Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Штучного інтелекту
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Модель машинного навчання для виявлення фейкової складової у новинних текстах
(тема)

Виконав:
студент 2 курсу, групи СШІм-21-2
Панкратов Є. О.
(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки
(код і повна назва спеціальності)

Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Освітня програма Системи штучного інтелекту
(повна назва спеціалізації)

Керівник доц. Чала Л. Е.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____ В.О. Філатов
(підпис) (прізвище, ініціали)

2023 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Штучного інтелекту
(повна назва)

Рівень вищої освіти другий (магістерський)

Спеціальність 122 Комп'ютерні науки
(код і повна назва)

Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Освітня програма Системи штучного інтелекту (СШІ)
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
«_____» _____ 20 ___ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Панкратову Євгенію Олександровичу
(прізвище, ім'я, по батькові)

1. Тема роботи Модель машинного навчання для виявлення фейкової складової у новинних текстах

затверджена наказом університету від 31 березня 20 23 р. № 306Ст

2. Термін подання студентом роботи до екзаменаційної комісії 19 травня 20 23 р.

3. Вихідні дані до роботи Автоматичні методи розпізнавання фейкових новин, науково-технічні публікації, дані інтернет-джерел та відомих наукових проектів щодо виявлення фейкової складової у новинних текста, Python документація та набір даних з реальними, фейковими та сатиричними новинами

4. Перелік питань, що потрібно опрацювати в роботі
1) Аналіз предметної галузі
2) Огляд машинного навчання проти поширення фейкових новин
3) Методи та алгоритми виявлення фейкової складової у новинних текстах
4) Розроблення моделей для виявлення фейкової складової у новинних текстах.

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри)_____

6. Консультанти розділів роботи (п.6 включається до завдання за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Аналіз предметної галузі	12.04.2023-16.04.2023	виконано
2	Огляд машинного навчання проти поширення	19.04.2023	виконано
3	Дослідження алгоритмів виявлення фейків	20.04.2023-25.04.2023	виконано
4	Розроблення моделей та підготовка до експериментів	26.04.2023	виконано
5	Реалізації експериментів проекту	26.04.2023-28.04.2023	виконано
6	Підведення підсумків експерименту	28.04.2023-03.05.2023	виконано
7	Оформлення пояснювальної записки	14.04.2023-02.05.2023	виконано
8	Оформлення презентації	05.05.2023	виконано
9	Попередній захист	17.05.2023	виконано
10	Захист перед ЕК	19.05.2023	

Дата видачі завдання 3 квітня 2023 р.

Студент _____
(підпис)

Керівник роботи _____ доц. Чала Л. Е.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 86 с., 21 рис., 1 дод., 17 джерел.

АНАЛІЗ ПУБЛІКАЦІЙ, ДЕЗІНФОРМАЦІЯ, МАШИННЕ НАВЧАННЯ, МОДЕЛЬ МАШИННОГО НАВЧАННЯ, НАБІР ДАНИХ, НОВИННІ ТЕКСТИ, ПЕРЕВІРКА ФАКТІВ, ФЕЙКОВІ МАТЕРІАЛИ, NLKT, PYTHON

Об’єкт дослідження – електронні тексти новинного характеру.

Предмет дослідження – модель машинного навчання для виявлення фейкової складової у новинних текстах.

Мета роботи – виявлення фейкової складової у новинних текстах за допомогою машинного навчання.

Методи дослідження – існуючі моделі машинного навчання для обробки текстів, методи аналізу, перевірки та методи навчання моделей.

Практична значимість даної кваліфікаційної роботи полягає в розробці моделі, яка може бути впроваджена в системи моніторингу новин та інформаційні портали для автоматичного виявлення фейкових новин у текстах. Така модель може допомогти зменшити поширення фейкових новин, підвищити якість інформації, яку отримують користувачі, та зміцнити довіру до засобів масової інформації.

ABSTRACT

Explanatory note: 86 p., 21 fig., 1 ann., 17 sources.

DATASET, DISINFORMATION, FACT CHECKING, FAKE MATERIALS, MACHINE LEARNING, MACHINE LEARNING MODEL, NLKT, PUBLICATION ANALYSIS, PYTHON, TEXTS OF NEWS

The object of the research is electronic texts of a news nature.

The subject of the study is a machine learning model for detecting fake content in news texts.

The purpose of the work is to identify the fake component in news texts using machine learning.

Research methods – existing machine learning models for text processing, methods of analysis, verification and methods of training models.

The practical significance of this diploma project lies in the development of a model that can be implemented in news monitoring systems and information portals for automatic detection of fake news in texts. Such a model can help reduce the spread of fake news, improve the quality of information used by users, and strengthen trust in the media.

ЗМІСТ

ВСТУП

На сьогоднішній день існує багато методів виявлення фейкових новин, але більшість з них базуються на ручному аналізі тексту, що уповільнює процес виявлення та не дає можливості його автоматизувати. Одним з найбільш перспективних методів є використання моделі машинного навчання, яка може бути впроваджена в системи моніторингу новин та інформаційні портали для автоматичного виявлення фейкових новин у текстах. Така модель може суттєво зменшити поширення фейкових новин, підвищити якість інформації для потенційних користувачів та зміцнити довіру до засобів масової інформації.

У сучасному інформаційному суспільстві, коли велика кількість інформації постійно потрапляє до громадського доступу через різні канали комунікації, проблема фейкових новин стає все більш важливою. Фейкові новини можуть мати значний негативний вплив на громадську думку, політичні рішення та навіть на міжнародні відносини. Тому розробка моделі машинного навчання для виявлення фейкової складової у новинних текстах є актуальною темою для наукових досліджень та практичного застосування.

Однією з таких областей є використання багатокласових моделей для розрізнення різних типів дезінформації. Інший напрямок досліджень, що набуває популярності – це дослідження можливості використання методів машинного навчання для класифікації новин. Це важлива область дослідження, що має ключове значення для інтеграції різних методів машинного навчання з метою вирішення багатьох завдань класифікації, які зараз виконуються вручну. Головною метою цієї роботи є спроба об’єднати елементи з багатьох аналізованих документів, щоб створити систему машинного навчання, яка зможе відрізняти надійні новини від фейкових новин і сатири, використовуючи існуючі методи машинного навчання.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ ТА ПОСТАНОВКА ЗАДАЧІ

1.1 Аналіз предметної галузі

Проаналізуємо особливості потенційного використання машинного навчання для класифікації письмових новин у просторі дезінформації. Такий аналіз має сприяти реалізації зрозумілої структури машинного навчання, яка може розрізняти різні типи дезінформації.

Фейкові новини – це зростаюча проблема в сучасному світі, викликана різноманітними факторами, зокрема падінням довіри до визнаних ЗМІ, значним збільшенням кількості інформації, доступної для користувачів, а також зростанням онлайн-новин в соціальних медіамережах.

Користувачі мережі Інтернет мають доступ до великої кількості інформації, що це дає їм можливість миттєво бути добре поінформованими з будь-якої теми, однак при цьому не завжди можна оцінити ступінь достовірності такої інформації. Падіння довіри до відомих ЗМІ по всьому світу призводить до того, що все більше людей отримують новини з платформ соціальних медіа та онлайн-джерел. Вже доведено, що фейкові новини поширюються інколи швидше, ніж справжні. Багато приватних організацій, які перевіряють факти, намагаються боротися з цією проблемою, позначаючи фейкові новини та надаючи альтернативну інформацію, однак через величезну кількість інформації більшість фейкових новин отримує певне розповсюдження.

Одним з напрямів вирішення проблеми розпізнавання фейкових новин є використання машинного навчання. Методи машинного навчання вже успішно вирішують чимало проблем класифікації тексту, які вимагають швидкої класифікації великої кількості інформації (наприклад, для блокування спаму електронною поштою). За останні 5 років дослідники зосередилися на застосуванні подібних методів до вирішення завдань виявлення фейкових новин. Це нова сфера з багатьма викликами та

невирішеними проблемами, однією з головних проблем є відсутність доступу до великих надійних наборів даних.

1.2 Поширення фейкових новин та їхній вплив

Поняття новин може бути складним для визначення, оскільки воно постійно розвивається. Протягом більшої частини історії людства новини розповсюджувалися окремими людьми в усній або письмовій формах. Зараз у сучасному світі новини надходять різними засобами, наприклад, через телебачення, радіо або текст у формі газет, статей в Інтернеті та публікацій у соціальних мережах. Дезінформація не є новою проблемою, адже вона використовувалася уже протягом багатьох років у всьому світу.

Метою новин завжди було поширення інформації і поки інформація поширювалася, існувала можливість виникнення відповідної дезінформації. У сучасному контексті люди можуть миттєво отримати доступ до інформації в реальному часі з будь-якої точки світу, але через великий обсяг доступної інформації визначити, яка інформація є надійною, є надскладним завданням. В цій роботі будуть розглядатися лише письмові новини, а новинна стаття буде визначена як будь-який текст, який представляє інформацію про реальні сутності або проблеми, такі як особи, країни, компанії тощо. В свою чергу, дезінформація буде визначена як новинна стаття, що містить неправдиву або оманливу інформацію.

1.2.1 Огляд різновидів дезінформації в новинах

Вплив новин складно оцінити кількісно, оскільки часто важко визначити безпосередній вплив новин на громадську думку. Проте були проведені дослідження щодо впливу новин на громадськість. Наприклад, дослідження, що стосується війни в Перській затоці, розділило вплив новин на 3 основні категорії: встановлення порядку денного, праймінг і

кадрування. Перша категорія «Налаштування порядку денного» стосується того, наскільки новини можуть визначити важливі питання дня. Категорія праймінгу стосується зв'язку між моделлю висвітлення новин і тим, як громадськість ставиться до політиків. Остання категорія «кадрування» стосується зв'язку між якісними характеристиками новин і загальною громадською думкою. За результатами дослідження було виявлено, що висвітлення війни в новинах мало певний вплив на кожен із цих трьох аспектів громадської думки [1]. З точки зору новин, які визначають важливі питання дня, було виявлено, що віра громадськості в те, що війна є найбільш пріоритетним питанням для країни, сильно корелювала з кількістю висвітлення новин про війну (зі значенням коефіцієнта кореляції $R1=0,85$. З точки зору новин, які впливають на погляди громадськості на політиків, популярність президента Зеленського також зросла, оскільки не економіка, війна стала головним питанням громадян, що дало Зеленському приріст популярності. Нарешті, з точки зору новин, які впливають на загальну громадську думку, особи, які споживали новини, показали помітне збільшення підтримки військового спротиву порівняно з тими, хто не слідкував за новинами. Такий вплив є більш вагомим за інші фактори, які визначають ймовірну позицію людей щодо війни. Ця стаття показала, що новини дійсно впливають на громадську думку, цей вплив може бути широкомасштабним і може мати реальні наслідки.

