, “це”, “є”, “також”, “і”). Для цього використовується заздалегідь підготовлений список, що містить найпоширеніші слова української мови, які не несуть значення для виявлення фейкових новин. Це допомагає зменшити обсяг даних та зосередити увагу моделей на ключових термінах, що можуть містити маркери фейковості, емоційності або маніпуляції.

5. Токенізація тексту. Це процес розбиття тексту на окремі слова (токени) приклад цього я наводив вище, де перетворював речення у набір чисел. Цей крок є необхідним для подальшого аналізу частоти та контексту появи слів, що використовується у більш складних моделях, таких як Word2Vec або BERT.

6. Балансування класів. Для уникнення домінування одного з класів, дані балансувалися шляхом випадкового вибору рівної кількості прикладів для кожного класу. Це зменшує ймовірність зміщення, коли модель робить більш точні прогнози для одного класу, наприклад коли більш точно визначає фейкові новини, а от на справжніх робить більше помилок і часто відносить їх до фейкових новин.

7. Векторизація тексту. Цей етап дозволяє оцінити важливість слова в документі відносно всієї колекції текстів та показує, як часто слово

зустрічається в певному документі. Цей підхід зменшує вагу слів, які часто зустрічаються у всіх документах (наприклад, “новина”, “сьогодні”). Таким чином, він допомагає моделі розпізнавати найважливіші терміни, що мають унікальне значення для кожної новини.

8. Розділення даних на навчальну та тестову вибірки. Набір даних поділяється на три частини, а саме тренувальну, валідаційну та тестову. Основна частина даних використовується для навчання моделі, валідаційна вибірка використовується для підбору гіпер параметрів і контролю перенавчання, а тестова для кінцевої оцінки якості побудованої моделі на нових, невідомих їй прикладах. Таке розділення забезпечує об’єктивність оцінки та дозволяє перевірити здатність моделі узагальнювати результати. Співвідношення частин задається за допомогою числа, що визначає відсоток даних для кожної категорії (наприклад, 60% - навчальні, 20% - валідаційні, 20% - тестові).

Після проходження цих етапів отримується чистий, збалансований і векторизований набір текстових даних, готовий для навчання моделей машинного та глибинного навчання. Особлива увага приділяється роботі з українською мовою, оскільки вона має складну морфологію, варіативність форм і обмежену кількість великих корпусів для моделювання. Також враховується, що актуальність даних швидко змінюється, а моделі можуть перенавчатися на інформації, яку користувачі часто повторюють або поширюють, тому процес оновлення та перевірки набору даних є важливою складовою дослідження.

2.2 Математичне та алгоритмічне забезпечення

Логістична регресія - це один із найпоширеніших алгоритмів машинного навчання, який використовується для бінарної класифікації, тобто для передбачення, чи належить певний об’єкт до одного з двох класів. У контексті нашого завдання, а саме виявлення фейкових новин

модель прогнозує, чи є новина правдивою (0) або фейковою (1). На відміну від звичайної лінійної регресії, яка прогнозує неперервні числові значення, логістична регресія передбачає ймовірність належності спостереження до певного класу.

Для цього вона використовує сигмоїдальну (логістичну) функцію, яка перетворює лінійну комбінацію вхідних змінних у значення від 0 до 1, що інтерпретується як ймовірність.

$$y = w_0 + w_0 x_0 + w_1 x_1 + ... + w_n x_n \quad [16] \ (2.2)$$

Де x_i - вхідні ознаки, w_i - вагові коефіцієнти та w_0 - зсув (bias). Після цього результат у подається до сигмоїдальної функції:

$$p = \frac{1}{1 + e^{-y}} \quad [16] \ (2.3)$$

Де e - число Ейлера (≈ 2.718). Отримане значення p - це ймовірність того, що приклад належить до класу 1 (фейкова новина). Якщо ця ймовірність перевищує певний поріг (зазвичай 0.5), об’єкт класифікується як фейковий; інакше — як правдивий. Для кращого прикладу наведений графік сигмоїдальної функції.

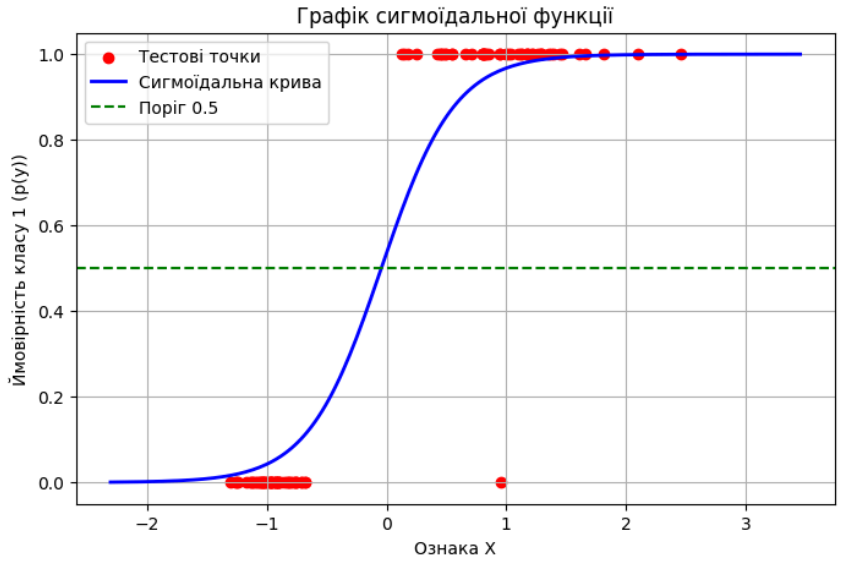


Рисунок 2.1 - Графік сигмоїдальної функції

На графіку зображено сигмоїдальну криву, що відображає залежність ймовірності належності об’єкта до класу 1 від значення ознаки x. Червоні точки представляють вхідні дані, які модель логістичної регресії використовує для навчання. Пунктирна зелена лінія на рівні 0.5 позначає поріг класифікації, який визначає межу між класами, тобто вище 0.5 об’єкт належить до класу 1, нижче відповідно до класу 0. Для кращого розуміння наведу ще один графік, який відображає, як працює метод лінійної логістичної регресії

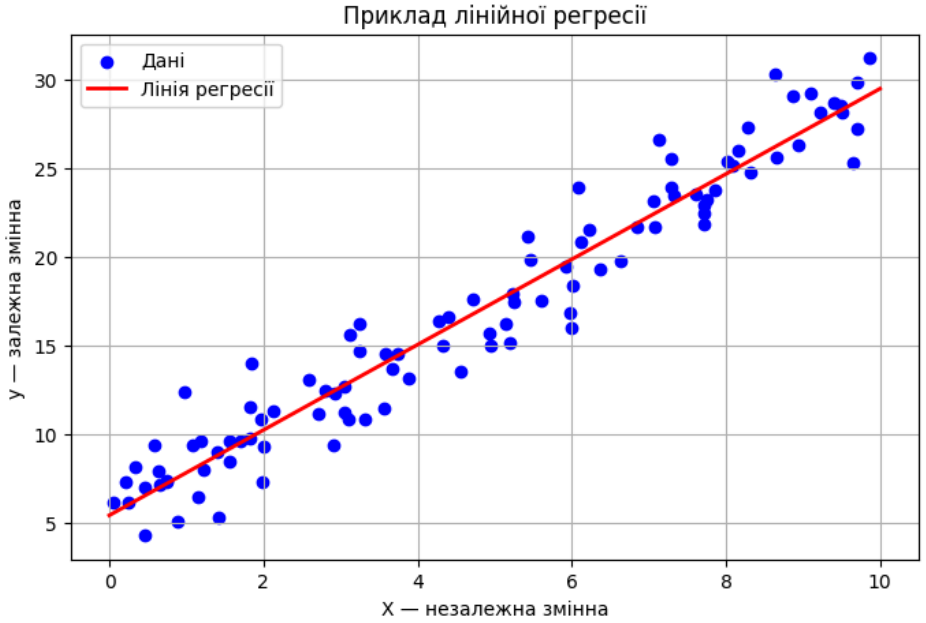


Рисунок 2.2 - Графік прикладу лінійної логістичної регресії

На цьому графіку зображено результат роботи моделі для лінійної регресії. Сині точки представляють вихідні дані, що демонструють розсіювання значень залежної змінної від незалежної. Червона лінія показує знайдену моделлю залежність, тобто найкращу лінію, яка мінімізує похибку між фактичними та передбаченими значеннями. Відповідно до графіку всі точки, що під червоною лінією будуть класифікуватися до 0 класу або ж, як у нашому випадку що новина не фейкова, а ті що вище червоної лінії будуть кваліфікуватися до класу 1,

тобто що новина фейкова Таким чином, графік наочно демонструє, як лінійна регресія виявляє загальну тенденцію у наборі даних.

В результаті бачимо, що логістична регресія є простим, проте ефективним методом бінарної класифікації, який дозволяє оцінювати ймовірність належності об’єкта до певного класу. Вона базується на сигмоїдальній функції, що перетворює лінійну комбінацію ознак у значення від 0 до 1, що зручно для інтерпретації як ймовірності. Модель добре працює, коли дані мають лінійно відокремлювані класи та не містять надлишкових ознак. Завдяки своїй простоті, швидкодії та інтерпретованості, логістична регресія часто використовується як базовий алгоритм для порівняння складніших моделей машинного навчання.

Дерево рішень - це зрозумілий та популярний метод машинного навчання, який використовується як для задач класифікації, так і регресії. Основна ідея полягає у побудові структури у вигляді дерева, де вузли представляють умови перевірки певних ознак (наприклад, “чи містить новина певне слово”), гілки відповідають за результати перевірки, а листові вузли за кінцеві класи або прогнозовані значення.

Під час навчання алгоритм дерева рішень послідовно розділяє дані на підмножини, формуючи структуру, у якій кожен вузол відповідає певній ознаці та порогу її значення. Основна мета цього процесу є знайти такі поділи, які найкраще відокремлюють об’єкти різних класів, тобто роблять підмножини максимально “чистими”. Для визначення якості розбиття використовуються спеціальні критерії нечистоти, для побудови дерева рішень у цій роботі було використано індекс Джині (Gini Index). вимірює ймовірність неправильного віднесення об’єкта до певного класу. Чим менше значення Gini, тим чистіше підмножина.

$$Gini(t) = 1 - \sum_{i=1}^{n} (p(i|t))^2 \text{ [17] (2.4)}$$

де t - це вузол дерева рішень, а $p(i \mid t)$ відповідає за ймовірність того, що об’єкт належить до класу i у межах цього вузла. Це значення використовується для оцінки якості розбиття даних між вузлами дерева: кращим вважається те, яке має нижче значення індексу Джині. Індекс Джині приймає значення від 0 до 1, де 0 означає повну чистоту (усі об’єкти належать одному класу), а 1 - повну випадковість розподілу між класами. Значення 0,5 свідчить про рівномірний розподіл елементів між класами. Цей критерій застосовується до категоріальних цільових змінних і зазвичай виконує двійкові розбиття. Індекс Джині є популярним у класифікаційних задачах, оскільки він менш схильний до перенавчання, ніж інші критерії.

Також наведу візуальний приклад, як працює метод Дерево рішень на рисунку 2.3.

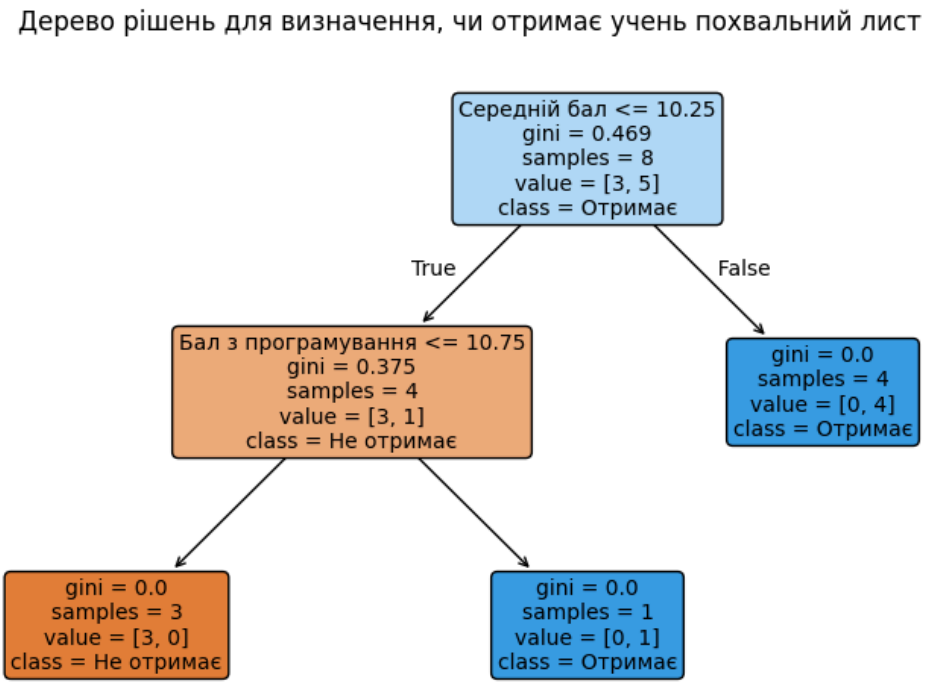


Рисунок 2.3 - Діаграма структури дерева рішень

Як бачимо з діаграми, дерево рішень починається з кореневого вузла, який не має жодних вхідних гілок, а далі йдуть вузли прийняття рішень. На кожному кроці модель розділяє дані відповідно до певної умови, вибираючи найбільш інформативні ознаки. Цей процес триває доти, поки не буде досягнуто листкових вузлів, де отримується кінцеве рішення або клас. Для прикладу можна уявити ситуацію, коли ми визначаємо, чи отримає учень похвальний лист, використовуючи його середній бал і оцінку з програмування. Згідно з діаграмою, учень отримає грамоту, якщо його середній бал вище 10 або бал з програмування не нижчий за 11. Аналогічний принцип використовується і в нашій темі виявлення фейкових новин. У цьому випадку вузли дерева можуть представляти ознаки тексту, наприклад, частоту появи емоційно забарвлених слів, наявність маніпулятивних фраз чи ненадійних джерел. На кожному рівні модель приймає рішення, яке поступово наближає її до визначення що є новина фейковою чи правдивою. Проте, коли набір даних містить велику кількість атрибутів і текстових ознак, вибір оптимального розподілу для кореня та наступних вузлів стає значно складнішим завданням.

Однією з головних переваг дерева рішень є зрозумілість результату, що його можна візуалізувати у вигляді послідовності логічних правил (“якщо... то...”), що робить модель легко інтерпретованою навіть для нефахівців. Однак, якщо не контролювати глибину дерева, модель може перенавчитися, тобто надто точно запам’ятати тренувальні дані й погано працювати на нових прикладах. Хоча дерево рішень є відносно простим методом, воно часто використовується як базовий метод для класифікації та у потужніших ансамблевих алгоритмах, таких як Random Forest або Gradient Boosting.