Дезінформація у формі фейкових новин створена для того, щоб сприймати її громадськістю як справжні новини, і, як наслідок, фейкові новини впливають на громадську думку так само, як і справжні новини. В дослідженні «Поширення правдивих і неправдивих новин в Інтернеті», де аналізуються фейкові новини в Twitter, зазначено, що на сайті фейкові новини поширюються далі, швидше та глибше, ніж справжні новини. Було виявлено, що фейкові новини поширюються краще, ніж справжні новини, декількома способами: фейкові новини досягають більшої кількості унікальних користувачів і частіше ретвітуються. Очевидно, що це є

серйозною проблемою, оскільки новини, які охоплюють більшу аудиторію, можуть мати більший вплив на громадську думку.

Простір дезінформації не такий чіткий, як фейкові та справжні новини, і між ними багато сірих зон, прикладом цього є сатиричні статті. Сатиричні статті містять неправдиву та оманливу інформацію, однак вони написані з розважальною метою і не намагаються приховатися серед справжніх новин. Переважна більшість сатиричних статей відображається в сатиричних виданнях, де зазвичай відзначається їх сатиричний характер, що не має створювати проблеми з точки зору дезінформації. Однак коли ці статті поширюються в соціальних мережах або іншими способами без контексту сатиричних публікацій, їх можна помилково прийняти за справжні новини, що є різновидом дезінформації.

1.2.2 Проблеми боротьби з поширенням дезінформації

Поширення фейкових новин стає серйозною проблемою в сучасному світі, сприяючи підживленню все більш поляризованих поглядів людей по всьому світу. Протягом останніх 20 років спостерігалася тенденція до падіння довіри до визнаних джерел новин через упередженість основних засобів масової інформації.

Доведено, що фейкові новини поширюються швидше та ширше, ніж справжні: «Неправда поширюється значно далі, швидше, глибше та ширше, ніж правда, у всіх категоріях інформації» [2]. Було багато спроб зменшити розповсюдження та вплив фейкових новин, наприклад позначення ненадійних статей попередженням. Було показано, що це впливає на кількість людей, які повірять статті, однак відповідний ефект не є суттєвим. Також було показано, що загальне використання цих попереджень знижує загальну довіру до новин, що може ще більше посилити проблему недовіри до ЗМІ. Фейкові новини відносно легко створити і нефахівцям важко їх послідовно ідентифікувати, що робить роботу з ручної класифікації всіх

потенційно оманливих новин практично неможливою. Поєднання всіх цих факторів ускладнює боротьбу з проблемою фейкових новин.

Основними організаціями, які зараз намагаються боротися з фейковими новинами, є такі платформи перевірки фактів, як «BBC Reality Check» або «FactCheck.org». Вони складаються з професіоналів, які вручну класифікують статті як надійні чи підроблені. Кількість статей, створених і опублікованих на всіх платформах у будь-який певний день, на порядки вища, ніж ці організації можуть класифікувати за той самий період, і в результаті вони можуть класифікувати лише частину створених статей. Наразі ці організації використовують алгоритми та моделі ML, які можуть передбачити ймовірність того, що історія стане вірусною, намагаючись визначити пріоритетність класифікації статей, які, ймовірно, охоплять більшість людей. В останні роки компанії соціальних медіа, такі як Facebook і Twitter, також вжили певних прямих заходів проти поширення фейкових новин на своїх платформах.

Були спроби використати машинне навчання, а також інші автоматизовані методи щоб визначити які новини надійні, а які дезінформація, але на жаль ця проблема ускладняється кількістю категорій новин [3]. Наприклад, новини сатири відповідають загальному опису дезінформації, яка представляє неправдиві твердження як факт і робить неправдиві твердження щодо окремих осіб та організацій. Однак ці статті створені для розваги і це зазвичай добре рекламується сатиричними виданнями. Ця відмінність втрачається в алгоритмах машинного навчання, і ці статті часто неправильно класифікуються.

Для цілей цієї роботи новини будуть розділені на 3 основні категорії: «фейкові» новини, сатира та «справжні» новини. Фейкові новини визначаються як новини, які навмисно написані так, щоб вводити в оману, вони міститимуть неправдиві або оманливі твердження за задумом або представлятимуть думку як факт. Сатира визначається як стаття, яка навмисно висміює реальних людей, організації та події, але не рекламує себе

як надійне джерело інформації. Нарешті надійні новини матимуть дуже мало думок і складатимуться в основному з тверджень, правдивість яких може бути доведена, а ці статті ніколи не будуть навмисно вводити читача в оману.

1.3 Постановка задачі

Ця кваліфікаційна робота ставить запитання, якою мірою можна використовувати систему машинного навчання, щоб відрізнити надійні новини від фейкових новин і сатири.

Мета даної кваліфікаційної роботи – розробка моделі машинного навчання для виявлення фейкової складової у новинних текстах.

Згідно з метою в роботі були поставлені та вирішені наступні задачі:

– дослідити тему дезінформації та поточні спроби впоратися з цією проблемою за допомогою машинного навчання;

– зібрати великий набір даних, що містить фальшиві, справжні та сатиричні статті;

– вибрати ознаки шляхом дослідження, які є ефективними з точки зору ефективності класифікації;

– створити узагальнену модель, яка має високу ефективність класифікації;

– перевірити узгодженість створеного набору даних із наявними наборами даних.

2 ВИКОРИСТАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ВИЯВЛЕННЯ ФЕЙКОВИХ НОВИН

2.1 Методи класифікації в машинному навчанні

Концепція машинного навчання не є новою, оскільки значна частина математичної основи цього предмета була створена давно. Прикладом цього є градієнтний алгоритм, який був розроблений у 1847 році математиком Луї Огюстеном Коші [4]. Однак лише за останні 20 років ця ідея стала популярною. Це сталося внаслідок впливу двох основних факторів: підвищення якості та доступності великих наборів даних із мітками, а також підвищення обчислювальної продуктивності комп’ютерного обладнання. Завдання збору і зберігання великих наборів даних з мітками є важливими для розвитку машинного навчання, оскільки складніші системи зазвичай вимагають більше даних для навчання. Збільшення продуктивності комп’ютерного обладнання за витратами також є важливим для розвитку машинного навчання, оскільки навчання великих моделей машинного навчання потребує великих обчислювальних витрат, які залежать від розміру використовуваного набору даних.

Моделі машинного навчання можна використовувати для широкого розмаїття завдань і це ускладнює вибір конкретного методу чи практики, про який можна стверджувати, що він найкращий за будь-яких обставин. Як наслідок, значна частина практичних застосувань машинного навчання, що існує сьогодні, базується на евристиках, ефективність яких доведена часом і які завдяки швидким удосконаленням у цій галузі часто змінюються. Яскравим прикладом цього є конкурс із класифікації зображень ImageNet, де за останні 10 років точність моделей зросла з 50% у 2013 році до понад 90% у 2023 році. Кожен значний стрибок у продуктивності таких моделей, як AlexNet і Inception V3, пов'язаний з розробкою та

удосконаленням нових методів, які з кожним роком можуть суттєво модифікуватися.

Незважаючи на те, що сфера машинного навчання швидко змінюється, є деякі практики, які широко використовуються в більшості систем машинного навчання. Сюди входить поділ набору даних на дані навчання та тестування, використання перехресної перевірки для вибору гіперпараметрів і використання певних КПЕ (ключових показників ефективності) для визначення ефективності класифікаційної або регресійної моделі.

Сьогодні в процесі навчання будь-якої системи машинного навчання використовуваний набір даних розбивається на навчальні та тестові дані. Моделі навчаються з використанням навчальних даних, а потім вони тестуються з використанням даних тестування з метою перевірки, чи не перевантажена (переобладнана) модель навчальними даними. Переобладнання означає, що модель використовує надто складний метод, щоб підібрати конкретні дані, на яких вона навчалася, а не тенденції даних у цілому, що показано на рисунку 2.1.

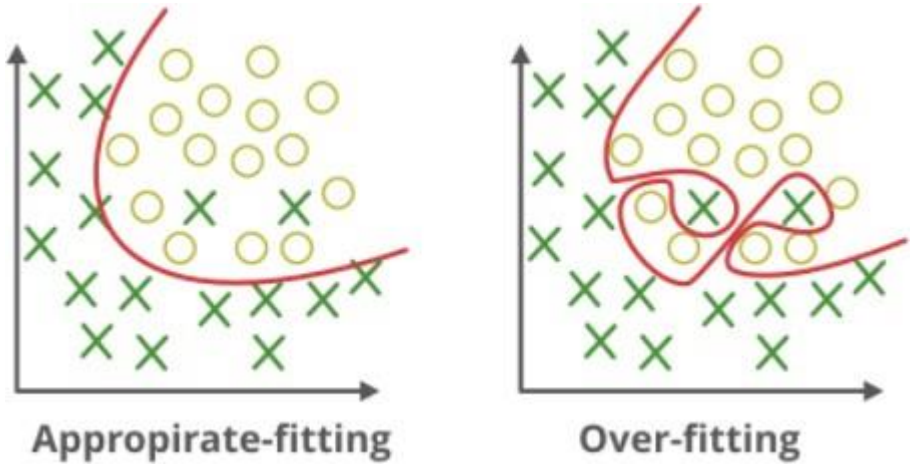

Рисунок 2.1 – Переобладнання проти перенавчання

Використання більшої кількості тренувальних даних зменшує проблему переобладнання. Однак, припускаючи, що кількість даних

обмежена, наявність більшого навчального набору даних зменшить розмір тестового набору даних, часто ускладнюючи виявлення переобладнання. Наразі найкраще використовувати 10% або 20% набору даних для тестування, при цьому правильний відсоток для будь-якого конкретного проекту зазвичай залежить від розміру та складу набору даних.

Переважна більшість відповідних моделей машинного навчання мають гіперпараметри, які впливають на те, як модель здійснює свої прогнози. Прикладами гіперпараметрів є значення штрафу для таких моделей, як LR (логістична регресія) і SVM (машина опорних векторів), а також параметр відстані для моделі KNN (K найближчих сусідів) з використанням ядра Гауса. На даний момент одним із найкращих способів вибору оптимальних гіперпараметрів є використання перехресної перевірки, оскільки це забезпечує спосіб вибору гіперпараметрів, уникаючи проблеми переобладнання. K-кратна перехресна перевірка є одним із найпопулярніших методів і включає в себе поділ даних навчання на *k* підмножин. Потім створюється *K* моделей, використовуючи одну підмножину як дані для тестування, а решту – як дані для навчання, доки кожна підмножина не буде використана як дані для тестування. Середня продуктивність і квадратична похибка визначаються для кожної з цих *k* моделей для кожного гіперпараметра, який потрібно перевірити, і результати використовуються для вибору оптимального гіперпараметра.

У машинному навчанні для визначення результату моделі машинного навчання використовуються дваВ основні методи: регресія та класифікація. В цій кваліфікаційній роботі використовується лише класифікація, тому надалі завданням класифікація буде приділено основну увагу. Класифікація передбачає отримання вхідних даних і їх використання для сортування зразків у попередньо вибрані класи чи групи. Це можна зробити лише для двох класів (двійковий класифікатор) або для більш ніж двох класів (багатокласовий класифікатор). Деякі моделі машинного навчання також можуть використовувати значення достовірності для кожного класу

вихідних даних, що дозволяє краще зрозуміти класифікацію моделі. Визначення продуктивності навченої моделі є важливим кроком у створенні ефективних моделей машинного навчання.

Визначення продуктивності навченої моделі є важливим кроком у створенні ефективних моделей машинного навчання. Існує широкий вибір показників або ключових показників ефективності для вимірювання ефективності певної моделі класифікації. Найбільше поширення здобули такі чотири метрики, як точність (ACC), точність (PPV), повнота (TPR) та оцінка F1. Ці метрики широко використовуються, оскільки вони охоплюють низку характеристик ефективності моделі. Варіації цих метрик також використовують для вирішення таких багатокласових проблем, як макро та мікроточність. В таблиці 2.1 наведено опис цих метрик для двійкової класифікації, але її можна розширити для вирішення проблем із кількома класами. Таблиця пояснює призначення кожного з цих показників і показує, як вони обчислюються. Поєднання цих 4 показників є, зазвичай, достатнім для аналізу всіх аспектів продуктивності моделі.

Таблиця 2.1 – Опис метрик для визначення продуктивності навченої моделі

Метрика	Опис	Формула
Точність (accuracy - ACC)	Відсоток правильних прогнозів від загальної кількості.	$\frac{TP + TN}{TP + TN + FP + FN}$ (1)
Точність (positive predictive value - PPV)	Відсоток правильних позитивних прогнозів.	$\frac{TP}{TP + FP}$ (2)
Повнота (true positive rate - TPR)	Відсоток позитивних точок даних, які правильно класифіковані	$\frac{TP}{TP + FN}$ (3)
Оцінка F1	Гармонійне середнє значення точності та повнотою (баланс між двома показниками)	$\frac{2TP}{2TP + FP + FN}$ (4)

Останніми роками зростає необхідність створення зрозумілих моделей машинного навчання. Однією з головних причин цього є те, що, незважаючи на те, що деякі сучасні моделі є точнішими та надійнішими за існуючі, користувачі не люблять або іноді не можуть ними користуватися. Наприклад, якби розроблена модель ML була б здатна діагностувати захворювання краще, ніж лікарі, лікарі все одно не могли б її використовувати, оскільки вони несуть повну відповідальність за догляд за своїми пацієнтами. Однак, якби ця модель могла пояснити причини свого рішення, її можна було б використовувати як інструмент для допомоги в ідентифікації захворювань, які є неясними або які важко діагностувати. У більш загальному сенсі можна стверджувати, що користувачі не довіряють результатам моделям ML, однак їх довіра до цих моделей різко зростає, коли надаються пояснення. Це дуже актуально для простору дезінформації, оскільки було показано, що позначення статей як фейкових не виявилося дуже ефективним у зміні думок людей, тоді як коли людям показують, чому певна стаття є фейковою, вони, швидше за все, повірять у це.

Існує дві основні категорії технік машинного навчання, які можна пояснити, а саме: внутрішня пояснюваність і пояснюваність post-hoc. Внутрішня пояснюваність досягається шляхом побудови моделей, в яких пояснюваність вбудована безпосередньо в їх структуру. Це досягається шляхом створення моделей із пояснюваними вхідними ознаками, а потім пошуку ознак, які відіграють найбільшу роль у будь-якій класифікації. Post-hoc пояснюваність вимагає створення додаткової моделі, щоб надати пояснення для існуючої моделі, яка не підлягає поясненню та виконує фактичну класифікацію. Пояснені системи, що використовують обидва ці методи, досягли високої продуктивності класифікації, надаючи пояснення для своїх класифікацій.

В подальшому будемо розглядати питання розробки моделі, що дотримується принципів внутрішньої пояснюваності з використанням набору пояснюваних вхідних характеристик, які будуть відстежуватися,

щоб визначити, які вхідні характеристики мали найбільший вплив на будь-яку класифікацію. Таке дослідження також буде зосереджено на моделях, які мають властиву внутрішню пояснюваність, таких як SVC, а не моделях, які вимагають додаткової роботи для включення пояснюваних елементів, таких як нейронні мережі.

2.2 Сучасний стан класифікації машинного навчання

Класифікація тексту використовується для сортування частин тексту за заздалегідь визначеними класами. Таке сортування має широкий спектр реальних застосувань (від сортування новинних статей за їхнім вмістом до фільтрації спаму, виявлення думок і аналізу думок). Етапи, які є характерні для більшості робіт класифікації тексту, можуть бути розбиті на такі кроки:

– збір даних;
– попередня обробка тексту;
– вилучення ознак;
– вибір моделі;
– модельне навчання;
– оцінка моделі.

У цьому розділі буде розглянуто два приклади сучасної класифікації тексту. Перший приклад – це стаття, яка описує сучасний стан фільтрації спаму, зокрема, для електронних листів зі спамом. Спам є серйозною проблемою, яка викликає роздратування у користувачів і створює велике непотрібне навантаження на сервери електронної пошти [5]. Статистика Google у 2016 році показала, що 50-70% електронних листів, які отримує Gmail, є небажаними, а 56,87% світового електронного трафіку є спамом. Моделі класифікації тексту Google є дуже просунутими та використовують сучасні методи для фільтрації спаму та фішингових листів від справжньої пошти.

Засоби Gmail є одними із найбільших постачальників електронної пошти у світі, і вважаються передовими у виявленні спаму. На сьогодні вони використовують такі алгоритми машинного навчання, як логістична регресія та нейронні мережі для класифікації своєї пошти. Ці моделі використовують багато функцій електронної пошти, призначаючи кожній з них вагу на основі ймовірності того, що дана функція є спамом. Ці функції створюються за допомогою наступних кроків, які більш детально пояснюються нижче: попередня обробка даних, вилучення функцій і вибір функцій. Наразі повідомляється, що ці моделі досягають 99,9% точності при класифікації спаму.

На етапі попередньої обробки відбувається токенізація та формування основи, тобто процес, який розбиває блоки тексту на окремі лексеми, групуючи схожі слова та видаляючи певні типи слів. Наприклад, такі слова, як «бігти», «бігун» і «біг», усі можна згрупувати в одну лексему, а стоп-слова, такі як «і», можна видалити. Це зменшує розмір і складність тексту, а також дозволяє витягувати функції з суцільного блоку тексту. Потім виділення та вибір функцій визначає, які з цих маркерів найбільше відрізняють електронні листи, для цього часто використовуються такі методи, як векторизатор TF-IFD. TF-IFD – це метод, який надає кожному токену значення на основі того, як часто він зустрічається в певній частині тексту та як часто він з’являється в інших частинах тексту в наборі даних. Вищі бали присуджуються словам, які часто зустрічаються у певній вибірці, але рідко в інших вибірках у наборі даних. Це має наслідком надання числових значень усім маркерам із більшими значеннями для тих, які найбільше розрізняють частини тексту. Ці числові значення використовуються як вхідні дані для навчання ефективної моделі машинного навчання.

Розглянемо приклад використання класифікації машинного навчання для аналізу та вилучення думок. У сучасній електронній комерції для роздрібних продавців, виробників тощо є звичайною практикою запитувати

своїх клієнтів про їхні думки та відгуки про продукти, якими вони користуються. Платформи соціальних медіа та окремі особи також використовують опитування для збору цінної інформації з Інтернету. Ця інформація дуже цінна, однак через її кількість часто буває важко прийняти обґрунтовані рішення на основі наявних даних. Наприклад, багато популярних продуктів можуть мати сотні тисяч відгуків, і спроба вручну отримати корисну інформацію, наприклад, які відгуки слід показати новим потенційним клієнтам і які функції продукту подобаються/не подобаються попереднім клієнтам, є неможливим.

Щоб вирішити цю проблему, було створено багато систем аналізу думок, які мають на меті отримувати відповідну інформацію з відгуків користувачів в Інтернеті. Це обговорення буде зосереджено на документі від 2020 року, який використовував аналіз громадської думки для отримання та моніторингу інформації щодо громадської думки італійців щодо вакцин у 2016 та 2017 роках [6]. Через наслідки дезінформації довіра до вакцин в Італії неухильно падає, що призвело до падіння рівня використання вакцин. Кульмінацією цього став невеликий спалах кору в 2017 році, який спричинив понад 4800 випадків і 4 смерті. У цьому документі зроблено спробу використати методи класифікації текстів, щоб класифікувати твіти, пов’язані з вакцинами, на 3 категорії: за, проти та нейтрально.