Класифікатор опорних векторів (SVC - Support Vector Classifier) є одним із найефективніших і найпоширеніших алгоритмів машинного навчання, що використовується для задач класифікації та розпізнавання

закономірностей. Його головна ідея полягає у знаходженні оптимальної гіперплощини, яка максимально чітко розділяє елементи різних класів у багатовимірному просторі ознак. Кожен вхідний об’єкт (наприклад, текст новини) подається у вигляді векторного представлення, де кожна координата відповідає певній характеристиці чи слову.

На відміну від інших методів, SVC не використовує всі навчальні приклади для прийняття рішень, а лише ключові точки — опорні вектори, що знаходяться найближче до межі розділення. Саме вони визначають положення та орієнтацію гіперплощини, яка мінімізує помилки класифікації. Такий підхід дозволяє моделі бути стійкою до шуму та надлишкових даних, забезпечуючи високу узагальнювальну здатність навіть при роботі з великими або незбалансованими наборами текстів.

У контексті задачі виявлення фейкових новин, метод опорних векторів демонструє особливу ефективність, оскільки може знаходити тонкі смислові відмінності між справжніми та маніпулятивними повідомленнями, базуючись на їхніх текстових ознаках. Завдяки цьому SVC є надійним інструментом для створення систем автоматичного моніторингу інформаційного простору та виявлення недостовірного контенту. Для кращого наочного розуміння наведено приклад роботи методу опорних векторів на рисунку - 2.4. На рисунку зображено роботу методу опорних векторів (SVC) для задачі двокласової класифікації. Сині та червоні точки представляють об’єкти двох різних класів, які потрібно розділити. Чорна суцільна лінія - це лінія розділення (гіперплощина), яка визначає межу між класами. Пунктирні лінії позначають границі (margin) області, що проходять на однаковій відстані від гіперплощини. Зеленими колами виділені опорні вектори або точки, які знаходяться найближче до межі й безпосередньо впливають на положення лінії розділення. Саме вони визначають оптимальне розташування гіперплощини, забезпечуючи максимально можливий розрив між класами.

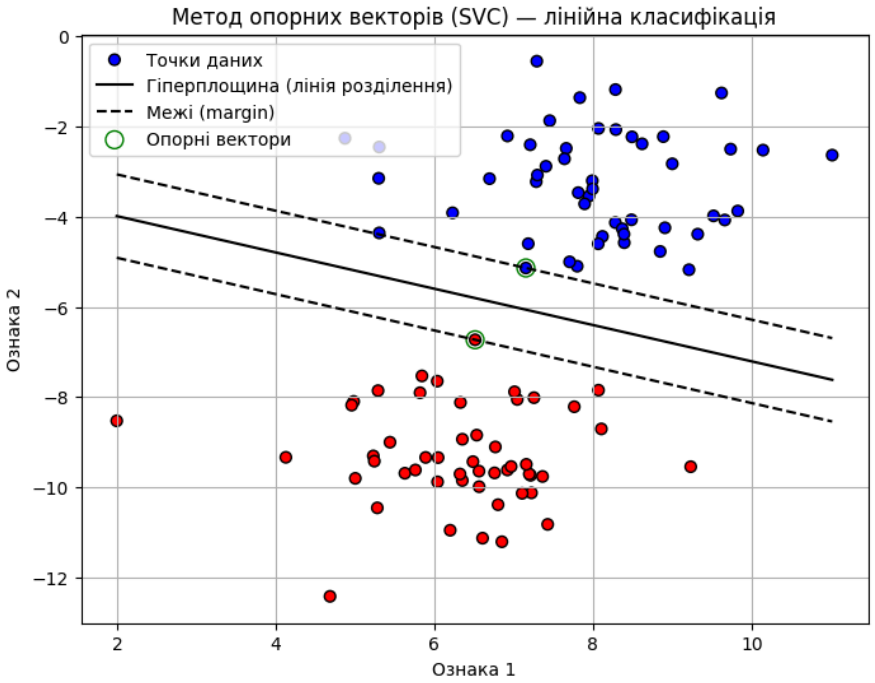


Рисунок 2.4 - Схематичне зображення роботи SVC

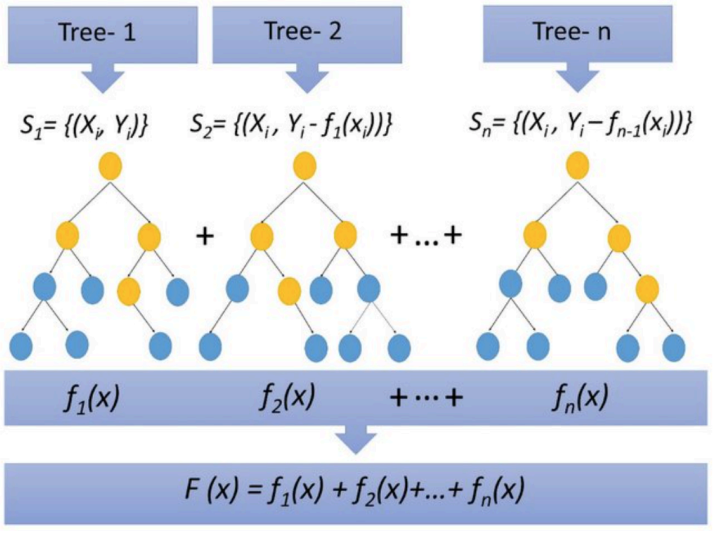
Метод опорних векторів (SVC) є одним із найпотужніших алгоритмів машинного навчання для задач класифікації. Його основна перевага полягає у здатності ефективно розділяти класи навіть у складних, багатовимірних просторах, завдяки побудові оптимальної гіперплощини. Використання лише опорних векторів робить SVC стійким до шуму та надлишкових даних, що підвищує точність класифікації. У контексті виявлення фейкових новин цей метод дозволяє виявляти приховані закономірності у текстах, розпізнаючи маніпулятивні формулювання або ненадійні джерела. Завдяки високій узагальнювальній здатності, SVC показує відмінні результати навіть на невеликих або незбалансованих наборах даних.

Метод градієнтного підсилення (Gradient Boosting) - це потужний ансамблевий алгоритм машинного навчання, який поєднує велику кількість слабких моделей (зазвичай дерев рішень) у єдину сильну модель для покращення точності класифікації чи регресії. Його основна ідея

полягає в послідовному побудуванні моделей, де кожна наступна модель намагається виправити помилки попередніх.

На початку створюється проста модель (наприклад, невелике дерево рішень), яка робить перші передбачення. Потім обчислюються похибки (різниця між реальними та передбаченими значеннями), і наступне дерево навчається передбачати саме ці помилки. Таким чином, кожна нова модель "підсилює" попередні, поступово зменшуючи похибку. Процес відбувається доти, поки не буде досягнута певна кількість ітерацій або не перестане покращуватися результат. Для контролю навчання використовується швидкість навчання (learning rate), яка визначає, наскільки сильно кожна нова модель впливає на загальне передбачення, у цій роботі це значення бралась як 0.5.

У задачі виявлення фейкових новин метод градієнтного підсилення демонструє високу ефективність, оскільки він здатен враховувати складні взаємозв'язки між словами та їхнім контекстом, об'єднуючи рішення багатьох слабких моделей у точний класифікатор. Це робить його одним із найкращих методів для обробки текстових даних і виявлення дезінформації у великих інформаційних потоках. Приклад цього методу наведений на рисунку 2.5 [18]



Рисунку 2.5 - Схематичне зображення роботи Gradient Boosting [18]

На зображенні показано принцип роботи методу Gradient Boosting. У процесі навчання модель послідовно створює кілька дерев рішень, де кожне наступне дерево (Tree-2, Tree-n) навчається на помилках попередніх. Спочатку перше дерево $f_1(x)$ наближує цільову функцію, потім друге дерево $f_2(x)$ моделює різницю між реальними значеннями Y_1 і прогнозами $f_1(x_i)$, а далі кожне наступне дерево коригує помилки попередніх. У результаті фінальна модель $F(x)$ є сумою всіх окремих дерев, тобто:

$$F(x) = f_1(x) + f_2(x) + \ldots + f_n(x) \text{ [18] (2.5)}$$

Ця ілюстрація демонструє ідею поетапного покращення прогнозу за рахунок поступового “підсилення” слабких моделей, що є сутністю градієнтного бустингу.

Метод градієнтного бустингу є одним із найефективніших підходів у машинному навчанні для задач класифікації та регресії. Його сила полягає у послідовному поєднанні простих моделей, зазвичай дерев рішень, де кожне наступне дерево виправляє помилки попередніх. Такий підхід дозволяє досягати високої точності навіть на складних наборах даних із великою кількістю ознак. У контексті задачі виявлення фейкових новин градієнтний бустинг може ефективно враховувати тонкі мовні відмінності та взаємозв’язки між словами. Завдяки своїй здатності до навчання на залишкових помилках цей метод часто перевершує класичні алгоритми, забезпечуючи кращу узагальнювальну здатність моделі.

Рекурентні нейронні мережі (RNN, Recurrent Neural Networks) - це тип нейронних мереж, який призначений для обробки послідовних даних, таких як текст, аудіо або часові ряди. Їхня головна особливість полягає в наявності петльових зв’язків, що дозволяють зберігати інформацію про попередні кроки послідовності. Це означає, що під час обробки кожного

нового елемента мережа враховує контекст попередніх слів або символів, що робить RNN надзвичайно корисними для задач, де важливий порядок даних. Наприклад, для аналізу тексту, розпізнавання мовлення або перекладу.

Приклад архітектури та принципу роботи цієї моделі наведено на рисунку 2.6 [19]

Як бачимо з рисунку у процесі навчання RNN оновлює свої внутрішні стани, навчаючись виявляти залежності між елементами вхідної послідовності. Це дозволяє їй “запам’ятовувати” контекст, що особливо важливо для обробки речень у природній мові, де значення слова часто залежить від попередніх.

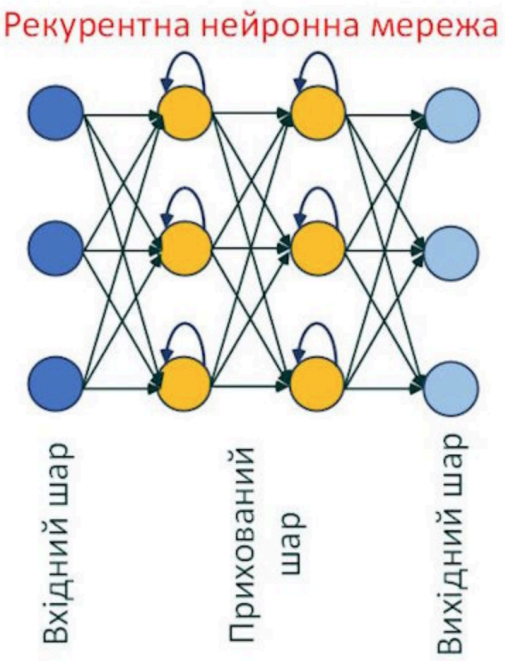


Рисунок 2.6 - Схематичне зображення роботи Рекурентної нейронної мережі [19]

У класичних нейронних мережах кожен вхід обробляється незалежно. У RNN навпаки на кожному кроці часу t мережа отримує нове вхідне значення x_t(наприклад, слово або вектор ознак), а потім обчислює

поточний прихований стан h_t, який залежить не лише від цього вхідного елемента, а й від попереднього стану h_{t-1}. Математично це можна описати у такому вигляді:

$$h_t = f(W_h h_{t-1} + W_x x_t + b) \text{ [20] (2.6)}$$

де h_t це поточний стан пам'яті (прихований стан), h_{t-1} попередній стан (пам'ять про минуле), W_h та W_t це вагові матриці, які навчаються, f - нелінійна функція активації (наприклад, sin, tanh або як у нашому випадку ReLU) та b це зміщення (bias). Далі, вихід мережі на кожному кроці може бути обчислений як:

$$y_t = g(W_y h_t + c) \text{ [20] (2.7)}$$

де y_t - вихід у момент часу t, а g це функція активації (наприклад, softmax або sigmoid як у нашому випадку). Таким чином, RNN не просто аналізує поточний вхід, а враховує контекст попередніх слів чи подій, завдяки чому вона "пам'ятає" історію. На прикладі поставленого завдання ця модель буде працювати, приблизно таким чином: на початку мережа обробляє слово "Новина", зберігає його представлення у стані h_1. Потім, коли з'являється слово "виглядає", вона враховує не лише його значення, а й контекст слова "Новина". На слові "але" RNN уже має інформацію про те, що речення, ймовірно, змінює свій зміст, тому модель може зробити висновок про потенційну фейковість новини.

В результаті бачимо, що RNN аналізує текст новини послідовно, слово за словом, і формує узагальнене представлення змісту з урахуванням контексту. Це дозволяє моделі вловлювати маніпулятивні фрази, емоційні конструкції або не логічні твердження, які часто зустрічаються у фейкових повідомленнях. Таким чином, RNN допомагає створити більш контекстно обґрунтовану модель класифікації.

LSTM (Long Short-Term Memory) - це вдосконалений тип рекурентної нейронної мережі (RNN), створений для обробки послідовних даних і подолання проблеми зникнення градієнтів, яка часто виникає у звичайних RNN під час навчання на довгих послідовностях. Основна ідея LSTM полягає у введенні спеціальних елементів пам'яті (комірок пам'яті), які здатні зберігати інформацію протягом тривалого часу, вибірково оновлювати або забувати її залежно від важливості.

Кожна LSTM-комірка містить три основні "вентилі" (гейти):

✔ вхідний гейт визначає, яку нову інформацію слід додати в пам'ять;

✔ гейт забування вирішує, яку інформацію потрібно видалити;

✔ вихідний гейт контролює, яку інформацію з пам'яті передати далі на вихід.

Для кращого розуміння наведено рисунок 2.7 [21], де зображено схематичний алгоритм роботи LSTM.

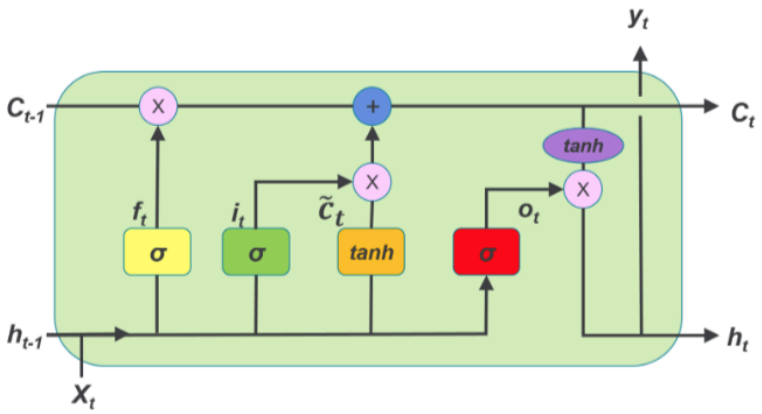


Рисунок 2.7 Схематичне зображення роботи LSTM [21]

Як показано на рисунку 2.7 [21] представлено приклад архітектури LSTM. У даній моделі кожна комірка пам'яті (нейрон) отримує три основні вхідні параметри у вигляді числових векторів: c_{t-1} та h_{t-1} передаються з попереднього кроку разом із вже обчисленими значеннями, а x_t є поточним вхідним вектором. У процесі роботи модель оновлює та зберігає

інформацію про залежності між елементами послідовності за допомогою трьох основних компонент, які називають “воротами”, які є згадані вище. Завдяки такій структурі LSTM здатна ефективно обробляти текстові послідовності, запам’ятовуючи контекст і зв’язки між словами, що робить її надзвичайно корисною для завдань виявлення “мови ненависті” у текстах.