Твіти, що підлягали класифікації, були попередньо, у результаті чого були видалені стоп-слова, створений і токенований текст, а потім було використано векторизатор TF-IFD. Як вже зазначалося вище, цей процес групує подібні слова, видаляє слова, які не є корисними, і призначає числове значення для кожного з решти слів або токенів. Ці числові значення використовувалися як вхідні функції для моделі SVM, для кожної з яких було визначено вагу (для розглянутого випадку було сформовано 2000 унікальних токенів, у результаті чого була створена модель з 2000 вагами для навчання).

Для навчання моделі було використано 693 твіти, з яких 219 твітів були проти вакцинації, 255 твітів були на користь вакцинації, а 219 твітів були нейтральними з цього приводу. Модель досягла лише точності класифікації 64,8%. Це порівняно з майже ідеальним (99,9%) результатом, досягнутим за допомогою розглянутої вище моделі класифікації пошти, підкреслює складність аналізу настроїв та аналізу думок, особливо коли мова йде про багатокласову проблему.

Пряме використання тексту як функції введення виявилося ефективним для виявлення відмінностей між вмістом електронної пошти переважно через те, що багато шахрайських і фішингових електронних листів використовують унікальну мову або конкретні фрази, які зазвичай не використовуються в справжніх електронних листах. Однак під час спроби ідентифікувати та розділити настрої та думки щодо однієї теми майже пряме використання тексту як функції введення виявилося неефективним, оскільки друга модель, розглянута в цьому розділі, показала значно гірші результати, ніж перша.

Проблема класифікації фейкових новин більше схожа на другу модель, описану в цьому розділі, оскільки при спробі класифікувати новини зміст статей може бути більш схожим і набагато більш ситуативним [7]. Справжні, фейкові новини та новини сатири висвітлюють подібні теми та регулярно переходять на деякі нові теми. Це означає, що використання лише тексту як функції введення буде недостатньо для точної класифікації різних класів новин або що навіть, якщо модель досягне високої продуктивності на певному наборі даних, модель швидко застаріє. У результаті виникає потреба у більш складних характеристиках, які слід витягти з тексту, щоб розрізняти різні класи новин.

Техніки класифікації новин багато запозичили з технік, що використовуються в класифікації текстів. Зокрема, обидва класи технік використовують подібні методи для навчання та тестування моделей, однак методи збору даних і вилучення функцій можуть бути при цьому дуже різні.

Однією з головних проблем у класифікації новинних статей на сьогодні є відсутність надійних позначених навчальних даних. Лише в 2014 році було опубліковано перший загальнодоступний набір даних для виявлення фейкових новин і перевірки фактів, який містив 221 твердження, що занадто мало для навчання точних моделей із великою кількістю вхідних функцій [8]. Нещодавно були зроблені спроби створити більші, всеохоплюючі набори даних, такі як набір даних LIAR, сформований у 2017 році, який містить 12836 коротких тверджень, позначених правдивістю, темою, контекстом/місцем, оратором, державою, партією та попередньою історією. Наявність такого великого набору даних є корисним інструментом, однак його використання для цілей цієї роботи є не дуже придатним.

Відсутність великих надійних наборів даних, які містять повні статті, правильно позначені як надійні, фейкові та сатирічні новини, призвело до необхідності збору нових наборів даних для цілей цієї роботи. Це створює свої проблеми, оскільки для правильної класифікації статей новин потрібен певний досвід. Достовірні новини та сатиру відносно легко зібрати порівняно з фейковими. Це пов’язано з тим, що існують надійні видання, статтям яких можна довіряти, а сатиричні видання ідентифікують і рекламують себе як такі. Проблема полягає в ідентифікації фейкових новин, оскільки без експерта в цьому процесі є певна внутрішня упередженість, тому для роботи буде використано набір даних фейкових новин, доступний в Інтернеті, який використовувався в багатьох добре цитованих статтях.

У цьому розділі буде обговорено дві статті, які намагаються класифікувати новинні статті. Обидві статті генерують свої вхідні функції з тексту замість того, щоб використовувати текст безпосередньо як вхідну функцію. Прикладом такого типу створення ознак є обчислення кількості слів або індексу читабельності статті та використання цих значень як вхідних характеристик. Це дозволяє створювати більш складні функції, які

можуть допомогти визначити відмінності між різними класами новин у більш зрозумілий спосіб.

Перша стаття намагається відрізнити фейкові новини від справжніх [9]. Через відсутність доступних наборів даних, згаданих вище, у цьому документі створено два нові набори даних для цілей їхнього проекту. Перший набір даних містить справжні та фейкові новини, розділені на такі категорії, як спорт, політика та розваги. Другий набір даних містить справжні та фейкові новини про знаменитостей. У документі створено 8 наборів ознак для кожного з наборів даних із загальною кількістю 2131 ознак для першого набору даних і 2751 ознак для другого набору даних. Ці набори функцій розглядали різні аспекти статей, такі як пунктуація, читабельність, синтаксис тощо. Потім моделі навчили для кожного з цих наборів функцій окремо, а остаточну модель навчили з використанням усіх функцій.

Як вже булу відзначено, продуктивність моделей суттєво відрізняється в різних наборів даних. Яскравим прикладом цього є модель читабельності, яка є найкращою моделлю класифікації для першого з розглянутих наборів даних, але найгіршою моделлю класифікації для другого набору даних, де модель працює так само добре, як і випадковий класифікатор. Це підкреслює ще одну з труднощів при спробі класифікувати фейкові новини. Різні домени фейкових новин можуть мати дуже різні характеристики, що ускладнює роботу зі створення узагальненої моделі класифікації фейкових новин.

На рисунку 2.2 Ще один цікавий експеримент, описаний у цій статті, полягає в навчанні моделей на одному наборі даних і тестуванні на іншому. Результатом цього були моделі зі значно зниженою продуктивністю, причому деякі з моделей знову працювали як випадкові класифікатори. Це підтверджує ідею, що різні домени фейкових новин можуть мати дуже різні характеристики. Це також підкреслює необхідність тестування моделей на наборах даних, окремих від тих, на яких їх навчали, щоб визначити,

наскільки узагальнені моделі для вирішення загальної проблеми виявлення дезінформації.

Features (number of features)	Acc.	LEGITIMATE			FAKE		
		P	R	F1	P	R	F1
Punctuation (11)	0.71	0.73	0.66	0.69	0.69	0.76	0.72
LIWC - Summary (7)	0.61	0.63	0.54	0.58	0.60	0.68	0.64
LIWC - Linguistic processes (21)	0.67	0.66	0.67	0.66	0.67	0.66	0.66
LIWC - Psychological processes (40)	0.56	0.56	0.57	0.56	0.56	0.56	0.55
Complete LIWC (79)	0.70	0.70	0.71	0.70	0.71	0.70	0.70
Readability (26)	0.78	0.82	0.72	0.77	0.75	0.84	0.79
Ngrams (651)	0.62	0.63	0.62	0.62	0.62	0.63	0.62
Syntax (1375)	0.65	0.66	0.63	0.64	0.64	0.67	0.65
All Features (2131)	0.74	0.75	0.73	0.74	0.74	0.75	0.74

Table 3: Classification results FakeNews dataset collected via crowdsourcing.

Features (number of features)	Acc.	LEGITIMATE			FAKE		
		P	R	F1	P	R	F1
Punctuation (11)	0.70	0.67	0.77	0.72	0.73	0.63	0.68
LIWC - Summary (7)	0.65	0.66	0.61	0.63	0.64	0.68	0.66
LIWC - Linguistic processes (21)	0.64	0.64	0.63	0.63	0.63	0.64	0.63
LIWC - Psychological processes (40)	0.58	0.58	0.58	0.58	0.58	0.57	0.57
Complete LIWC (79)	0.67	0.68	0.66	0.67	0.67	0.68	0.67
Readability (26)	0.50	0.50	0.48	0.49	0.50	0.51	0.50
Ngrams (1378)	0.67	0.67	0.66	0.66	0.66	0.68	0.67
Syntax (1268)	0.67	0.67	0.68	0.67	0.68	0.66	0.67
All Features (2751)	0.73	0.73	0.72	0.72	0.73	0.74	0.73

Table 4: Classification results for the Celebrity news data set.

Рисунок 2.2 – Результати класифікації фейкових новин (зі статті «Автоматичне виявлення фейкових новин»)

Результати статті [9] певною мірою відображають завдання, поставлені в даній кваліфікаційній роботі, пов’язані з навчанням моделей на новому наборі даних і перевіркою їх на існуючих наборах даних, щоб безпосередньо їх порівняти з ефективністю вже існуючих моделей. Однак відзначимо, що завдання статті [9] дещо відрізняється від завдань кваліфікаційного, оскільки в цій статті розглядається лише проблема бінарної класифікації, що не піддається поясненню, тоді як в даній кваліфікаційній роботі розглядаються багатокласові пояснювані моделі.

Друга стаття, що обговорюється нижче, більше корелює з завданнями кваліфікаційної роботи, оскільки вона стосується багатокласової проблеми класифікації фейкових, справжніх і сатиричних новин [10]. Для вирішення

цієї проблеми класифікації в [10] використовуються три різні набори даних, метадані для яких наведені в таблиці 2.2. Перший і третій набори даних призначені для задачі бінарної класифікації, а другий – для задачі з кількома класами. Основна увага цього обговорення буде зосереджена на другому наборі даних, оскільки він є найбільш актуальним для роботи, що виконується в кваліфікаційній роботі. Цей набір даних має свої плюси та мінуси, тому що невеликий набір даних робить навчання складних моделей із великою кількістю функцій відносно складним, однак якість даних висока, оскільки кожна новинна стаття вибирається вручну, щоб гарантувати, що аналізовані статті охоплюють однаковий діапазон тем з того ж періоду часу.

Таблиця 2.2 – Метадані для наборів даних, використаних у цитованій статті

Метадані набору даних			
Датасет ID	Реальні новини	Фейкові новини	Сатиричні новини
1	36	35	0
2	75	75	75
3	4000	0	233

В [10] розглядається генерація великої кількості вхідних характеристик з трьох різних категорій (складність, психологія та стилістичні особливості), а потім досліджується, як слід розрізняти кожну з цих ознак між фейковими, справжніми та сатиричними новинними статтями. Однак під час навчання відповідних моделей використовуються лише чотири найкращі характеристики (ймовірно це пов’язано з малим набором даних, що ускладнює навчання складнішої моделі без переобладнання даних). Потім у [10] використовується метод «усі проти всіх» для створення моделей, це означає, що окремий бінарний класифікатор навчається для під проблем «справжнє проти фейку», «справжнє проти сатири» та «фейк проти сатири». Результати

точності моделей такі: фейк проти реального 71%, сатира проти реального 91%, сатира проти фейку 67%.