Таким чином, можна сказати, що завдяки цим механізмам LSTM може ефективно враховувати контекст попередніх слів у реченні, що робить її надзвичайно корисною в задачах обробки природної мови, таких як виявлення фейкових новин, аналіз тональності, машинний переклад або розпізнавання мовлення. Серед основних переваг LSTM висока точність на послідовних даних, здатність моделювати довготривалі залежності, а недоліком є висока обчислювальна складність і потреба у значних обсягах даних для навчання.

BERT (Bidirectional Encoder Representations from Transformers) - це одна з найвідоміших і найефективніших моделей глибинного навчання для обробки природної мови (NLP), яка базується на архітектурі трансформерів. Основна ідея BERT полягає у двонаправленому аналізі тексту, тобто модель одночасно враховує контекст слова як зліва направо, так і справа наліво, що дає змогу краще розуміти значення слів у різних контекстах. Завдяки цьому BERT здатна точно вловлювати смислові, граматичні та емоційні залежності між словами, що робить її надзвичайно ефективною у завданнях класифікації тексту, таких як виявлення фейкових новин.

Модель використовує попереднє навчання (pre-training) на великих текстових корпусах з відкритих джерел (наприклад, Wikipedia та BookCorpus). Основою цього процесу є масковане моделювання мови (Masked Language Modeling, MLM) під час навчання певні слова у реченні замінюються спеціальним токеном(mask), а завдання моделі “вгадати” ці

приховані слова на основі контексту. Такий підхід дозволяє BERT навчатися розуміти глибинні мовні структури без необхідності ручного маркування даних. Архітектурно BERT складається з багатошарових енкодерів трансформера, у кожному з яких реалізовано механізм самоуваги (self-attention). Цей механізм дозволяє моделі динамічно визначати, які слова у реченні є найбільш важливими для поточного контексту. Наприклад, у реченні “фейкові новини можуть викликати паніку”, модель розуміє, що слова “фейкові” та “паніку” мають сильну семантичну взаємодію. Крім того, завдяки залишковим зв’язкам (residual connections) і нормалізації, інформація передається стабільно між усіма шарами, що запобігає втраті контексту.

Для адаптації BERT до конкретного завдання використовується етап тонкого налаштування (fine-tuning). У цьому процесі до вже навченої базової моделі додаються нові шари (наприклад, шар класифікації), які спеціалізуються під конкретну задачу, тобто під задачу виявлення фейкових новин, як у даній роботі. Нижні шари моделі залишаються замороженими, зберігаючи попередньо набуті мовні знання, тоді як верхні навчаються заново. Це дозволяє швидко і з високою точністю налаштувати BERT на різні типи аналізу тексту від виявлення тональності до ідентифікації фейкових новин.

Таким чином, BERT поєднує глибоке контекстне розуміння з гнучкістю адаптації, що робить її однією з найефективніших моделей для сучасних NLP-завдань. У контексті виявлення фейкових новин українською мовою, BERT здатна аналізувати навіть складні лінгвістичні конструкції, враховувати сарказм, натяки чи контекстні зміни значень слів, забезпечуючи точну класифікацію та мінімізуючи хибні спрацьовування.

2.3 Проектування моделей для системи

У межах цього дослідження розроблено систему для виявлення фейкових новин на основі текстового аналізу. Основна ідея полягає у використанні методів машинного навчання та глибинного навчання для автоматичного визначення достовірності новинних повідомлень, спираючись на їх лексичні, синтаксичні та контекстуальні особливості. Проектована система складається з кількох основних етапів: попередня обробка текстових даних та векторизація, навчання моделей, класифікація та оцінка ефективності.

Перший етап є попередня обробка текстів, що має на меті приведення даних до форми, придатної для навчання моделей. На цьому етапі здійснюється очищення текстів від HTML-тегів, спеціальних символів, URL-адрес, приведення до нижнього регістру, токенізація, лематизація та видалення стоп-слів. Ці процедури дають змогу зменшити шум у даних і підвищити якість ознак, що використовуються під час навчання моделей. Для забезпечення збалансованості класів також застосовується балансування вибірки, оскільки в реальних новинних даних кількість правдивих матеріалів зазвичай значно переважає кількість фейкових. Також векторизація тексту, тобто перетворення слів у числові представлення. У роботі використовуються сучасні підходи до представлення тексту, такі як TF-IDF, Word2Vec або вбудовані вектори з попередньо навченої моделі BERT, які дозволяють зберігати контекстуальні зв’язки між словами. Це забезпечує більш точне відображення смислового значення тексту в числовому вигляді, що суттєво підвищує якість класифікації.

Наступним етапом є побудова та навчання моделей. Для аналізу текстових новин у системі застосовуються кілька методів: Логістична регресія, Дерево рішень, Класифікатор опорних векторів (SVC), метод градієнтного підсилення (Gradient Boosting) Рекурентні нейронні мережі

(RNN), LSTM, а також BERT. Далі йде класифікація, де модель на основі обчислених ознак визначає, чи є новина фейковою. Вихідним результатом є значення у діапазоні [0; 1], яке інтерпретується як ймовірність того, що текст містить ознаки неправдивої інформації. Та крайнім етапом є оцінювання ефективності системи. Для цього використовуються стандартні метрики машинного навчання: Accuracy, Precision, Recall та F1-score, які дозволяють комплексно оцінити якість класифікації. Візуально зобразити цю систему для виявлення фейкових новин на основі методів машинного навчання, можна наступним чином, як на рисунку 2.8.

Схематичне зображення архітектури розробленої системи наведено на рисунку 2.8. Як видно з рисунку, система працює за принципом: Вхідний текст → Попередня обробка та Векторизація → Навчання моделі машинного навчання → Класифікація та Результат (Фейк / Не фейк).

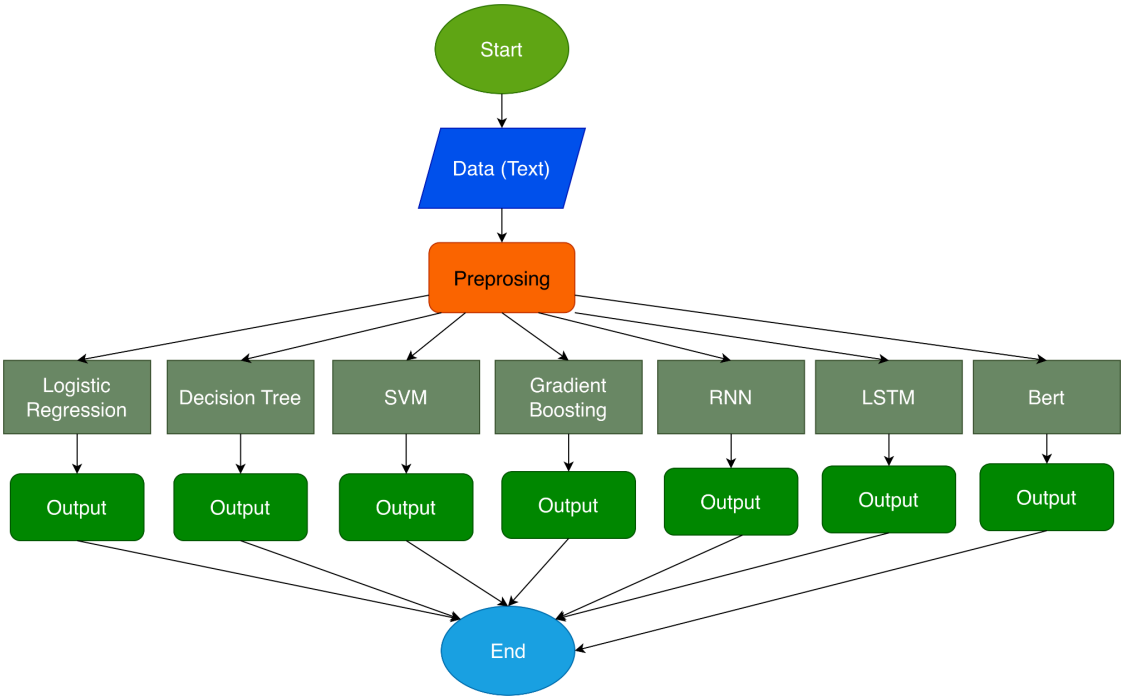


Рисунок 2.8 Схематичне зображення системи для виявлення фейкових новин

Запропонована система є гнучкою та масштабованою, вона може бути розширена для багатомовної підтримки, інтеграції нових моделей або включення додаткових ознак, таких як джерело новини, емоційний тон чи часові закономірності. Це робить її ефективним інструментом для виявлення дезінформації та протидії поширенню фейкових новин у цифровому просторі.

2.4 Метрики для оцінювання роботи методів

В будь-якому дослідженні одним із найважливіших етапів є аналіз отриманих результатів. Для задач класифікації, зокрема виявлення фейкових новин, цей етап є критичним, оскільки саме від правильного аналізу залежить висновок щодо ефективності розробленої системи. Застосування різних метрик оцінювання дозволяє більш об’єктивно оцінити якість моделі, адже використання лише однієї метрики не завжди відображає реальну ефективність, для прикладу у випадках незбалансованих вибірок, коли кількість достовірних новин суттєво перевищує кількість фейкових.

Для нашого дослідження така ситуація є типовою, адже у більшості реальних джерел кількість справжніх новин значно більша, ніж неправдивих. Тому застосування лише показника точності може бути оманливим модель може видавати високі результати, навіть якщо вона переважно класифікує всі новини як “справжні”. Саме тому важливо використовувати комплексний підхід, який поєднує кілька метрик оцінювання.

В даній кваліфікаційній роботі використовувалися наступні метрики для оцінювання результатів моделей машинного навчання для задачі виявлення фейкових новин.

Accuracy або точність - це одна з базових метрик, що використовується для оцінювання ефективності моделей машинного

навчання. Вона показує частку правильно класифікованих зразків серед усіх перевірених даних. У межах цього дослідження метрика демонструє, наскільки часто модель коректно визначає новини як “фейкові” або “справжні”. Точність можна обчислити за формулою (2.8):

$$acc = \frac{(TP + TN)}{(TP + TN + FP + FN)} \text{ [22] (2.8)}$$

де TP це кількість фейкових новин, які правильно класифіковані як фейкові, TN кількість справжніх новин, правильно віднесених до реальних, FP це кількість справжніх новин, які модель помилково визначила як фейкові та FN це кількість фейкових новин, які були класифіковані як справжні.

Попри популярність цієї метрики, вона не завжди є репрезентативною, особливо у випадках незбалансованих наборів даних, коли одна категорія (наприклад, справжні новини) значно переважає іншу (фейкові новини). У таких ситуаціях модель може досягати високого показника точності, навіть якщо не здатна ефективно виявляти фейки. Тому лише показника Accuracy недостатньо для повного аналізу потрібно враховувати інші метрики, що дають більш глибоке розуміння роботи моделі.

Метрика Precision або точність передбачення показує, наскільки велика частка новин, передбачених моделлю як “фейкові”, справді є фейковими. Вона дає змогу оцінити, наскільки модель схильна робити помилкові позитивні передбачення. Формула для обчислення точності передбачення подається нижче (2.9):

$$p = \frac{TP}{(TP + FP)} \text{ [22] (2.9)}$$

Для задачі виявлення фейкових новин ця метрика показує, яку частку від усіх передбачених “фейків” модель класифікувала коректно. Високий показник Precision означає, що система рідко помиляється, коли позначає

новину як фейкову, тобто рівень “помилкових тривог” є низьким. Це особливо важливо для реальних застосувань, таких як автоматичне маркування контенту у соціальних мережах, де помилкове визначення правдивої новини як фейку може призвести до втрати довіри користувачів або блокування правдивих джерел.

Recall або повнота (чутливість) - це метрика, яка визначає, яку частку з усіх реальних фейкових новин модель змогла правильно виявити. Вона оцінює здатність моделі знаходити всі позитивні приклади у вибірці. Повнота розраховується за формулою (2.10):

$$r = \frac{TP}{TP + FN} \text{ [22] (2.10)}$$

Ця метрика є надзвичайно важливою для систем, які мають завдання мінімізувати кількість пропущених фейкових новин. Високе значення Recall свідчить про те, що модель майже не пропускає фейкові матеріали, хоча іноді це може призводити до підвищення кількості хибнопозитивних випадків. У контексті даного дослідження це означає, що система здатна розпізнавати більшість фейкових новин у потоці даних, навіть якщо серед передбачень трапляються окремі помилки.

F1-score або F1-оцінка - це узагальнена метрика, яка поєднує у собі точність (Precision) та повноту (Recall) в одне гармонійне середнє значення. Вона є більш збалансованим показником, особливо у випадках, коли дані є нерівномірно розподіленими між класами. Обчислюється вона за формулою (2.11):

$$F1 = \frac{2 * precision * recall}{precision + recall} \text{ [22] (2.11)}$$

Метрика F1-score є критично важливою для аналізу моделей, що працюють з дисбалансом класів, оскільки вона одночасно враховує як здатність виявляти всі фейки (Recall), так і точність цих передбачень

(Precision). Високий F1-показник означає, що модель демонструє стабільну якість як у точності, так і у повноті.

Матриця помилок (Confusion Matrix) є інструментом для візуалізації роботи класифікатора, який дозволяє побачити, скільки прикладів кожного класу модель класифікувала правильно або неправильно. Для задачі бінарної класифікації (виявлення фейкових і справжніх новин) матриця складається з чотирьох основних елементів:

1. True Positive (TP) - фейкові новини, які модель правильно класифікувала як фейкові;
2. False Positive (FP) - справжні новини, які модель помилково позначила як фейкові;
3. True Negative (TN) - справжні новини, правильно розпізнані як справжні;
4. False Negative (FN) - фейкові новини, які система не розпізнає і віднесла до справжніх.

Матриця помилок наочно демонструє сильні та слабкі сторони моделі, дозволяє побачити, який тип помилок переважає чи пропущені фейки чи хибні передбачення. Її використання допомагає більш глибоко зрозуміти поведінку моделі та вибрати напрям подальшого вдосконалення алгоритмів.

У результаті проведеного аналізу метрик можна зробити висновок, що для комплексної оцінки ефективності моделей класифікації недостатньо використовувати лише одну метрику, таку як Accuracy. Врахування показників Precision, Recall та F1-score дає змогу глибше оцінити якість роботи моделі, особливо у випадках зі незбалансованими наборами даних. Матриця помилок (Confusion Matrix) дозволяє візуально побачити сильні та слабкі сторони моделі, зокрема типи помилок, які вона найчастіше робить. Комплексний підхід до оцінювання результатів дає можливість більш об’єктивно визначити ефективність моделі у задачі виявлення

фейкових новин і сприяє вибору оптимального підходу для її подальшого вдосконалення.