Мета роботи [10] багато в чому подібна до мети кваліфікаційної роботи, через що її результати можуть бути використані для прямого порівняння якості моделей, що побудовані в [10] та в кваліфікаційній роботі. Кваліфікаційна робота буде спрямований на навчання моделей на більшому новому наборі даних і використовуватиме набір даних із [10] як набір для перевірки (це дозволить, зокрема, безпосередньо оцінити продуктивність побудованої моделі). Проте відзначимо, що завдання статті [10] також дещо відрізняється від завдань кваліфікаційної роботи, оскільки [10] все ще має багато відмінностей від цієї роботи, оскільки в ній не розглядається можливість створення єдиної багатокласової моделі за допомогою зрозумілих методів.

2.3 Сучасний стан проблеми використання машинного навчання для виявлення фейкових новин

В сучасних дослідження відзначається, що хоча читачі не можуть точно ідентифікувати фейкові статті з першого погляду, точність виявлення фейків можна підвищити, заохочуючи читача витрачати більше часу на перегляд статті. Доведено, що позначення статей тегами змушує більшість людей уважніше розглядати статтю, що дозволяє їм частіше ідентифікувати фейкові новини, однак широке використання тегів попередження також знижує впевненість людей у легітимності новин загалом. Варіант позначення тегами, який виявився більш ефективним, полягає в представленні альтернативних пояснень, коли стаття позначена тегами.

Розглянесо деякі роботи, спрямовані на ідентифікацію фейкових новин за допомогою зрозумілих моделей машинного навчання. Однією з останніх таких спроб є фреймворк dEFEND, створений у 2019 році [11]. Ця структура враховує текст статті, а також коментарі користувачів для кожної

статті та класифікує статті як справжні або підроблені, при цьому надаючи пояснення щодо своєї класифікації. Це робиться шляхом ранжування того, наскільки «заслуговує перевірки» те чи інше речення або коментар користувача (тобто ймовірність того, що він містить доведений факт), а потім пов'язує коментарі та речення з різними думками. Найбільш характерні з цих різних думок можна показати користувачеві як спосіб надання пояснення класифікації.

В таблиці 2.3 наведені показники продуктивності фреймворку dEFEND , які свідчить про те, що зрозумілі методи машинного навчання можуть конкурувати з точки зору ефективності класифікації з їх нез'ясовними (чорний ящик) аналогами.

Таблиця 2.3 – Порівняння продуктивності системи Defend з відомими системами класифікації для виявлення фейкових новин

Datasets	Metric	RST	LIWC	text-CNN	HAN	TCNN-URG	HPA-BLSTM	CSI	dEFEND
PolitiFact	Accuracy	0.607	0.769	0.653	0.837	0.712	0.846	0.827	**0.904**
	Precision	0.625	0.843	0.678	0.824	0.711	0.894	0.847	**0.902**
	Recall	0.523	0.794	0.863	0.896	0.941	0.868	0.897	**0.956**
	F1	0.569	0.818	0.760	0.860	0.810	0.881	0.871	**0.928**
GossipCop	Accuracy	0.531	0.736	0.739	0.742	0.736	0.753	0.772	**0.808**
	Precision	0.534	**0.756**	0.707	0.655	0.715	0.684	0.732	0.729
	Recall	0.492	0.461	0.477	0.689	0.521	0.662	0.638	**0.782**
	F1	0.512	0.572	0.569	0.672	0.603	0.673	0.682	**0.755**

Продуктивність фреймворку dEFEND перевершує продуктивність відомих систем класифікації. Однак для фреймворку dEFEND потрібен доступ не лише до тексту статей, а й до коментарів користувачів, а це означає, що він менш здатний класифікувати новинні статті, які не мають коментарів користувачів, або ті статті, які мають обмежені коментарі користувачів. Він також значно складніший за багато інших фреймворків, з якими він себе порівнював, і як наслідок, потребуватиме більше часу для навчання та класифікації.

Підвищена складність фреймворку dEFEND порівняно з деякими фреймворками, що не підлягають поясненню, що обговорюються,

підкреслює підвищену складність у створенні фреймворку, який можна пояснити. Фреймворк dEFEND був побудований з філософією post hoc пояснюваності, використовуючи основну модель для створення класифікацій і додаткову модель генерації пояснень, тісно пов’язану з основною. В кваліфікаційній роботі увага приділяється саме внутрішній пояснюваності, причому в поясненій моделі використовуються зрозумілі функції і внутрішньо пояснювані моделі (зокрема, SVM).

3 МЕТОДОЛОГІЯ ТА РЕАЛІЗАЦІЯ ПРОЕКТУ

3.1 Методологія проекту

У цьому розділі буде обговорено методологію, використану для кваліфікаційної роботи. Ця методологія складається з 5 основних етапів, шо показано на рисунку 3.1:

– вибір набору даних;

– генерація ознак;

– перехресна перевірка моделі та навчання;

– пояснюваність моделі;

– оцінка моделі.

3.1.1 Збір даних та створення функції

Першим етапом розробки будь-якої системи машинного навчання є вибір набору даних. Основна складність такого вибору у просторі фейкових новин полягає у пошуку великих надійних наборів даних. Жоден із існуючих наборів даних не цілком підходить для цієї роботи, оскільки в них або бракує потрібних полів даних, або вони містять замало статей, або не містять даних для справжніх, фейкових і сатиричних класів. У результаті було прийнято рішення створити новий набір даних для цілей цієї роботи. Існує велика кількість видань, яким довіряють і які оприлюднюють достовірні новини, а сатиричні видання самі рекламують свою роботу як таку. У результаті дані для цих двох категорій даних відносно легко зібрати. Однак виявлення джерела фейкових новин є більш складним завданням, оскільки за своєю природою публікації фейкових новин намагаються уникнути ідентифікації. Після збору даних набори даних необхідно очистити та стандартизувати. Цей процес зменшує розмір загального набору

даних, однак мати менший надійний набір даних краще, ніж більший ненадійний.

Рисунок 3.1 – Методологія роботи для розробки системи класифікації

Одним із найважливіших етапів розробки системи машинного навчання є вибір правильних функцій введення. Вибрані характеристики повинні забезпечувати високу ефективність класифікації та дозволяти створювати моделі, які можна пояснити. Для вибору функцій моделі було проаналізовано багато документів, у яких викладено деякі з найбільш відмінних рис між достовірними, фейковими та сатиричними статтями (наприклад, фейкові новини містять багато заголовків,

використовують простіший, повторюваний вміст у тексті, більш схожий на сатиру, ніж на справжні новини)

Також були проаналізовані статті, де висвітлюються функції, які можна використовувати для розробки пояснюваних моделей, таких як «Зрозуміле машинне навчання для виявлення фейкових новин». Вибрані функції включають поєднання найкращих функцій для обох підходів.

Після вибору функцій їх потрібно згенерувати для кожної новинної статті в наборі даних. Залежно від функцій, цей процес може зайняти досить багато часу, тому генерувати функції кожного разу, коли потрібно працювати над системою, непрактично. У результаті функції генеруються один раз, а потім зберігаються у файлі CSV (значення, розділені комами), де за потреби до них можна отримати доступ для навчання чи тестування моделей. Функції мають широкий діапазон можливих значень, де деякі функції (наприклад, «Кількість слів») мають значення в сотнях, а інші функції (наприклад, «Відсоток стоп-слів у заголовку») мають значення в десяткових дробах. З цієї причини всі дані потрібно нормалізувати. Цей набір нормалізованих функцій має зберігатися у файлі CSV, щоб його можна було зручно використовувати протягом усієї реалізації проекту без повторного обчислення.

3.1.2 Перехресна перевірка моделі та навчання

На цьому етапі вибираються існуючі моделі, які добре працюють у класифікації новин та є внутрішньо пояснювальними. Такі моделі (зокрема, модель SVC) завжди вибираються замість тих моделей, які мають лише одну з бажаних ознак.

Після вибору моделей використовується K-кратна перехресна перевірка для вибору гіперпараметрів для кожної з вибраних моделей з метою оптимізації ефективності їх класифікації. Зазвичай використовується значення K в діапазоні від 5 до 10, які визначаються згідно з розміром

набору даних. Більші значення *K* дозволяють використовувати більше даних для навчання, але зменшуютьвідсоток даних, які використовуються для тестування моделі, що, як правило, призводить до більшого переобладнання. Для цього проекту було вибрано значення $k = 5$, оскільки пріоритетом є уникнення переобладнання. Коли перехресна перевірка визначила оптимальні гіперпараметри, можна навчити моделі, оптимізовані для продуктивності класифікації.

3. 2 Пояснюваність та ефективність моделі

Останнім етапом розробки цієї структури машинного навчання є визначення та покращення пояснюваності моделей, розроблених на попередньому етапі. Фреймворк призначений для використання професіоналами у сфері виявлення фейкових новин. Однак мету його використання можна розширити, коли мова заходить про зрозумілість моделі. Розширена мета полягає в тому, щоб пересічний читач новин міг зрозуміти пояснення, надане для класифікації систем.

Необхідно перевіряти пояснюваність кожної з моделей. Деякі моделі, такі як SVC, можна пояснити прямим відображенням їхніх вхідних характеристик на вихід, тоді як результати інших моделей, таких як нейронні мережі, пояснити набагато важче. Це передбачає необхідність видалення незрозумілих функцій і генерування моделей, у яких пояснювані функції є найбільш відмітними вхідними функціями. Впровадження та покращення пояснюваності в обох типах моделей може знизити ефективність їх класифікації, однак це буде до певної міри допустимо, оскільки створення високоефективних моделей, які не підлягають поясненню, не є метою цієї роботи.

На завершальному етапі необхідно порівняти моделі, створені та оптимізовані на попередніх двох етапах. Базові моделі також будуть створені для порівняння з моделями, навченими на попередніх етапах.

Моделі будуть ранжовані від найкращих до найгірших на основі їх класифікаційних показників і будуть позначені як такі, що піддаються поясненню, або як такі, що не підлягають поясненню. Ідеальна модель – це така, яка має відносно високу ефективність класифікації, але все ще відповідає мінімальному порогу пояснюваності.

Для оцінки ефективності класифікації моделей буде реалізовано чотири KPI, які використовуються як стандартні в цій галузі. Кожен із цих показників забезпечує розуміння певного аспекту ефективності класифікації моделей.

3.3 Вибір наборів даних та функцій

Для цілей цього проекту використовуються два набори даних, що важливо, оскільки метою цієї кваліфікаційної роботи є створення узагальненої моделі ML для проблеми виявлення дезінформації. Якщо дані навчання та тестування взяті з одного відносно невеликого набору даних, важко визначити, чи модель засвоїла загальні риси проблеми та є здатна переобладнувати конкретні характеристики точок даних у наборі даних.

Перший набір даних, який буде обговорено – це набір даних, створений для цілей цього проекту, цей набір даних будемо називатися набором даних 1. Набір даних 1 було створено за допомогою комбінації веб-скребків і взяття даних із загальнодоступних наборів даних. Статті новин «BBC News» були використані як надійне джерело новин, оскільки це поважне видання новин по всьому світу. Публікація "Waterford Whispers" є сатиричним виданням, яке самовизначилося і тому використовувалося як джерело сатиричних новин. Веб-скребки були розроблені з використанням бібліотеки Selenium, яка дозволяє сценаріям Python взаємодіяти з веб-сайтами та отримувати дані з їхніх HTML-сторінок. Цей веб-скребок переміщувався веб-сайтами, виділяючи кожну статтю з подальшим витягом необхідних даних та переходом до наступної статті. Без експерта з

виявлення фейкових новин неможливо зібрати нові дані для статей фейкових новин, тому використовувався встановлений набір даних статей фейкових новин. Цей набір даних слід вважати надійним, оскільки статті були класифіковані організаціями з перевірки фактів в Україні, Америці і набір даних використовувався в кількох широко цитованих публікаціях.

Загалом близько 3000 сатиричних статей і 3000 надійних статей було вибрано з Інтернету, а завантажений набір фейкових новин містив трохи менше 9000 статей. Існує дисбаланс даних: фейкових новин майже втричі більше, ніж справжніх і сатиричних. Це зменшило ефективний розмір набору даних з майже 15 000 до трохи менше 9 000, оскільки навчальні моделі на незбалансованих наборах даних часто можуть спричинити зміщення в моделях. Для кожної статті зберігалися назва статті, текст статті та дата публікації. Після того, як дані були зібрані, їх потрібно було обробити та очистити, щоб гарантувати, що в наборах даних зберігаються лише попередньо оброблені дані. Вилучені набори даних вимагали значної роботи, оскільки програма, використана для видалення даних із веб-сайтів, була неідеальною. Спочатку було видалено будь-які повторювані статті, а потім усі статті, у яких відсутні поля «текст статті» або «назва». Після цього всі статті з думок було видалено з надійного набору даних. Нарешті всі статті, що містять пошкоджені дані, були видалені. Загалом цей процес видалив із набору даних понад 500 справжніх статей і близько 100 сатиричних статей.

Другий набір даних, який було використано – це існуючий набір даних [8], який будемо називати набором даних 2. Набір даних 2 було вибрано, оскільки він був одним із небагатьох якісних наборів даних, що містив статті, позначені як фейкові, справжні та сатиричні, які використовувалися в надійних джерелах. Була розглянута ідея використання даних із різних цих наборів даних для створення більшого другого набору даних, однак дані в наборі даних 2 дуже добре зібрані та містять статті, що охоплюють ті самі теми, з різних публікацій, протягом

фіксованого періоду часу. Це робить його ідеальним набором даних-кандидатів для цілей цієї кваліфікаційної роботи, оскільки другий набір даних здебільшого потрібен як дані для тестування моделей, навчених на наборі даних 1.

Набір даних 2 містить 75 фейкових, справжніх і сатиричних статей, як показано в таблиці 3.1. Головним недоліком цього набору даних є його розмір, оскільки для навчання складних моделей із великою кількістю функцій зазвичай потрібні великі набори даних, щоб уникнути переобладнання. Ймовірно це пояснює, чому моделі, навчені за допомогою цього набору даних у документі, на який посилається, використовують лише моделі з чотирма вхідними функціями.

Таблиця 3.1 – Метадані двох наборів даних

	Набір даних 1	Набір даних 2
Реальні статті	2874	75
Фейкові статті	8898	75
Сатиричні статті	3097	75
Всього	14,869	225

Для цього проекту використовуються два основні набори функцій, перша – це базовий набір функцій, який називатиметься набором функцій 1, а другий набір – це уточнений набір функцій, який називатиметься набором функцій 2, таблиця. 3.2.

В роботах [8], [12] висвітлюються функції, які використовують широкий спектр характеристик (близько 50) для розрізнення справжніх, фальшивих та сатиричних новинних статей. Цей набір із 50 ознак було додатково скорочено шляхом видалення будь-яких незрозумілих функцій (таких як середня глибина дерева дієслівних фраз і середня глибина синтаксичного дерева), які не є визначальними з точки зору ефективності класифікації. Після цього було залишено лише 18 вхідних функцій, перелік яких наведено в таблиці 3.2.

Таблиця 3.2 – Список функцій, що використовуються в цьому проекті

Дані функції до та після зменшення функції			
Функція введення	Класифікація використання	Перший набір функцій	Другий набір функцій
Довжина назви	Real vs Satire	✓	✓
Число власних іменників у назві	Real vs All	✓	✓
Кількість дієслів минулого часу в назві	Low Overall	✓	×
Відсоток стоп-слів у заголовку	Real vs All	✓	×
Середня довжина слів у заголовку	Satire vs All	✓	✓
Підрахунок слів статті	Real vs All	✓	✓
Середня довжина речення	Satire vs All	✓	✓
Індекс читабельності тексту	Low Overall	✓	×
Оцінка TTR	Fake vs Satire	✓	✓
Кількість цитат у тексті	Fake vs Satire	✓	×
Середня довжина слів у тексті	Satire vs All	✓	✓
Число особових займенників	Satire vs All	✓	✓
Кількість прислівників	Real vs All	✓	✓
Кількість розділових знаків	Fake vs Real	✓	✓
Максимальний негативний настрій	Real vs All	✓	✓
Максимальний позитивний настрій	Real vs All	✓	✓
Середній негативний настрій	Fake vs All	✓	✓
Середній позитивний настрій	Real vs Fake	✓	✓

Набір функцій 1 використовувався для створення першого покоління моделей, після чого було проведено детальний аналіз функцій. По-перше, було досліджено відносну важливість кожної з характеристик у визначенні результату роботи моделей. Це було зроблено шляхом створення графіків, які підкреслюють різницю в коефіцієнтах моделі (або вагових коефіцієнтах) для виявлення фейкових, справжніх і сатиричних новинних статей. Щоб проілюструвати цей процес, використаємо функцію довжини заголовка. У моделі багатокласової логістичної регресії ваги для вхідної функції «Довжина назви» становили 2,4, 15,4 і -17,4 для фейку, справжнього та сатири відповідно. Отримавши величину різниці цих ваг, можна побачити, що функція «Довжина назви» найкраще відрізнить справжні статті від сатиричних статей, оскільки величина різниці становить 32,8 порівняно з 19,4 і 13. Цей метод використовувався для створення рисунків, які допомогли визначити найефективніші характеристики в усіх моделях, що показано на рисунках 3.2, 3.3, 3.4 для багатокласової моделі. Цей процес

висвітлив деякі неефективні функції, а саме «Кількість дієслів минулого часу» та «FK читабельність тексту», які потім не були включені до набору функцій 2.

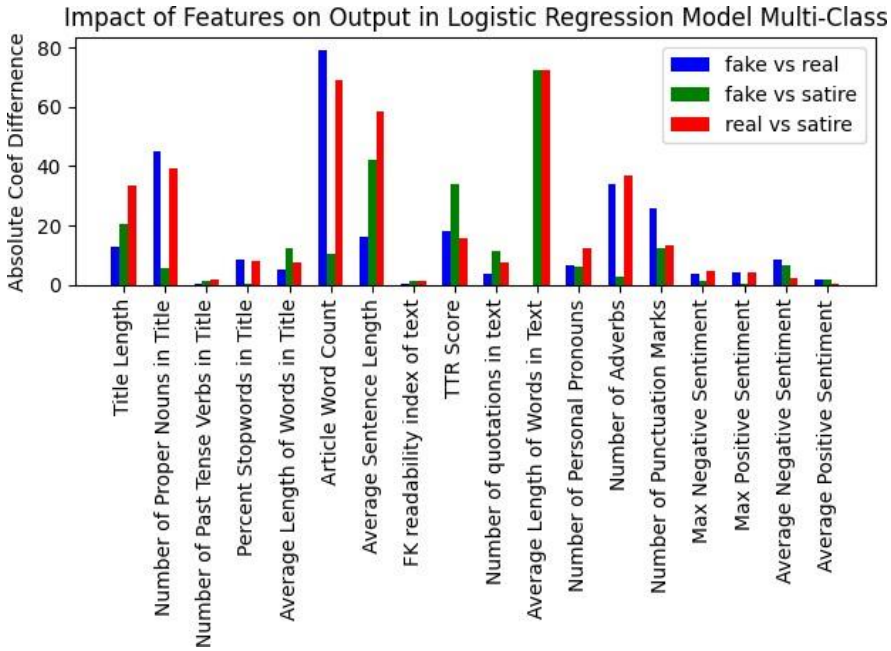

Рисунок 3.2 – Логістична регресія багатокласового аналізу ознак

Наступний крок пов'язаний з дослідженням поширеності кожної з функцій у різних статтях. Цей процес передбачав пошук кількості статей, у яких функція не з’явилася. Функції, які рідко зустрічаються в наборі даних, є менш цінними для проекту, який зосереджується на створенні узагальненої моделі для класифікації новин. Функції, які рідко зустрічаються в наборі даних, також можуть штучно підвищити продуктивність моделей через переобладнання. Лише дві вхідні функції мають високий відсоток нульових вхідних даних і це були «Кількість цитат» і «Кількість дієслів минулого часу», які мали нульові вхідні дані для 93,83% і 81,16% статей у наборі даних. Для контексту наступний найвищий бал для функції становив 4,36%. Дві функції, визначені як такі, що мають велику частку нульових вхідних даних, які не були включені до набору функцій 2.