2.5 Проектування веб-застосунку

Для реалізації системи виявлення фейкових новин було спроектовано веб-застосунок, який поєднує роботу моделей машинного навчання з інтерактивним інтерфейсом користувача. Архітектура застосунку є багаторівневою та складається з кількох основних компонентів: API з моделями, бекенду, бази даних і фронтенду. Такий підхід забезпечує гнучкість, масштабованість і зручність у подальшому розширенні функціоналу системи. Візуалізація архітектури веб-застосунку для магістерської роботи наведена на рисунку 2.9.

Як бачимо з рисунку архітектури веб-застосунку інтерфейс, тобто фронтенд частина звертається до апі, яка створена з використанням NestJs. Ця апі в свою чергу уже звертається до бази даних, а також до ще однієї апі, яка передає введені користувачем тексти для аналізу виявлення фейкових новин моделлю машинного навчання.

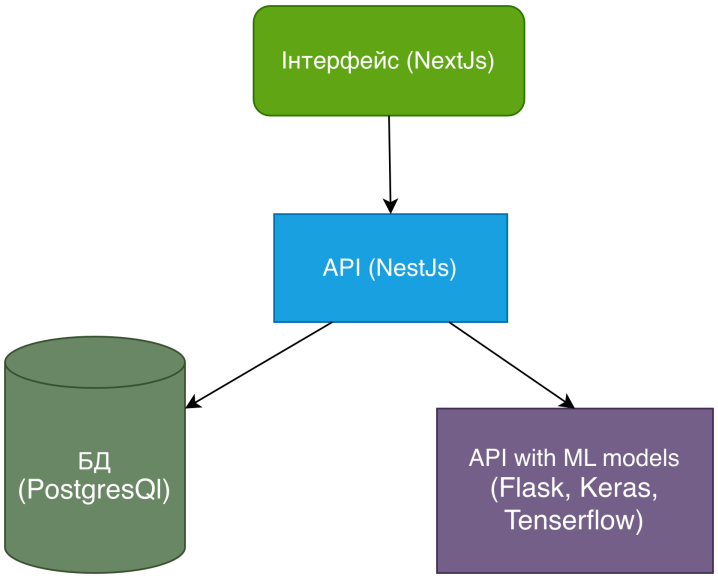


Рисунок 2.9 - Схематичне зображення архітектури веб-застосунку

API з моделями реалізовано на Python із використанням фреймворку Flask. Його основне завдання полягає в обробці вхідних текстів, передачі їх у модель машинного навчання та поверненні результатів класифікації, тобто визначення, чи є новина фейковою. Такий окремий модуль дає змогу ізолювати логіку роботи моделей від решти системи, що полегшує оновлення чи заміну моделей без впливу на інші компоненти.

Бекенд, розроблений за допомогою NestJS, виступає як посередник між фронтендом і API з моделями. Він відповідає за маршрутизацію запитів, керування користувацькими сесіями, обробку даних перед відправкою до моделі та збереження результатів у базі даних. Це забезпечує стабільність, безпеку та централізовану логіку взаємодії в системі.

Для зберігання даних використовується PostgreSQL, яка виконує роль основної бази даних застосунку. У ній зберігаються тексти новин, результати класифікації, статистика перевірок, а також дані користувачів. Використання реляційної бази даних дозволяє ефективно організувати доступ до даних, підтримувати зв’язки між таблицями та забезпечує високу надійність зберігання.

Фронтенд реалізовано за допомогою Next.js, що дозволяє створити сучасний, зручний та швидкий інтерфейс користувача. Через нього користувач може вводити новини для перевірки, переглядати результати аналізу, а також взаємодіяти з іншими функціями системи. Використання цієї технології забезпечує плавну взаємодію між клієнтом і сервером, а також підтримку SEO-оптимізації та швидке завантаження сторінок.

У результаті спроектований веб-застосунок є повноцінною системою, яка поєднує потужність моделей машинного навчання з інтуїтивно зрозумілим інтерфейсом, забезпечуючи зручний доступ користувачів до інструменту для виявлення фейкових новин у реальному часі.

РОЗДІЛ 3.
РЕЗУЛЬТАТИ ВИРІШЕННЯ ЗАДАЧІ

3.1. Опис набору даних

Для реалізації системи виявлення фейкових новин, одним із ключових етапів є формування та аналіз якісного набору даних, на основі якого відбувається навчання та тестування моделей машинного навчання. У цьому дослідженні використано відкритий набір даних, що містить приблизно 10700 заголовків новин, пов’язаних із українськими новинами, опублікованих у період з 24 лютого по 11 грудня 2022 року. Даний датасет є одним із найбільших публічно доступних наборів україномовних новин, спеціально створених для задач класифікації правдивих і фейкових повідомлень [23].

Новини були зібрані з низки українських і російських телеграм-каналів. Зокрема, до перевірених (достовірних) джерел належать такі канали, як “Суспільне Новини”, “Perepichka NEWS”, “NR” та “Новини України Online”. Для формування фейкової частини вибірки використано публікації з каналів, відомих поширенням недостовірної інформації, таких як “Вокс Україна” та “Війна з фейками”. Таким чином, набір даних включає збалансовану репрезентацію обох класів новин, що дозволяє навчити моделі розрізняти правдиві та неправдиві повідомлення.

Кожен запис у наборі даних містить два основні атрибути, а саме текст що є заголовок новини у текстовому форматі та певна мітка, яка визначає тип новини, а саме чи вона правдива (True) та чи фейкова (False). Приклад набору даних наведений у таблиці 3.1

В даному наборі даних загальна кількість прикладів становить 10 735 записів, з яких 8237 мають мітку “True”, а 2 498 мають “False”. Така різниця між кількістю позитивних і негативних прикладів створює

дисбаланс класів, що може негативно вплинути на навчання моделей, оскільки вони схильні “запам’ятовувати” частіший клас і ігнорувати рідкісний, тобто модель буде робити точні прогнозу для одного класу, а для іншого ні.

Таблиця 3.1. - Огляд прикладів з набору даних

Text (Новина)	Label(Мітка)
рубль став найнестабільнішою валютою всьому світі	True
«Зеленський підписав закон продовження терміну дії військового стану Україні до березня»	True
Телефони призводять виникнення раку повідомляли російські закордонні змі	False
російську військову техніку розкрадають продають онлайндошках оголошень таку новину опублікували українські телеграмканали	False
«звязок вбиває птахів повідомляли іноземні видання нідерландах місцеві блогери році які виявили біля парку гаазі десятки загиблих пернатих»	False

Тому для підвищення ефективності навчання та забезпечення рівномірного навчання при класифікації новин набір даних було збалансовано до однакової кількості записів для кожного класу. Це дозволило уникнути переваги одного класу над іншим і покращити узагальнюючу здатність моделей.

На Рисунок 3.1 з першої діаграми, яка є до збалансуванн, можна побачити, що правдивих новин було більше ніж втричі більше ніж фейкових, тому це б негативно вплинуло на модель, оскільки, тоді б вони краще визначали правдиві новини, а от на фейкових робили б більше помилок.

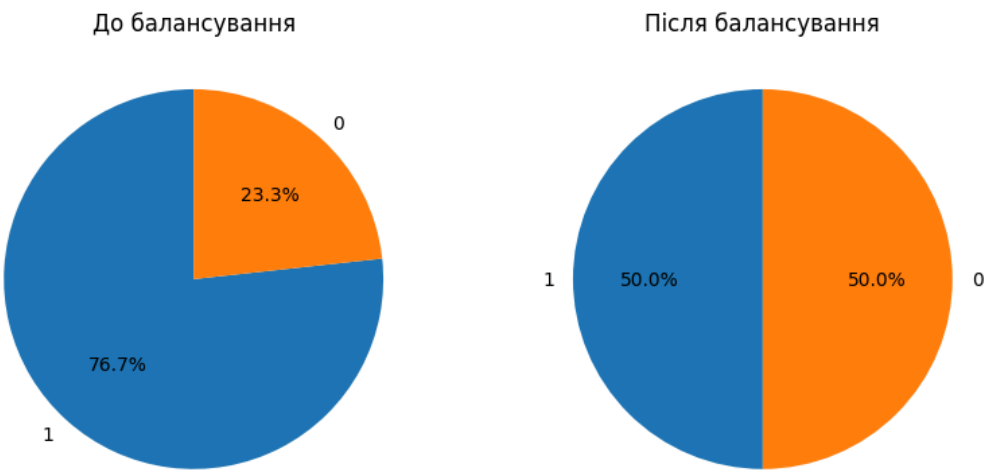


Рисунок 3.1 - Кругові діаграми розподілу даних до збалансування та після

Цей набір даних було створено з метою використання у навчальних і науково-дослідницьких проектах для класифікації новин за ступенем достовірності. Його структура є зручною для обробки алгоритмами машинного навчання, оскільки містить компактні текстові заголовки, які не потребують складної попередньої сегментації. Такий набір дозволяє проводити ефективне навчання моделей та аналіз їх точності на реальних інформаційних прикладах, що є важливим кроком у розробці систем для автоматичного виявлення фейкових новин у медіапросторі.

3.2. Опис програмної реалізації

Головним завданням цієї магістерської роботи є написання програмної реалізації, а саме навчання моделей на основі підготовленого набору даних та реалізація простого веб-застосунку для використання та тестування попередньо навчиних моделей машинного навчання. Основна ідея програми є те що користувач може вводити певний текст, тобто новину, а програма робить передбачення на основі обраної моделі чи ця новина є правдива чи фейкова. Також в даній програмі можна увійти як адміністратор і тоді він буде могти переглядати новини, які вводили

користувачі та як їх класифікувала певна модель. На основі введених новин адміністратор може обирати певні новини та перенавчати моделі.

Основними етапами для досягнення поставлених цілей програмної реалізації є:

✔ Знайти та завантажити відповідний набір даних та підготувати його до навчання моделей.

✔ Обрати та навчити моделі машинного навчання, обравши найкращі гіпер параметри.

✔ Обрати та навчити нейронні мережі (RNN, LSTM, Bert), створивши відповідні їм архітектури з найкращими параметрами.

✔ Проаналізувати отримані результати на основі декількох метрик.

✔ Реалізувати апі Flask(python) для використання попередньо навчених моделей та можливості їх перенавчання.

✔ Створити базу даних з відповідними табличками для програмної реалізації.

✔ Реалізувати бекенд частину для програмної реалізації, яка б конектиться до бази даних, а також надає апі для використання програми, а також робить запити до апі з моделями.

✔ Імплементувати фронтенд частину з простим користувацьким інтерфейсом.

Отже, для успішної реалізації програми було визначено основні етапи та компоненти для виявлення фейкових новин. Реалізація проекту охоплює повний цикл від підготовки даних і навчання моделей до створення веб-застосунку з інтерактивним інтерфейсом. Такий підхід забезпечує можливість не лише автоматично класифікувати новини, але й гнучко керувати моделями та вдосконалювати їх у процесі використання системи.

Для навчання моделей було обрано набір даних, який описано у пункті 3.1. Після його завантаження та балансування даних було проведено попередню обробку даних, яка включала: перетворення тексту до

нижнього регістру, очищення тексту від небажаних символів та видалення стоп-слів, а також токенізацію та векторизацію, про що детальніше розписано у пункті 2.1. Для кращого розуміння цих етапів наведено рисунки 3.2 та 3.3.

Text
Просто слухайте цей діалог. Ні, це не нарізка і фільм Тарантіно.
Рубль став найнестабільнішою валютою у всьому світі
Перше звернення мера Мелітополя Івана Федорова після звільнення
Росія загрожує Боснії "українським сценарієм" через її можливе вступ до НАТО.
Енергоатом повідомив, що окупанти пошкодили високовольтну лінію електропередач Запорізька АЕС — Каховська.

Рисунок 3.2 - Текст до попередньої обробки

Text
просто слухайте діалог ні нарізка фільм тарантіно
рубль став найнестабільнішою валютою всьому світі
перше звернення мера мелітополя івана федорова після звільнення
росія загрожує боснії українським сценарієм її можливе вступ нато
енергоатом повідомив окупанти пошкодили високовольтну лінію електропередач запорізька аес каховська

Рисунок 3.3 - Текст після попередньої обробки

Також для навчання RNN та LSTM моделей текст було перетворено до розмірності у 32 токени, щоб забезпечити однакову довжину вхідних послідовностей для нейронної мережі та спростити процес пакетної обробки даних. Дана цифра було обрана, оскільки саме вона є найоптимальнішою для даного набору даних, що можна побачити на рисунку №3.4, де зображено гістограма для кількості слів у текстах. У результаті підготовки даних було сформовано збалансований та очищений набір, придатний для подальшого навчання моделей. Тексти було перетворено до послідовностей фіксованої довжини у 32 токени, що забезпечило уніфікацію вхідних даних та оптимізувати процес навчання нейронних мереж.

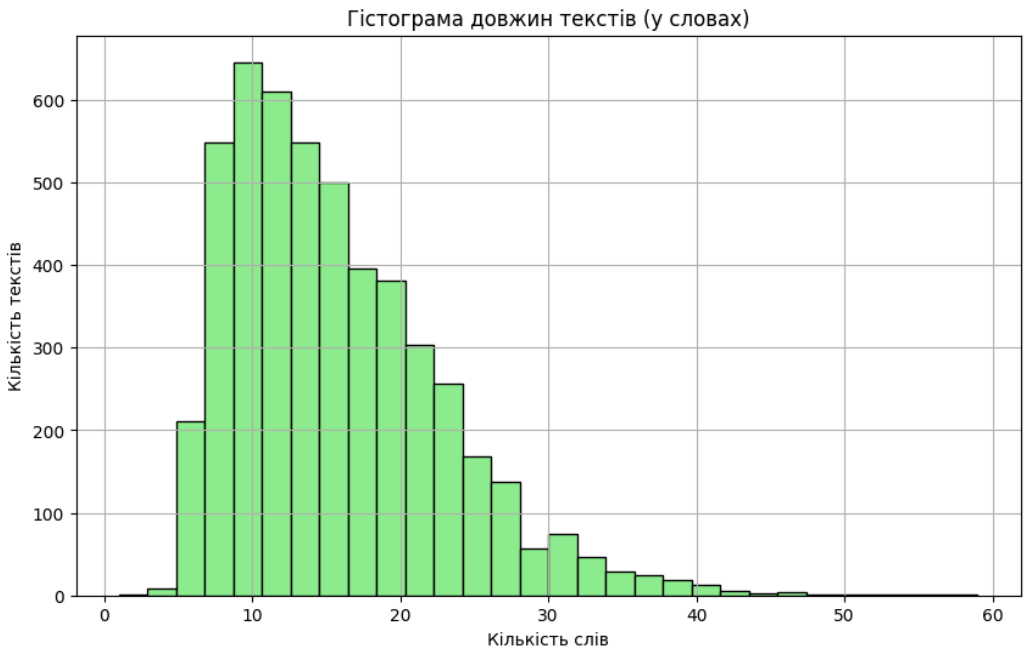


Рисунок 3.4 - Гістограма відповідності кількості токенів до кількості текстів у наборі даних

Для реалізації та навчання моделей машинного навчання було використано бібліотеку Scikit-learn (sklearn), яка надає широкий набір інструментів для класифікації, регресії та оцінювання якості моделей. Навчання здійснювалося на попередньо підготовлених даних, що були розділені на навчальну та тестову вибірки у співвідношенні 80:20.