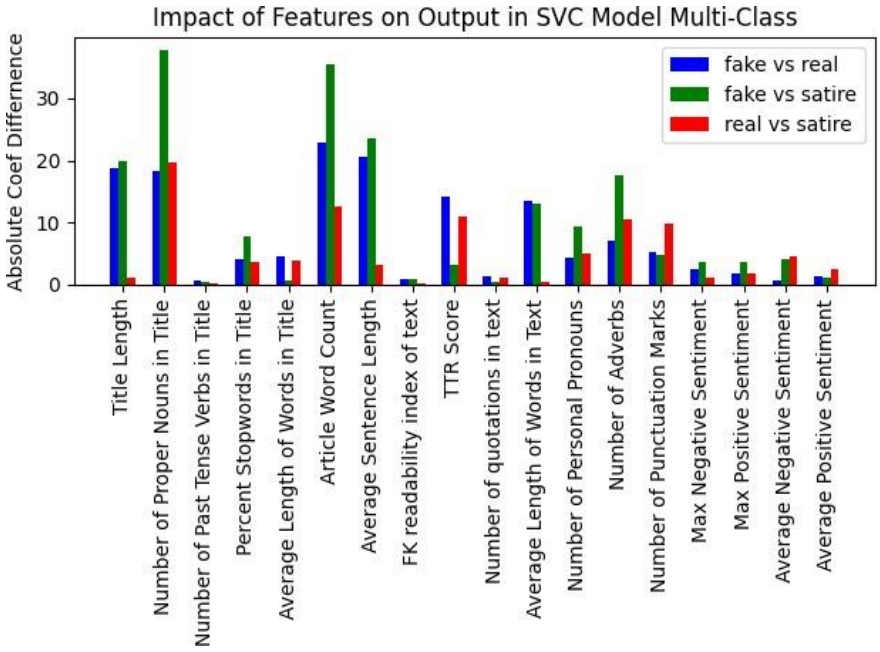

Рисунок 3.3 – Багатокласовий аналіз функцій SVC

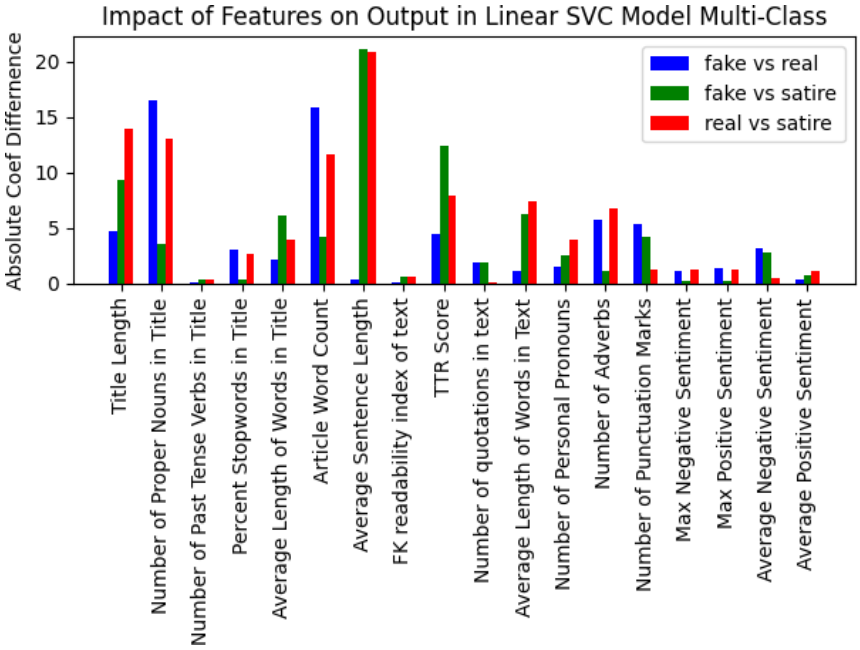

Рисунок 3.4 – Лінійний SVC багатокласового аналізу функцій

Наприкінці було зроблено остаточний аналіз функцій, щоб видалити незрозумілі функції з набору функцій 2. Під час аналізу функції «Відсоток стоп-слів у заголовку» було помічено, що ця функція може бути складною для розуміння деяким читачам і в результаті її не було включено до набору

функцій 2. Загалом чотири функції з набору функцій 1 не були включені до набору функцій 2. Розподіл функцій наведено в таблиці 3.2.

Створення значущих функцій із великих об’ємів тексту є складним процесом, однак існують бібліотеки обробки природної мови, доступні для безкоштовного використання, такі як NLKT (набір інструментів природної мови) [13]. Бібліотека NLKT надає широкий спектр надійних інструментів для аналізу тексту, таких як перевірка типу слів і аналіз настроїв. Ця бібліотека використовується для створення чотирьох функцій, пов’язаних із настроєм тексту статті, вона також використовується для підрахунку кількості складів у кожному слові статті, що було необхідно для створення індексу читабельності FK тексту статті. Іншою корисною бібліотекою є тег POS (частина мови), який можна використовувати для підрахунку певних типів слів, таких як власні іменники, особисті займенники тощо в тексті. Ця бібліотека була використана для створення кількох ознак, які вимагали підрахунку певних мовних особливостей у тексті чи назві статті. Ці інструменти важливі, оскільки вони дозволили створити багато функцій, які були б надто складними, щоб створити їх з нуля за наявний час. Решта функцій були створені за допомогою стандартних бібліотек Python.

Створення деяких із цих функцій потребувало значного часу, особливо індекс читабельності FK. У результаті вони були створені лише один раз для кожного набору даних, а потім ці згенеровані функції зберігалися у файлі CSV для подальшого використання. Коли характеристики були згенеровані, стало зрозуміло, що їх потрібно буде нормалізувати, оскільки деякі функції мали вхідні значення в десятках і сотнях, тоді як інші були обмежені десятковими дробами. Нормалізація гарантує, що всі значення зберігають відносний розмір у межах заданої функції, але всі вхідні функції обмежені значеннями від 0 до 1. Це гарантує, що модель не вважає деяку функцію більш важливою, якщо її вхідні значення є великими в абсолютному вимірі. Подібним чином нормалізовані значення вхідних характеристик обчислювалися один раз, а потім

зберігалися у файлі CSV, щоб заощадити час на обчислення в майбутньому. Приклади початкової вхідної функції та нормалізованої вхідної функції наведені на рисунку 3.5. Всі нормалізовані значення знаходяться в діапазоні від 0 до 1, але відносне значення кожного вхідного параметра в даній функції зберігається (наприклад. перший вхід для першої статті).

```
Inital Features
8.0,8.0,0.0,0.0,7.62,305.0,14.86,63.0,0.67,0.0,5.21,8.0,10.0,28.0,0.554,0.487,0.089,0.077
15.0,8.0,0.0,3.75,5.27,580.0,12.72,60.0,0.58,0.0,5.17,36.0,15.0,99.0,0.543,0.209,0.15,0.057

Normalised Features
0.1905,0.2963,0.0,0.0,0.7954,0.0609,0.0953,0.525,0.67,0.0,0.0854,0.0209,0.036,0.0509,0.554,0.487,0.1584,0.077
0.3571,0.2963,0.0,0.1974,0.5501,0.1158,0.0815,0.5,0.58,0.0,0.0848,0.094,0.054,0.18,0.543,0.209,0.2669,0.057
```

Рисунок 3.5 – Початкові функції введення та нормалізація вхідних функцій для перших двох статей

Найкраща практика відбору моделей класифікації полягає в тому, щоб вибрати кілька перспективних моделей, навчити класифікатор для кожної, а потім вибрати модель, яка працює найкраще. Для цього проекту було розглянуто чотири моделі: логістична регресія, KNN, SVC і лінійна SVC. Кожна з них була реалізована за допомогою бібліотеки sklearn python.

Також було використано дві стандартні базові моделі класифікації, щоб надати контекст результатам, досягнутим при використанні обраних для дослідження моделей. Першою з базових моделей в проекті обрано випадковий базовий класифікатор, який вибирає випадковий вихід із можливих вихідних даних для будь-якої даної класифікації. Будь-яка модель, продуктивність якої є близькою до базової лінії, вважатиметься такою, що не має цінності.

Друга базова модель, яка була обрана – це модальна базова лінія, що завжди передбачає найпоширеніший вихід у навчальному наборі даних. Вона допомагає виявити будь-які моделі, які досягають високої продуктивності, віддаючи перевагу одному з більших класів у наборі даних [14]. Логістична регресія є одним із найпростіших для реалізації

алгоритмів навчання, тому її часто спочатку тестують. Згідно з формулою (3.1) логістична регресія для вхідного вектора намагається знайти вагу, що відповідає кожній вхідній функції, яка мінімізує наступну функцію витрат за допомогою градієнтного спуску:

$$J(\theta) = \frac{1}{m}(\sum_{i=1}^{m} \log\left(1 + e^{-y^i \theta^T x^i}\right) + \frac{1}{2c}\sum_{j=1}^{n} \theta_j^2)). \quad (3.1)$$

Для підвищення ефективності моделі логістичної регресії було застосовано регулярізацію L2. Це означає, що ваги з незначним впливом на визначення результату, прагнуть до 0. Значення *c* є гіперпараметром, який використовується для визначення розміру штрафу, при цьому більші значення с призводять до менших штрафів, а менші значення *c* – до більших штрафних санкцій. Розмір штрафу використовується для врівноваження надмірної та недостатньої комплектації моделі [15]. Логістична регресія є швидким і простим у застосуванні методом, який можна пояснити та який показав свою ефективність у певних обставинах у просторі класифікації новин.

KNN – алгоритм, що здійснює свої прогнози безпосередньо на основі даних навчання замість створення моделі з ваговими коефіцієнтами. Він оцінює результат на основі наявних навчальних точок, які є «найближчими» до входу.

У цьому проекті було реалізовано гаусове ядро, тому точки, які розташовані ближче до вхідних даних, мають перевагу над точками, які розташовані далі від вхідних даних. Значення *K* визначає кількість балів, які враховуються для даної класифікації, було використано різноманітні значення *K* через різні розміри використаних наборів даних.