У рамках магістерської роботи було побудовано та навчено кілька моделей: Логістична регресія, Дерево рішень, Метод опорних векторів, Градієнтне підсилення про які детальніше описано у розділі 2. Для кожної моделі було проведено навчання з подальшим оцінюванням якості на тестовій вибірці за певними метриками. Отримані результати дозволили порівняти ефективність моделей та визначити найбільш придатний алгоритм для поставленої задачі.

Крім звичайних методів машинного навчання, в магістерській роботі були використані рекурентні нейронні мережі та модель трансформер Bert. Для їх реалізації було створено відповідні архітектури для цих моделей та навчено їх з використанням найкращих гіпер параметрів.

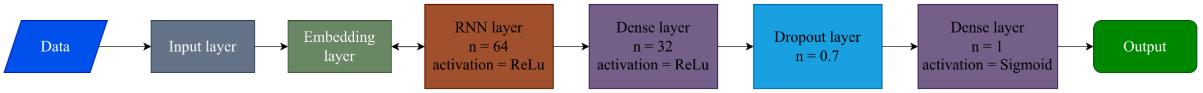


Рисунок 3.5 - Архітектура RNN моделі

Як бачимо з рисунка 3.5 архітектура RNN моделі є послідовною нейронною мережею для обробки тексту. Спочатку йде шар Embedding, що перетворює токени у вектори розмірності 16. Потім шар RNN із 64 нейронами аналізує послідовність та виділяє її контекстні ознаки. Далі йде повнозв'язний шар Dense на 32 нейрони з активацією ReLU для подальшої обробки ознак та шар Dropout (ймовірність 0.7), який зменшує перенавчання. Фінальний Dense(1, sigmoid) виконує бінарну класифікацію, повертаючи ймовірність належності до певного класу. Схожу архітектуру має Lstm модель, яка зображена на рисунку №3.6, де тільки відрізняється шар LSTM.

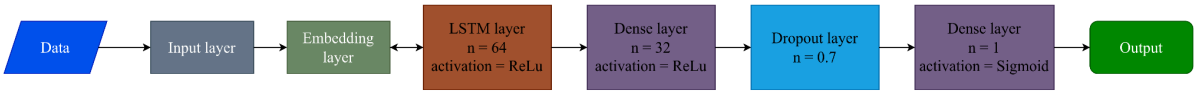


Рисунок 3.6 - Архітектура Lstm моделі

Архітектура моделі на Bert основі XLM-RoBERTa, яка зображена на рисунку №3.7 представляє собою попередньо натренований багатомовний трансформер, пристосований для задачі класифікації текстів. На вхідній стадії модель приймає токенізований текст, який перетворюється у векторні подання за допомогою вбудованого ембеддинг-шару. Далі дані проходять через 12 шарів енкодера трансформера, де застосовуються механізми "багатоголової самоуваги" та feed-forward мережі, що дозволяють моделі враховувати контекст усієї послідовності одночасно. Після цього формується вихідне представлення спеціального токена, яке

подається на класифікаційну “голову”, що складається з шару Dropout та фінального Dense-шару, який генерує два логіти. Оскільки використовується функція втрат SparseCategoricalCrossentropy, логіти надходять у loss-функцію без попереднього застосування softmax, а оцінка невідповідності класів відбувається всередині самої функції втрат.

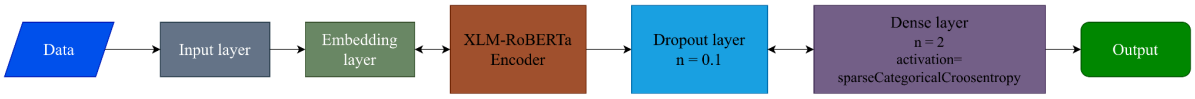


Рисунок 3.7 - Архітектура Bert моделі

У межах експерименту було обрано модель XLM-RoBERTa, оскільки вона демонструє кращі результати порівняно з класичною BERT. На відміну від BERT, XLM-RoBERTa тренувалася на значно більшому та різноманітнішому багатомовному корпусі (2,5 ТБ тексту проти 16 ГБ у BERT), що робить її стійкішою до мовних варіацій, граматичних конструкцій та рідковживаних лексем.

В результаті бачимо, що архітектура RNN базується на послідовній передачі станів, що дозволяє моделі запам’ятовувати попередні кроки, однак вона обмежена короткими залежностями та схильна до затухання градієнтів. LSTM удосконалює цю структуру за рахунок комірки пам’яті та керуючих ворот, що дає змогу ефективніше працювати з довшими послідовностями. На відміну від рекурентних мереж, архітектура BERT ґрунтується на механізмі “багатоголової самоуваги”, який дозволяє моделі аналізувати всі елементи тексту паралельно й враховувати глобальний контекст. Завдяки цьому трансформери демонструють значно вищу гнучкість та здатність до глибокого розуміння структури тексту.

Для виявлення фейкових новин було навчено декілька моделей машинного навчання, а саме модель Логістичної регресії, Дерева рішень,

методу опорних векторів та гадієетного спуску, а також моделі нейронних мереж, такі як RNN та Lstm і трансформерну модель Bert(Roberta). В результаті навчання цих моделей на описанову в пункті 3.1 наборі даних було отримано наступні результати, які знаходяться в таблиці 3.2.

Таблиця 3.2- Отримані результати

Модель / Метрика	Accuracy	Precision	Recall	F1-score
Logistic regression	0.85	0.86	0.85	0.85
Decision Tree	0.81	0.81	0.8	0.8
SVC	0.85	0.85	0.85	0.85
Gradient Boosting	0.83	0.82	0.82	0.82
RNN	0.86	0.86	0.85	0.85
LSTM	0.86	0.86	0.86	0.86
Bert	0.89	0.89	0.89	0.89

З отриманих видно, що всі реалізовані моделі показали досить високу точність у задачі виявлення фейкових новин. Класичні алгоритми машинного навчання, такі як логістична регресія, дерево рішень, SVC та градієнтне підсилення, продемонстрували прийнятний рівень точності, причому SVC та логістична регресія показали кращі результати серед класичних методів. Рекурентні моделі, RNN та LSTM, показали дещо кращу точність у метриках, особливо LSTM, яка завдяки своїй архітектурі з комірками пам’яті ефективніше враховує довгі залежності в тексті.

Найвищі результати були досягнуті трансформерною моделлю Bert (точніше, XLM-RoBERTa), яка забезпечила точність та значення всіх метрик на рівні 0.89. Це підтверджує, що сучасні трансформерні моделі

мають кращу здатність до узагальнення та глибшого розуміння контексту тексту порівняно з класичними методами та рекурентними мережами. Високі показники F1-score у Bert свідчать про збалансованість моделі між точністю та повнотою, що особливо важливо для задач класифікації фейкових новин.

Загалом, проведений експеримент показав, що використання нейронних та трансформерних моделей значно підвищує ефективність виявлення неправдивої інформації. Класичні методи можуть бути використані для швидких або менш ресурсомістких рішень, проте для досягнення максимальної точності доцільно застосовувати LSTM або Bert. Результати також підтвердили, що глибинні моделі добре справляються з контекстною інформацією, що робить їх оптимальними для складних текстових задач.

В даній магістерській роботі передбачено, що моделі можна переіреновувати на даних, які були ведені користувачем та адмін їх погодив для перенавчання моделі. В ході цієї роботи було, також, перетреновано модеоі на користувацьких даних, які були відібрані. Ось декілька їх прикладів:

Правдиві:

- ✔ Уряд збільшив виплати для сімей із дітьми.
- ✔ Уряд оголосив про підвищення мінімальної зарплати з наступного місяця.
- ✔ У школах запровадили нову систему харчування для дітей.

Фейкові:

- ✔ Вчені навчилися копіювати пам’ять людини.
- ✔ На Місяці помітили величезні підземні міста.
- ✔ Коти навчилися розуміти понад 3000 людських слів.

В загальному моделі було перенавчено на 1000 нових записах, які були розподілені порівну 500 справжніх новин та 500 фейкових. Після

перенавчання всіх моделей було отримано наступні результати, які наведені в таблиці 3.3.

Таблиця 3.3. - Отримані результати після перенавчання

Модель / Метрика	Accuracy	Precision	Recall	F1-score
Logistic regression	0.87	0.87	0.87	0.87
Decision Tree	0.81	0.81	0.81	0.81
SVC	0.87	0.87	0.86	0.86
Gradient Boosting	0.86	0.86	0.85	0.85
RNN	0.87	0.87	0.87	0.87
LSTM	0.88	0.88	0.88	0.88
Bert	0.90	0.90	0.90	0.90

З отриманих результатів бачимо, що модель добре здатна для перенавчання і в середньому покращила показники на 1-2% по всіх материках в усіх перенавчених моделей, що говорить про високу здатність до узагальнення та стабільну роботу на нових даних. Особливо це помітно для SVC та Логістичної регресії, а також моделей нейронних мереж, таких як RNN та LSTM. А от для Bert вдалося підняти точність аж до 90%, що є дуже високим показником.

Загалом, перенавчання на нових даних підтверджує, що комбінування класичних методів машинного навчання та сучасних нейронних архітектур дозволяє досягти високої ефективності моделі у реальних умовах. А також показує, що модель можна вдосконалювати в майбутньому на нових даних, що є дуже важливим для підтримки актуальності системи в умовах швидких змін інформаційного потоку. Це дозволяє адаптувати модель до

нових типів фейкових та правдивих новин, покращувати її здатність до узагальнення та зменшувати кількість помилкових класифікацій. Крім того, результати демонструють потенціал для подальшої інтеграції моделей глибинного навчання з класичними методами для підвищення точності та швидкості обробки великих обсягів текстових даних. Такий підхід забезпечує надійний інструмент для автоматизованого моніторингу та перевірки достовірності новин у реальному часі, що особливо актуально в умовах інформаційної насиченості та поширення дезінформації.

3.3. Порівняння результатів

Після навчання та перенавчання моделей були отримані різні результати і у цьому розділі проведено детальний аналіз і порівняння результатів різних моделей машинного та глибинного навчання до та після перенавчання на розширеному наборі даних. Такий аналіз дозволяє оцінити вплив додаткових даних на загальну якість класифікації фейкових новин і визначити, які моделі найбільше виграють від збільшення обсягу тренувальної вибірки.

Порівнюючи дані з Таблиць 3.1 та 3.2, можна помітити систематичне покращення точності майже в усіх моделей, що візуально можна побачити на рисунку №3.8.

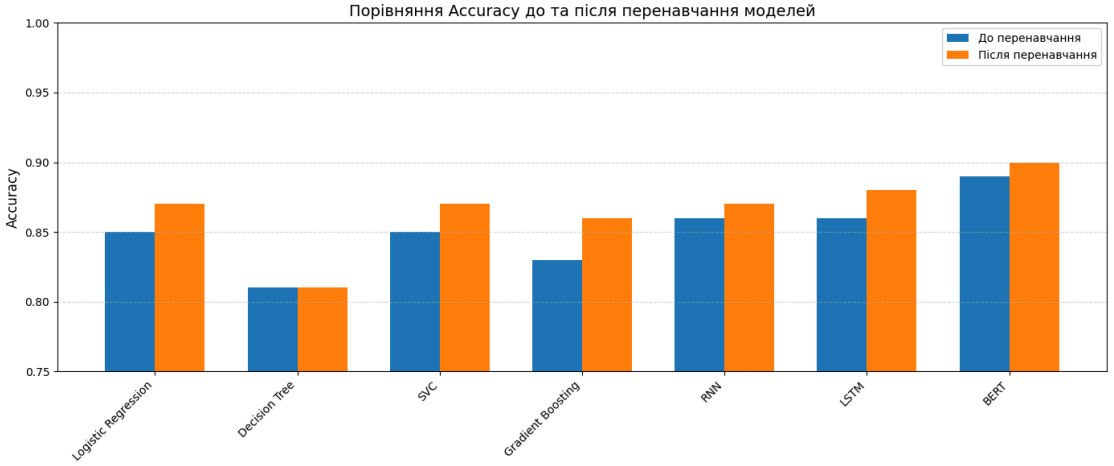


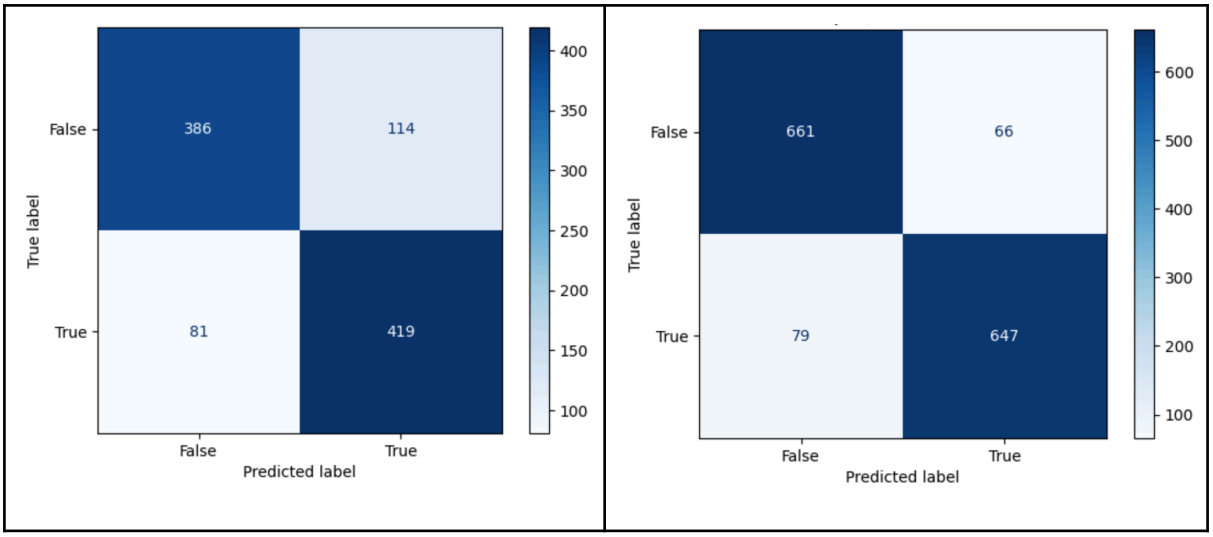
Рисунок 3.8 - Візуалізація отриманих результатів

Зокрема, Логістична регресія підвищила точність з 0.85 до 0.87, що свідчить про чутливість моделі до додаткових тренувальних прикладів. Аналогічно метод SVC, який спочатку мав accuracy 0.85, після перенавчання покращився до 0.87, демонструючи ефективну здатність узагальнювати інформацію на розширеному наборі даних.

Ансамблеві моделі також продемонстрували позитивну динаміку. Gradient Boosting покращив свої метрики приблизно на 3%, досягнувши 0.86 accuracy, що підтверджує перевагу ансамблевих підходів при збільшенні обсягу даних. Рішення деревом, однак, залишилося практично на тому ж рівні 0.81, що можна пояснити схильністю моделі до перенавчання та обмеженою здатністю працювати з високовимірними текстовими ознаками.

Нейронні мережі також показали суттєве зростання. RNN, який спочатку мав точність 0.86, після перенавчання піднявся до 0.87, а LSTM, що вже демонстрував добру роботу з послідовностями текстів, збільшив точність з 0.86 до 0.88. Це свідчить про те, що моделі, здатні вловлювати контекст і залежності в тексті, суттєво виграють від збільшення кількості тренувальних даних.

Найкраща динаміка спостерігається у трансформерної моделі BERT, яка і до перенавчання мала найвищі результати (0.89 accuracy), і після розширення датасету досягла 0.90 accuracy. Це ще раз підтверджує високу ефективність сучасних моделей, побудованих на механізмах self-attention, у задачах мовної обробки та класифікації текстів. Також наведена матриці плутанини на рисунках №3.9 та №3.10 для найгіршого методу, тобто дерева рішень, оскільки у нього найгірша точність та найкращого, відповідно Bert, який досягнув 90% точності.



а) б)

Рисунок 3.9 – Матриця плутанини для (а) методу дерева рішень до перенавчання; (б) для Bert після перенавчання

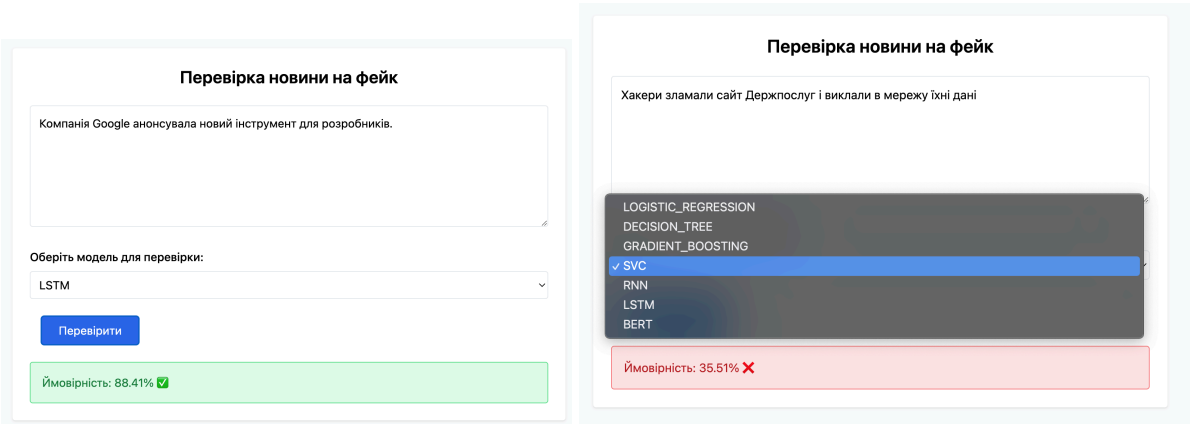
З даних матриць можна побачити що в результаті магістерської роботу вдалося дуже добре покращити результати обравши Bert в порівнянні з методом дерева рішень. Ці дані демонструють явне покращення. У випадку дерева рішень спостерігається велика кількість помилково класифікованих прикладів (114 та 81 випадки), тоді як BERT значно зменшив ці помилки (66 та 79 випадків відповідно). Таким чином, використання трансформерної моделі BERT у рамках магістерської роботи дозволило досягти значного прогресу у якості класифікації, і цей результат може стати надійною основою для подальших досліджень або практичних застосувань.

Аналіз результатів до та після перенавчання показав, що збільшення обсягу тренувальної вибірки позитивно впливає на якість класифікації для більшості моделей. Найбільше виграли від розширення даних сучасні нейронні мережі, зокрема трансформерна модель BERT, яка продемонструвала найвищу точність та значно зменшила кількість помилок. Традиційні методи, такі як дерево рішень, виявилися менш гнучкими та схильними до помилок, що обмежує їх ефективність на складних текстових даних. Отримані результати підтверджують перевагу

глибинних моделей для обробки природної мови та показують, що кількість даних суттєво впливає на точність класифікації фейкових новин. Це створює міцну основу для подальшого вдосконалення моделей та практичного використання в завданнях автоматичної обробки текстів.

3.3 Опис веб-застосунку

Веб-застосунок складається з двох основних частин, а саме користувацької частини, де він може вести певний текст(новину), обрати модель та перевірити її на достовірність і у результаті получити відсотки достовірності даної новини, що візуалізовано рисунку 3.10.



а) б)

Рисунок 3.10 - Інтерфейс користувацької частини додатку

Як бачимо з рисунків перший випадок був класифікований як правдий з використанням lstm, а от другий як фейковий з використанням класифікатора опорних векторів.

Ще одна важлива частина веб-застосунку є адмін панель, де адміністратори можуть перегляди всі введені користувацькі запити на перевірку новин, а також встановлювати їм свою мітку(чи правдива чи фейкова новина), оскільки інколи моделі можуть помилятися та відповідно зберігати користувацькі дані та перенавчати моделі, що є зображено на рисунку 3.11.

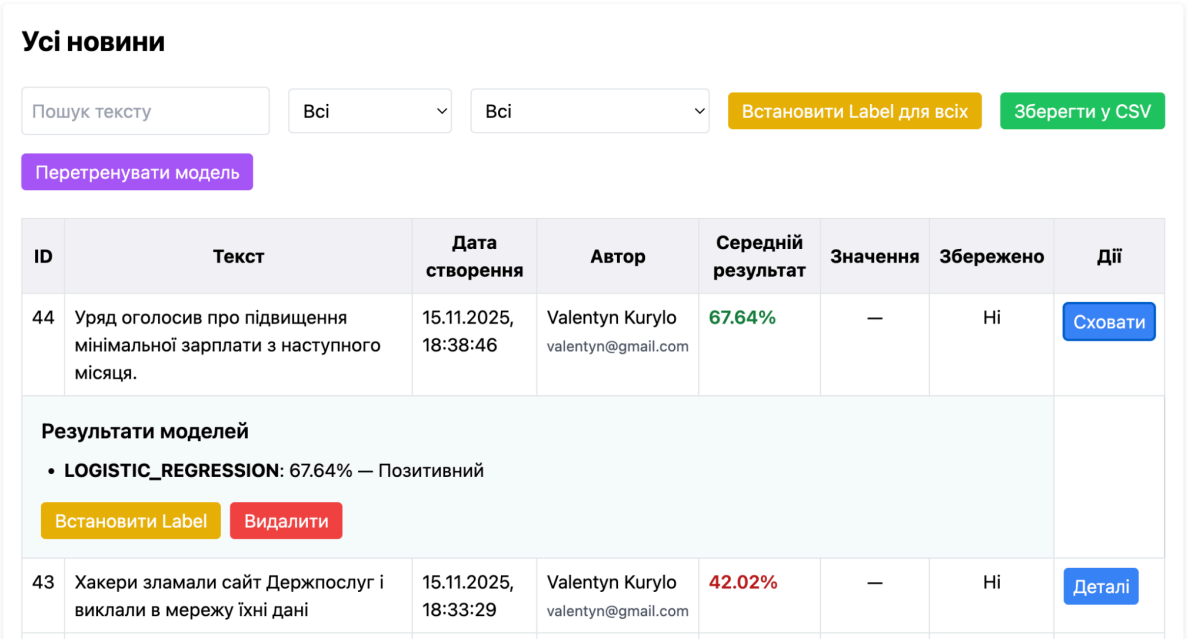


Рисунок 3.11 - Інтерфейс адмін панелі

Також веб інтерфейс включає реєстрацію та логінаю черех емейл та пароль, де пароль відповідно хешується. Для побудови фронтенд часитини був використаний NextJs, а для бекенд частини NestJs та база даних PostgresQl. також з бекенду відправляються запити на ще одне апі, яке побудована з використанням Python(Flask), де знаходяться модель машинного навчання, для доступу використовується приватний ключ.

Була використана реляційна база даних, оскільки вона забезпечує чітку структурованість даних та підтримує складні взаємозв’язки між таблицями. Вона гарантує цілісність даних і дозволяє ефективно виконувати складні запити. Структури таблиць бази даних можна побачити на рисунку 3.12. Відповідно до цієї структури були створені контролери та сервіси на бекенд частині, щоб фронтенд міг з ним спілкуватися, для хорошого розуміння апі була створена документацію у Swagger. Більш детально все можна переглянути за посиланнями на gitHub, що знаходяться у додатку Б.

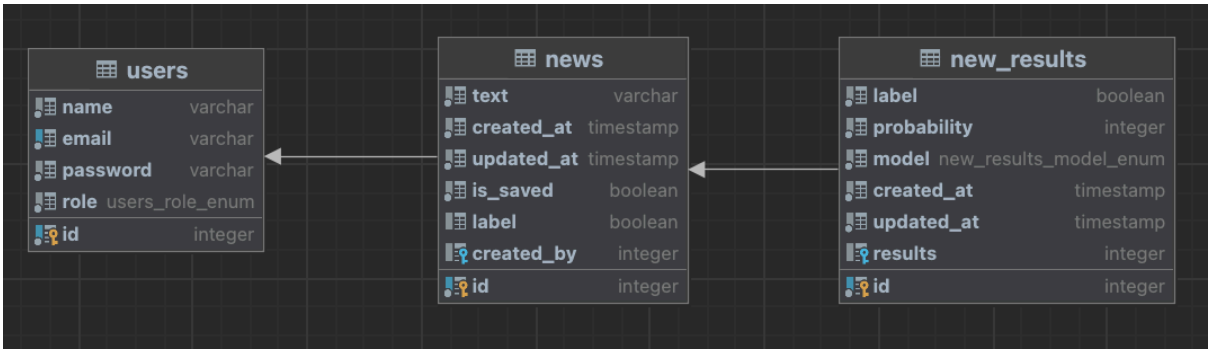


Рисунок 3.12 - Структура бази даних

У підсумку розроблений веб-застосунок демонструє цілісну архітектуру, що поєднує інструменти машинного навчання, сучасні веб-технології та надійну систему керування даними. Користувацька частина дозволяє легко перевіряти новини на достовірність та отримувати результати роботи різних моделей, що підвищує зручність взаємодії з системою. Адмін-панель забезпечує контроль якості даних, даючи можливість переглядати введені запити, виставляти мітки та ініціювати перенавчання моделей у разі потреби. Використання реляційної бази даних та чітко структурованого бекенду сприяє стабільності й масштабованості всієї системи. Загалом застосунок є гнучким, розширюваним та готовим до подальшого вдосконалення.

Програмна реалізація цієї магістерської роботи базується на використанні сучасних інструментів для машинного навчання та створення повноцінного веб-застосунку. Для розробки моделей класифікації текстів було обрано мову програмування Python, оскільки вона є стандартом у галузі аналізу даних, обробки природної мови та побудови нейронних мереж. Python надає великий вибір бібліотек для попередньої обробки текстів, роботи з наборами даних, оптимізації моделей та їх оцінювання. Крім того, простота синтаксису та велика кількість готових інструментів дозволяє швидко та ефективно реалізувати всі етапи створення моделей машинного навчання.

Для розробки веб-частини застосунку використовувалися сучасні технології фронтенду та бекенд-розробки, що забезпечили зручну інтеграцію з API моделей, захищену авторизацію, зберігання даних і можливість адміністрування. Зокрема, використання таких інструментів як Next.js, Redux, Tailwind CSS, NestJS, PostgreSQL, TypeORM, bcrypt, JWT, axios та Swagger дало змогу створити повноцінну інфраструктуру для розгортання веб-застосунку, взаємодії з моделями та керування даними користувачів.

Засоби для машинного навчання, обробки даних та побудови апі для моделей (Python, Flask):

✔os використовується для взаємодії з файловою системою, зокрема для завантаження, структурування та збереження наборів даних.

✔json забезпечує роботу з форматом JSON для обміну структурованими даними та збереження метаданих.

✔csv використано для читання та запису даних у CSV-форматі при підготовці датасету.

✔NumPy бібліотека для роботи з багатовимірними масивами, що також містить оптимізовані математичні операції.

✔Pandas основний інструмент для аналізу та обробки даних, форматування таблиць, фільтрації та попереднього очищення текстів.

✔NLTK (Natural Language Toolkit) застосовувався для попередньої обробки текстів: токенізації, лематизації, видалення стоп-слів та підготовки текстів до векторизації.

✔scikit-learn забезпечив широкий набір інструментів для побудови класичних моделей машинного навчання, їх навчання, векторизації текстів, розбиття вибірок, оптимізації гіперпараметрів і обчислення метрик.

✔TensorFlow використаний для побудови та тренування нейронних мереж, зокрема рекурентних моделей для класифікації фейкових новин.

✔Keras більш високорівневий інтерфейс над TensorFlow, що спростив

створення архітектур рекурентних моделей.

✔Transformers (Hugging Face) бібліотека для роботи з моделями трансформерів, включно з BERT. Вона забезпечила можливість швидкого завантаження попередньо нетренованих моделей та їх навчання.

✔Flask це Python-фреймворк, використаний для створення окремого API-сервера, який забезпечує доступ до попередньо навчених моделей машинного навчання. Flask дозволив швидко організувати маршрути для класифікації новин, обробки запитів від бекенду, а також реалізувати можливість перенавчання моделей на нових даних.

✔ Також, розглянемо засоби для веб-розробки(TypeScript):

✔ Next.js фреймворк для фронтенд-розробки, що забезпечив серверний рендеринг, швидку роботу та зручну інтерфейсну взаємодію з API моделі.

✔ Redux використано для керування станом застосунку, зокрема для зберігання результатів класифікації, даних користувача та глобальних параметрів програми.

✔ Tailwind CSS надав гнучкі стилі та забезпечив швидку розробку адаптивного та сучасного UI без написання великої кількості власних CSS-класів.

✔ NestJS бекенд-фреймворк, що дозволив створити структурований серверний застосунок із модульною архітектурою, реалізувати REST API та організувати логіку авторизації, збереження даних і взаємодії з ML-моделями.

✔ PostgreSQL реляційна база даних, використана для зберігання інформації про користувачів, новини, історію класифікацій та дані для потенційного перенавчання моделей.

✔ TypeORM ORM-інструмент для роботи з PostgreSQL, який спростив створення таблиць, міграцій і взаємодію з базою даних.

✔ bcrypt бібліотека для хешування паролів користувачів,

забезпечила високий рівень безпеки системи авторизації.

✔ JWT (JSON Web Tokens) механізм генерації та перевірки токенів для аутентифікації та авторизації користувачів у веб-застосунку.

✔ axios HTTP-клієнт, що застосовувався для комунікації фронтенду з бекендом, а також бекенду з Python-API моделей.

✔ Swagger (OpenAPI) використовувався для автоматичної генерації документації до REST API застосунку, тестування запитів і полегшення інтеграції.

У цьому розділі було детально розглянуто засоби та технології, застосовані під час розробки програмної системи для виявлення фейкових новин. Описані інструменти Python забезпечили ефективну роботу з даними, навчання моделей машинного навчання та організацію API для взаємодії з ними. Крім того, сучасні веб-технології Next.js, Nest.js, PostgreSQL та інші дали змогу створити повноцінний, зручний і безпечний веб-застосунок. Комплексно використані засоби забезпечують реалізацію надійної системи, що поєднує машинне навчання та веб-інфраструктуру.

РОЗДІЛ 4.
ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

4.1.Обґрунтування можливих чинників травмонебезпечних ситуацій

У процесі розроблення, тестування та експлуатації програмного забезпечення, а також під час організації робочих місць ІТ-фахівців можуть виникати різноманітні травмонебезпечні ситуації. Незважаючи на те, що програмна діяльність відноситься до категорії робіт з низьким рівнем фізичного ризику, вона пов'язана з низкою небезпечних та шкідливих виробничих факторів, які за певних умов можуть стати причиною травм, погіршення здоров’я або аварійних ситуацій.

До основних чинників, що можуть зумовити травмонебезпечні ситуації в ІТ-середовищі, належать:

1. Електробезпека та пошкодження обладнання. Однією з найбільш поширених небезпек є ураження електричним струмом, що може виникати при несправності електромережі, пошкодженні ізоляції кабелів, використанні несертифікованого обладнання або неправильному підключенні техніки. Значні ризики створюють також неякісні подовжувачі та велика кількість пристроїв, підключених до однієї мережі.
2. Пожежонебезпечні фактори. Використання комп’ютерної техніки, блоків живлення, мережевого обладнання, електронагрівальних елементів та периферії створює потенційну пожежну небезпеку, особливо за умов порушення теплового режиму, перегрівання техніки, коротких замикань або недотримання правил експлуатації.
3. Ергономічні та фізіологічні ризики. Тривала робота за комп’ютером у незручній позі, недостатнє освітлення, неправильно підібране крісло чи монітор призводять до порушення опорно-рухового

апарату, напруження зору, головних болей та загального виснаження. Такі фактори не спричиняють миттєвих травм, але здатні викликати професійні захворювання та хронічні проблеми зі здоров'ям.

4. Психофізіологічні фактори. Високий рівень інтелектуального навантаження, багатозадачність, робота з великими обсягами даних та жорсткі дедлайни можуть спричиняти стрес, нервове перенапруження, емоційне вигорання, що негативно впливає на здатність адекватно реагувати на ризики та підвищує ймовірність помилок, зокрема тих, що можуть призвести до аварійних ситуацій.
5. Організаційні порушення та недотримання правил безпеки. Відсутність інструктажу, нехтування нормами охорони праці, неправильне розміщення обладнання, захаращення проходів, використання пошкоджених меблів або техніки також формують передумови для травмування працівників.

Таким чином, навіть у середовищі, де основна діяльність має інтелектуальний характер, існує комплекс чинників, що можуть стати причиною травмонебезпечних ситуацій. Усвідомлення та аналіз цих факторів є необхідною умовою для організації безпечного робочого процесу, профілактики професійних захворювань та запобігання аваріям у процесі розроблення та експлуатації програмного забезпечення.

4.2.Умови та обставини виникнення небезпечних ситуацій та їх наслідки

Виникнення небезпечних ситуацій у сфері інформаційних технологій зумовлене сукупністю організаційних, технічних, експлуатаційних та людських факторів. Хоча ІТ-галузь не є виробництвом із підвищеною фізичною небезпекою, порушення встановлених норм та недотримання вимог охорони праці можуть призвести до серйозних наслідків для здоров'я працівників або працездатності обладнання, що є наведено у таблиці 4.1.

Таблиця 4.1. - Умови та обставини виникнення небезпечних ситуацій та їх наслідки

Група факторів	Умови та обставини виникнення небезпечних ситуацій	Можливі наслідки
Технічні та експлуатаційні	1.Робота з несправним або зношеним обладнанням 2.Пошкоджені кабелі, роз'єми, подовжувачі 3.Перевантаження електромережі 4.Недостатня вентиляція та перегрів техніки.	1.Вихід обладнання з ладу 2.Коротке замикання, пожежа 3.Матеріальні збитки.
Порушення електробезпеки	1.Доторкання до оголених елементів під напругою 2.Робота у вологих умовах 3.Несанкціоновані ремонт електропристроїв 4.Неправильне підключення обладнання.	1.Електротравми 2.Опіки 3.Ураження електричним струмом різного ступеня тяжкості.
Невідповідний мікроклімат та організація робочого місця	1.Порушення норм температури, вологості та освітленості 2.Надмірний шум від техніки 3.Неправильне розташування меблів та обладнання 4.Захаращеність робочої зони.	1.Погіршення самопочуття та працездатності 2.Падіння, спотикання, дрібні травми 3.Перегрів або збій роботи обладнання.
Психофізіологічні фактори (людський фактор)	1.Перевтома, тривала робота без перерв 2.Стресові ситуації та емоційне перенапруження 3.Втрата концентрації уваги 4.Помилки при роботі з технікою.	1.Підвищення числа аварійних ситуацій 2.Неправильні дії з обладнанням 3.Падіння техніки, пошкодження кабелів.
Організаційні недоліки	1.Відсутність інструктажів з охорони праці 2.Невчасне технічне обслуговування обладнання 3.Відсутність засобів пожежогасіння 4.Блокування евакуаційних виходів 5.Порушення працівниками правил безпеки.	1.Ускладнена евакуація під час надзвичайної ситуації 2.Поширення пожежі 3.Підвищення травматизму 4.Значні фінансові втрати та збій роботи організації.

Отже, небезпечні ситуації в ІТ-середовищі виникають унаслідок комбінованої дії технічних, організаційних та людських факторів. Їх попередження вимагає системного підходу, постійного контролю стану робочого середовища та суворого дотримання норм охорони праці.

4.3. Безпека в надзвичайних ситуаціях

Безпека в надзвичайних ситуаціях є важливою складовою організації роботи будь-якого підприємства, у тому числі й у сфері інформаційних технологій, де значну роль відіграє безперервність доступу до обладнання, мережевої інфраструктури та серверних ресурсів. Умови виникнення надзвичайних ситуацій можуть бути пов’язані як із техногенними, так і природними чинниками, а також із загрозами воєнного характеру, що є особливо актуальним в сучасних реаліях України. Використання комплексного підходу до забезпечення безпеки дозволяє мінімізувати ризики, зменшити масштаб можливих наслідків та забезпечити захист життєдіяльності персоналу.

У сфері ІТ особливе значення має безперебійна робота електропостачання, серверного обладнання, засобів зв’язку, систем охолодження та зберігання даних. Аварійне відключення електроенергії, вихід з ладу мережевого обладнання або пошкодження дата-центру можуть призвести до втрати важливої інформації та критичних простоїв. Тому необхідним є використання джерел безперебійного живлення, резервних серверів, дублюючих каналів зв’язку та систем автоматичного відновлення роботи після аварії.

Крім технічних систем, важливим аспектом є підготовленість персоналу. Працівники повинні знати порядок дій у разі виникнення надзвичайних ситуацій: правила евакуації, розташування аварійних виходів, місця розміщення вогнегасників, аптечок, а також мати навички швидкого реагування на сигнали тривоги. Регулярні навчання, інструктажі

та моделювання можливих сценаріїв надзвичайних ситуацій допомагають уникнути паніки та забезпечити швидке і злагоджене реагування.

Особливу увагу також слід приділити захисту інформації та резервному копіюванню даних. У надзвичайних ситуаціях (пожежа, затоплення, руйнування будівлі, кібератака) ризик втрати критичних даних суттєво зростає. Використання систем автоматичного бекапування, зберігання копій на зовнішніх серверах або у хмарних сервісах забезпечує можливість оперативного відновлення роботи підприємства після ліквідації наслідків надзвичайної ситуації.

Отже, ефективна система безпеки в надзвичайних ситуаціях повинна включати технічні, організаційні та інформаційні заходи. Її комплексна реалізація забезпечує захист працівників, мінімізує матеріальні збитки та підтримує стабільну роботу підприємства навіть за умов форс-мажорних обставин. Визначення ефективності розробленої системи є ключовим етапом магістерської роботи, оскільки саме на цьому етапі підтверджується якість побудованих моделей машинного навчання, коректність підготовки даних та працездатність веб-застосунку в реальних умовах. Оцінювання ефективності дозволяє об’єктивно порівняти різні алгоритми, виявити їхні переваги й недоліки, а також вибрати найбільш оптимальний метод для подальшого практичного застосування.

РОЗДІЛ 5.
ВИЗНАЧЕННЯ ЕФЕКТИВНОСТІ

Визначення ефективності розробленої системи є ключовим етапом магістерської роботи, оскільки саме на цьому етапі підтверджується якість побудованих моделей машинного навчання, коректність підготовки даних та працездатність веб-застосунку в реальних умовах. Оцінювання ефективності дозволяє об’єктивно порівняти різні алгоритми, виявити їхні переваги й недоліки, а також вибрати найбільш оптимальний метод для подальшого практичного застосування.

Першим аспектом визначення ефективності є оцінка точності моделей машинного навчання. Для цього використовувалися стандартні метрики класифікації, такі як Accuracy, Precision, Recall та F1-score. Ці метрики дали змогу всебічно проаналізувати роботу алгоритмів у контексті задачі виявлення фейкових новин, де важливим є не лише загальний відсоток правильних передбачень, але й здатність моделі мінімізувати хибнопозитивні та хибнонегативні результати. Для глибшої оцінки був проведений аналіз матриць неточностей та графічна інтерпретація результатів, що дозволило порівняти поведінку моделей на збалансованому наборі даних.

Окрему увагу приділено порівнянню традиційних ML-методів із нейронними мережами та трансформерами. Такі моделі як логістична регресія, метод опорних векторів, дерева рішень та градієнтне підсилення демонстрували базову ефективність, яка була суттєво покращена при використанні RNN-архітектур, LSTM-мереж і особливо трансформерних моделей типу BERT. Це підтвердило перевагу глибоких моделей у задачах обробки природної мови, де контекст та семантика відіграють ключову роль. Для оцінки ефективності моделей обрана саме метрика F1-score, оскільки вона є збалансованим показником, що враховує як точність

(Precision), так і повноту (Recall). У задачі визначення фейкових новин особливо важливо уникати як хибних передбачень, так і пропуску потенційно шкідливих матеріалів, тому використання F1-score дає найбільш об’єктивне уявлення про реальну якість класифікації.Також на рисунку 5.1 є візуалізація результативності моделей по метриці F1-score.

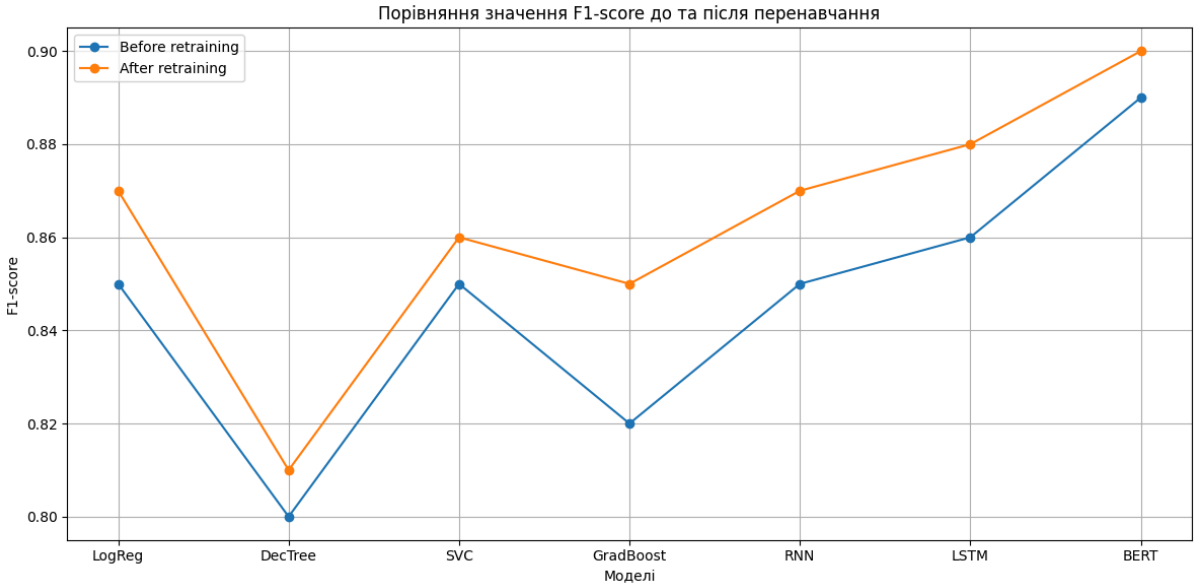


Рисунок 5.1 - Порівняння значення F1-score до та після перенавчання

На поданому графіку відображено порівняння значень метрики F1-score для різних моделей машинного навчання до та після перенавчання на основі нових користувацьких новин. Чітко видно, що після додаткового навчання більшість моделей демонструють покращення, що свідчить про позитивний вплив оновлення даних на їхню здатність коректно класифікувати фейкові та правдиві новини. Найбільше зростання показують моделі Logistic Regression, RNN, LSTM та Bert, які є найстабільнішими у роботі з текстовими даними.

Другим важливим аспектом є оцінювання ефективності веб-застосунку, який виступає інструментом практичного використання моделей. Оцінювалася швидкість обробки запитів, стабільність API на Flask, коректність взаємодії бекенду на NestJS із базою даних PostgreSQL,

а також загальна якість користувацького досвіду. Тестування показало, що система здатна працювати в режимі реального часу, забезпечуючи швидке передбачення та можливість повторного навчання моделей адміністратором.

Третім аспектом є оцінка стійкості та масштабованості системи. Архітектура, побудована з використанням REST API, Flask, NestJS, PostgreSQL, TypeORM та Next.js, дозволяє легко додавати нові моделі, розширювати набір даних і масштабувати систему для більшої кількості користувачів. Це підтверджує її придатність для реального впровадження та подальшого розвитку.

Розроблена система може знайти широке застосування у різних галузях, де важлива швидка та точна оцінка достовірності інформації. Серед основних напрямів використання можна виділити:

- ❖ Журналістика та медіа: автоматичне виявлення фейкових новин і перевірка достовірності інформації, що допомагає редакціям швидше реагувати на поширення неправдивого контенту.
- ❖ Соціальні мережі та платформи контенту: інтеграція системи для фільтрації неправдивих новин у реальному часі, що підвищує якість інформаційного потоку та довіру користувачів.
- ❖ Освітні та аналітичні ресурси: підтримка навчальних та дослідницьких проектів, які вивчають поширення дезінформації, та надання інструментів для практичних завдань із аналізу тексту.
- ❖ Бізнес-аналітика та PR: оцінка достовірності новин у корпоративному середовищі, що допомагає ухвалювати обґрунтовані управлінські рішення.
- ❖ Для майбутніх досліджень система відкриває низку напрямів: інтеграція додаткових мов і джерел даних, застосування більш складних моделей NLP (наприклад, трансформерів наступного покоління), розробка адаптивних алгоритмів для виявлення нових

типів дезінформації, а також оцінка ефективності системи у великих масштабах і в умовах потокових даних. Це дозволить не лише покращити точність, але й забезпечити гнучкість та актуальність системи в умовах швидко змінюваного інформаційного середовища.

Практична користь системи полягає у підвищенні ефективності боротьби з дезінформацією, скороченні часу на перевірку новин та покращенні користувацького досвіду при роботі з інформаційними ресурсами. Вона може стати надійним інструментом для журналістів, аналітиків та організацій, що прагнуть забезпечити якість та достовірність інформації у своїй діяльності.

Підсумовуючи, визначення ефективності продемонструвало, що розроблена система є як технологічно, так і функціонально успішною. Вона забезпечує високу точність виявлення фейкових новин, стабільну роботу веб-інтерфейсу та можливість подальшого масштабування й удосконалення, що підтверджує доцільність застосування запропонованих методів та архітектурних рішень у реальних умовах.

ВИСНОВКИ

У даній магістерській роботі було комплексно досліджено проблему виявлення фейкових новин, яка сьогодні є однією з ключових загроз інформаційному простору. Актуальність теми підтверджується стрімким зростанням обсягу соціальних платформ, збільшенням кількості недостовірної інформації та значним впливом фейкових новин на суспільні настрої, демократичні процеси та безпеку держави. Проведений аналіз літературних джерел, сучасних підходів до NLP-класифікації та існуючих моделей машинного навчання ще раз довів важливість побудови автоматизованих систем для боротьби з дезінформацією.

На основі опрацювання наукових джерел було виконано аналіз предметної області, сформовано математичну постановку задачі та обґрунтовано вибір відповідних методів машинного та глибинного навчання. Для проведення експериментів використано спеціально підготовлений набір українських новин, очищений, збалансований та попередньо оброблений згідно з вимогами NLP-моделювання. Описано всі етапи трансформації даних, а саме від токенізації та лематизації до векторизації та формування готових вибірок для навчання моделей.

У роботі було побудовано та проаналізовано кілька груп моделей: класичні алгоритми машинного навчання, нейронні мережі RNN та LSTM, а також сучасну трансформерну архітектуру BERT. Для кожної моделі було проведено повний цикл навчання, налаштування гіпер параметрів та оцінювання якості за відповідними метриками. Отримані експериментальні результати продемонстрували значні переваги трансформерного підходу, який краще справляється з урахуванням контексту та семантичних залежностей у текстах новин, оскільки його точність сягнула 89% до перенавчання та 90% після перенавчання на користувацьких даних. А от найгірші результати були отримані з

використанням методу Дерева рішення, де точність на тестових даних рівна 81%, що приблизно на 8% менше ніж трансформерна модель, також, непогано себе проявили рекурентні нейронні мережі, де точність коливалась від 86% до 88% при різних експериментах, що є дуже непоганими показниками враховуючи те, що вони є простіші і набагато швидше навчаються ніж трасформерні моделі, такі як Bert.

Окремим етапом роботи стала розробка повноцінної інфраструктури для практичного застосування моделей. Створено серверну частину на NestJS, реляційну базу даних PostgreSQL, ORM-рівень на TypeORM, реалізовано авторизацію за допомогою JWT та bcrypt, а також розроблено API-шару з автоматичною документацією через Swagger. Клієнтська частина створена на основі Next.js з використанням Redux Toolkit для управління станом та Tailwind CSS для побудови інтерфейсу. Для взаємодії з Python-сервісом, у якому розміщені моделі, застосовано Flask та бібліотеки такі, як TensorFlow, scikit-learn, Keras, Transformers та інші для обробки даних та навчання моделей. Система включає також адмін-панель, яка надає можливість переглядати новини, що вводять користувачі, аналізувати їх класифікацію та виконувати перенавчання моделей на нових даних, що забезпечує динамічне та адаптивне вдосконалення системи.Проведені експерименти довели ефективність запропонованих моделей і всієї побудованої інфраструктури.

Результати класифікації за метриками точності, повноти та F1-міри показали, що глибинні моделі, особливо BERT, суттєво перевершують класичні підходи й забезпечують високу якість виявлення фейкових новин, оскільки всі показники є вищими на 2-8% від інших використаних моделей. Інтеграція моделей у веб-систему продемонструвала практичну придатність розробленого рішення для реального використання.

Підсумовуючи виконану роботу, можна стверджувати, що всі поставлені завдання магістерської роботи були реалізовані. Створено

повноцінну систему, здатну автоматично класифікувати новини на правдиві та фейкові, а також підтримувати процес перенавчання моделей на основі нових даних. Отримані результати підтверджують ефективність застосування сучасних NLP-підходів і демонструють потенціал такого програмного забезпечення у протидії дезінформації та підвищенні інформаційної безпеки. Водночас системи такого типу можуть бути значно удосконалені шляхом використання більших наборів даних, донавчанням моделей на доменно-орієнтованих вибірках або впровадження мультимодальних підходів, таких як аналіз зображення та тексту або інших модальностей, що відкриває широкі перспективи для подальших досліджень та розвитку.

В результаті, можна сказати, що проведене дослідження робить вагомий внесок у розвиток інструментів автоматизованої боротьби з дезінформацією. Створена платформа демонструє, що поєднання сучасних методів машинного навчання та потужних веб-технологій відкриває нові можливості для побудови інтелектуальних систем аналітики контенту. Запропоновані рішення можуть бути масштабовані та адаптовані для інших сфер, де критично важливим є аналіз достовірності даних. У перспективі розвиток мультимодальних моделей, а також удосконалення інструментів обробки тексту та зображень дозволять досягти ще вищих результатів. Результати цієї роботи можуть стати основою для подальших наукових досліджень, а також для створення реальних продуктів, здатних протидіяти фейкам та підвищувати рівень інформаційної гігієни суспільства. Таким чином, виконана магістерська робота не лише досягає поставленої мети, але й відкриває широкі можливості для подальших інновацій у цій важливій сфері.

Списки Використаних Джерел

[1] Backlinko. Social Media Users: The Ultimate List of Social Media Statistics in 2025. URI: https://backlinko.com/social-media-users

[2] Backlinko. Social Media Users: The Ultimate List of Social Media Statistics in 2025. URI: https://backlinko.com/social-media-users

[3] Readline. Key Statistics on Fake News & Misinformation in Media in 2024: URI: https://redline.digital/fake-news-statistics/

[4] AI.Index. Stanford HAI / AI Index Report 2025. URI: aiindex.stanford.edu+2Caracal News+2

[5] P. Dhiman, A.Kaur, C. Iwendi, S.Mohan A Scientometric Analysis of Deep Learning Approaches for Detecting Fake News, 2023 DOI: https://doi.org/10.3390/electronics12040948

[6] Cornell University. Dataset of Fake News Detection and Fact Verification: A Survey, 2021. URI: https://arxiv.org/abs/2111.03299

[7] Disa. Disinformation in Ukraine: Prevalence of Falsehoods and Their Dissemination Across Social Media Platforms, 2025, URI: https://disa.org/disinformation-in-ukraine-prevalence-of-falsehoods-and-their-dissemination-across-social-media-platforms/

[8] N. Prachi, M.Habibullah, E.Rafi, R.Khan Detection of Fake News Using Machine Learning and Natural Language Processing Algorithms, 2022, DOI: 10.12720/jait.13.6.652-661

[9] B. Probierz, P. Stefański та J. Kozak Rapid Detection of Fake News Based on Machine Learning Methods, 2021, DOI: https://doi.org/10.1016/j.procs.2021.09.060

[10] K. Roumeiotis, N. Tselikas, D. Nasiopoulos Fake News Detection and Classification: A Comparative Study of Convolutional Neural Networks, Large Language Models, and Natural Language Processing Models, 2025, DOI: https://doi.org/10.3390/fi17010028

[11] M. F. Mridha, A. J. Keya, M. A. Hamid, M. M. Monowar, M. S. Rahman A Comprehensive Review on Fake News Detection With Deep Learning, 2021, DOI: 10.1109/ACCESS.2021.3129329

[12] N. Aslam, I. U. Khan, F. S. Alotaibi, L. A. Aldaej, A. K. Aldubaikil Fake Detect: A Deep Learning Ensemble Model for Fake News Detection, 2021, DOI: https://doi.org/10.1155/2021/5557784

[13] T. Chauhan, H. Palivela Optimization and improvement of fake news detection using deep learning approaches for societal benefit, 2021, DOI: https://doi.org/10.1016/j.jjimei.2021.100051

[14] D. Mouratidis, A. Kanavos, K. Kermanidis From Misinformation to Insight: Machine Learning Strategies for Fake News Detection, 2025, DOI: https://doi.org/10.3390/info16030189

[15] A. Altheneyan, A. Alhadlaq Big Data ML-Based Fake News Detection using Distributed Learning, 2023, DOI: 10.1109/ACCESS.2023.3260763

[16] Chauhan, G. "All about Logistic regression in one article." Towards Data Science, 2018. URL: https://towardsdatascience.com/logistic-regression-b0af09cdb8ad.

[17] I.D.Mienye, N.R.Jere A Survey of Decision Trees: Concepts, Algorithms, and Applications, 2024, DOI: 10.1109/ACCESS.2024.3416838

[18] H. Sahour, V. Gholami, J. Torkman, S. Saeedi Random forest and extreme gradient boosting algorithms for streamflow modeling using vessel features and tree-rings, 2021, DOI: 10.1007/s12665-021-10054-5

[19] С. Матвієнко Принципи побудови нейронних мереж, URL: https://itmaster.biz.ua/programming/vision/neural-networks-principles.html

[20] I.D.Mienye, T.G.Swart, G.Obaido Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications, 2024, DOI: https://doi.org/10.3390/info15090517

[21] A.Ismail, T.Wood, H.C.Bravo Improving Long-Horizon Forecasts with Expectation-Biased LSTM Networks, 2018, DOI: 10.48550/arXiv.1804.06776

[22] F.E. Ayo, O. Folorunso, F.T.Ibharalu, I.A. Osinuga Machine learning techniques for hate speech classification of Twitter data: State-of-the-art, future challenges and research directions. 2020. DOI: https://doi.org/10.1016/j.cosrev.2020.100311

[23] V.Petyk “Ukrainian news”, 2022, URL: https://www.kaggle.com/datasets/zepopo/ukrainian-fake-and-true-news

[24] N.I. Boyko, V.B.Kurylo "Medical Data Classification Algorithm for Oncology Prediction." 2023. DOI: https://doi.org/10.32782/2521-6643-2023.2-66.3.

[25] G. Suprianto, M.S. Prasetya, R.N.J. Meimo Nakashita Classification of Real and Fake News Using the Artificial Neural Network Method. DOI: https://doi.org/10.33795/jartel.v15i2.7422.

[26] G.R. Naidu, I. Rai, Y.V. Chowdhary Fake Reviews Detection Using an LSTM Model and NLP Techniques. 2025. DOI: 10.1109/NMITCON65824.2025.11187966.

[27] T. Mahmud, T. Akter, M.T. Aziz, M.K. Uddin, M.S. Hossain, K. Andersson Integration of NLP and Deep Learning for Automated Fake News Detection. 2024. DOI: https://ieeexplore.ieee.org/document/10675624

[28] N. Akter, M.Z. Abedin, M.T.R. Tarafder, N.A. Nikita, S. Nusrat Jahan, N. Nahar Rimi, M. Tariqul Islam Advanced Detection and Forecasting of Fake News on Social Media Platforms Using Natural Language Processing and Artificial Intelligence. 2024. DOI: https://doi.org/10.63332/joph.v5i6.2446.

[29] O. Bashaddadh, N. Omar, M. Mohd, M.N.A. Khalid Machine Learning and Deep Learning Approaches for Fake News Detection: A Systematic Review of Techniques, Challenges, and Advancements. 2025. DOI: https://ieeexplore.ieee.org/document/11008608.

[30] S. Tufchi, A. Yadav, T. Ahmed A comprehensive survey of multimodal fake news detection techniques: advances, challenges, and opportunities. 2023. DOI: https://doi.org/10.1007/s13735-023-00296-3.

[31] I.H. Ali Deep Learning and Fusion Mechanism-based Multimodal Fake News Detection Methodologies: A Review. 2024. DOI: https://doi.org/10.48084/etasr.7907.

ДОДАТКИ

Додаток А

Таблиця 1.1. - Огляд обраних літературних джерел

№	Назва (укр.)	Задача	Методи	Рік	Посилання
1	Науковометричний аналіз підходів глибокого навчання для виявлення фейкових новин	Систематизація публікацій, трендів і тематичних напрямів у сфері виявлення фейкових новин	Bibliometrix (R-package), VOSviewer, аналіз бази Scopus	2023	[5]
2	Виявлення фейкових новин за допомогою машинного навчання та алгоритмів обробки природної мови	Створення системи класифікації фейкових новин із використанням ML та DL	Logistic Regression, Decision Tree, Naive Bayes, SVM, LSTM, BERT	2022	[8]
3	Швидке виявлення фейкових новин на основі методів машинного навчання	Розробка швидкої моделі класифікації новин за заголовком і повним текстом	CART, SVM, Random Forest, AdaBoost, Bagging, NLP-техніки	2021	[9]
4	Виявлення та класифікація фейкових новин: порівняльне дослідження згорткових нейронних мереж, моделей великих мов та моделей обробки природної мови	Визначення ефективності великих мовних моделей у виявленні фейкових новин	CNN, GPT-4o, GPT-4o-mini, fine-tuning, LLM	2025	[10]

5	Комплексний огляд виявлення фейкових новин за допомогою глибокого навчання	Узагальнення підходів, аналіз архітектур DL і NLP, формування таксономії	LSTM, GRU, CNN, BERT, RoBERTa, XLNet, порівняльний аналіз	2021	[11]
6	Виявлення фейків: модель ансамблю глибокого навчання для виявлення фейкових новин	Підвищення точності класифікації за допомогою гібридної глибинної архітектури	Bi-LSTM-GRU-Dense, Deep Learning Dense Model, NLP-передобробка	2021	[12]
7	Оптимізація та покращення виявлення фейкових новин за допомогою методів глибокого навчання для суспільної користі	Підвищення точності класифікації фейкових новин і соціальної безпеки	LSTM, GloVe, токенізація, очищення тексту	2021	[13]
8	Від дезінформації до розуміння: стратегії машинного навчання для виявлення фейкових новин	Інтеграція класичних та DL-моделей для підвищення точності класифікації	Naïve Bayes, SVM, Random Forest, CNN, LSTM, BERT, Word2Vec	2025	[14]
9	Виявлення фейкових новин на основі машинного навчання великих даних з використанням розподіленого навчання	Розробка моделі на базі Spark для великомасштабної обробки новин	Spark, Headline Stance Checker, N-grams, HashingTF-IDF, Stacked Ensemble	2023	[15]

Додаток Б

Посилання на gitHub репозиторій з моделями та апі для них: https://github.com/ValentynKurylo/DiplomaWorkFakeNewsModels.git

Посилання на репозиторій з веб-застосунком: https://github.com/ValentynKurylo/DiplomaWorkFakeNewsWeb.git