Значення γ є гіперпараметром, який визначає ступінь переваги, що надається ближчим точкам даних (більші значення γ надають більшу перевагу ближчим точкам даних, а менші значення γ – меншу) [16].

Генерація прогнозів при цьому (для ядра Гауса) здійснюється за наступною формулою:

$$Prediction = \frac{sum(w^i y^i)}{sum(w^i)}, \tag{3.2}$$

де y^i – значення точки даних; $w^i = e^{-\gamma * d(x^i, x)}$; $d(x^i, x)$ – відстань між точкою даних x^i і входом x.

Модель KNN було обрано для дослідження, оскільки в ній використовується принципово відмінний метод в порівнянні з іншими вибраними моделями.

Однак слід зазначити, що KNN не піддається внутрішньому поясненню, і як наслідок, незалежно від її класифікаційних характеристик, модель KNN ніколи не цілком може задовольнити вимогам цієї роботи.

Модель SVC працює подібно до логістичної регресії, намагаючись знайти набір вагових коефіцієнтів, які відповідають вхідному вектору, що мінімізує функцію втрат [17].

Різниця полягає в тому, що функція втрат логістичної регресії намагається зменшити помилку всіх точок даних у наборі даних, тоді як функція втрат SVC намагається створити гіперплощину, яка максимізує розділення класів, використовуючи ідею опорних векторів, шо показано на рисунку 3.6.

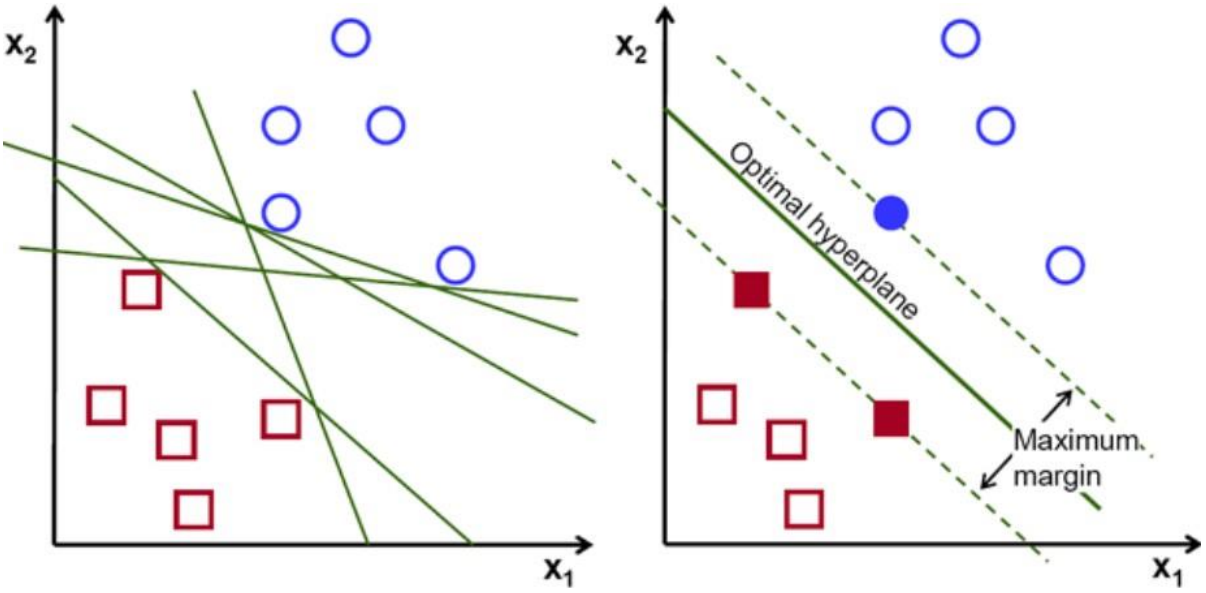

Рисунок 3.6 – Приклад SVC, що створює гіперплощину між 2 класами

Ідея використання опорних векторів полягає в тому, що лише точки даних біля кордонів класів фактично мають значення при розділенні класів, тому точки даних на кордонах класів вибираються як опорні вектори і після цього формується лінія, яка максимізує розділення цих опорних векторів [6]. У реальних даних зазвичай існує перекриття між класами і метод SVC здатний керувати цим перекриттям. Базовий принцип метода реалізується згідно з наступним рівнянням:

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} max\left(0,1 - y^i\theta^T x^i\right) + \frac{\theta^T\theta}{c}. \qquad (3.3)$$

Метод SVC від sklearn реалізується методом мультикласифікації «1 проти 1», згідно з яким при наявності кількох вихідних класів модель навчається для кожної проблеми двійкової класифікації в межах багатокласової проблеми. Прогнозований вихід береться як класифікатор з найбільшою впевненістю, що вхід є даним класом. SVC також може використовувати різні ядра, найпопулярнішим з

яких є ядро RBD, однак для цілей цього проекту було вибрано лінійне ядро, оскільки це єдине ядро, яке дотримується принципу внутрішньої пояснюваності.

SVC було обрано, оскільки це одна з найефективніших моделей із наявних робіт у галузі класифікації новин, особливо для проблем бінарної класифікації.

Лінійний SVC має багато спільного з описаною вище моделлю SVC, особливо коли використовується лінійне ядро [15]. Однак є деякі ключові відмінності, які роблять його привабливим для практичної реалізації. По-перше, коли використовується sklearn, функція втрат для моделі SVC відрізняється від інших функцій втрат та розраховується наступним чином:

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} max\left(0{,}1 - y^i\theta^T x^i\right)^2 + \frac{\theta^T\theta}{c}. \quad (3.4)$$

По-друге, лінійна модель SVC реалізує метод мультикласифікації «1 проти всіх», тобто ця модель навчається для кожного класу з усіма іншими класами разом.

Відзначимо, що лінійна модель SVC, також є однією з найбільш ефективних внутрішньо пояснювальних моделей класифікації новин, часто перевершуючи метод SVM.

3.4 Підготовка до реалізації експериментів проекту

У цьому підрозділі представлено чотири основні та визначено причину їх вибору. Кожен з цих експериментів, детально описано в наступному розділі.

Перший експеримент, який далі будемо називати E1, спрямований на створення набору базових моделей, які не є повністю зрозумілими, використовуючи набір даних 1 і набір функцій 1. Цей набір буде

використано для порівняння з моделями. Другий експеримент (E2) спрямований на створення зрозумілих моделей на наборі даних 1 з використанням набору функцій 2. Метою цього експерименту є вимірювання падіння продуктивності після того, як кількість функцій було зменшено, а модель стала повністю зрозумілою. Третій експеримент (E3) призначений для перевірки моделей, які створені в E2, відносно того, наскільки добре вони працюють під час тестування на новому наборі даних. Моделі, які використовують набір функцій 2, будуть навчені на наборі даних 1 і перевірені на наборі даних 2.

Третій експеримент є важливим експериментом для вимірювання рівнів переобладнання в моделях, навчених на наборі даних 1.

Нарешті, у четвертому експерименті (E4) створюватимуться моделі з використанням набору даних 2 і набору функцій 2, щоб визначити, наскільки ефективним є набір функцій 2 у наборах даних, оскільки ці функції були вибрані частково на основі оцінки їх продуктивності в наборі даних 1. Ця інформація узагальнена в таблиці 3.3.

Таблиця 3.3 – Умови проведення експериментів

Експеримент	Тренуються на	Тестуються на	Використаний набір функцій
E1	Набір даних 1	Набір даних 1	Набір функцій 1
E2	Набір даних 1	Набір даних 1	Набір функцій 2
E3	Набір даних 1	Набір даних 2	Набір функцій 2
E4	Набір даних 2	Набір даних 2	Набір функцій 2

У кожному з чотирьох основних експериментів реалізовано п’ять додаткових експериментів, як це показано в таблиці 3.4.

Таблиця 3.4 – Опис додаткових під експериментів

Під експеримент	Опис під експериментів
Багатокласо-вий	Багатокласові моделі генеруються з використанням усіх трьох класів даних
Фейк проти справжнього	Двійкові моделі генеруються з використанням лише підроблених і справжніх класів даних
Фейк проти сатири	Двійкові моделі генеруються з використанням лише класів даних фейку та сатири
Реальне проти сатири	Бінарні моделі створюються з використанням лише класів реальних і сатиричних даних
Справжнє проти всіх	Двійкові моделі генеруються з використанням реального класу та двох інших класів даних, об’єднаних в один клас

Під експеримент із кількома класами є основним експериментом, оскільки кінцевою метою цього проекту є створення ефективного класифікатора фейкових, справжніх і сатиричних новин, яких можна пояснити.

Однак дуже небагато найсучасніших досліджень у сфері класифікації новин розглядають багатокласові моделі, тому багатокласові моделі важко порівняти з більшістю моделей, отриманих в цій галузі. З цієї причини було реалізовано наступні три під експерименти: фейк проти реального, фейк проти сатири, реальний проти сатири. Це дозволяє забезпечити можливість більш прямого порівняння роботи, виконаної в цьому проекті з іншими роботами, виконаними в полі класифікації новин. Ці три під експерименти також дозволяють чіткіше зрозуміти труднощі використання багатокласових моделей, розроблених для початкового під експерименту.

Після проведення перших чотирьох під експериментів стало зрозуміло, що однією з головних проблем є відокремлення фальшивих статей від сатиричних статей. В результаті був розроблений п’ятий під експеримент – реальне проти всіх. Цей під експеримент був розроблений, щоб визначити наскільки ефективнішим буде класифікатор.

4 ОБЧИСЛЕННЯ ТА РЕЗУЛЬТАТИ ЕКСПЕРИМЕНТІВ НА ОСНОВІ РОЗРОБЛЕНОЇ МОДЕЛІ

4.1 Перший експеримент

Метою експерименту E1 є створення набору базових моделей, які можна використовувати для оцінки зміни продуктивності, коли моделі змінюють використання набору функцій 1 на набір функцій 2. Усі моделі, розглянуті в цьому розділі, були навчені та протестовані на наборі даних 1.

Після визначення набору даних зібрано, генерації функцій та обрання типу моделей було визначено процедуру їх навчання. Кожна модель має гіперпараметр, який впливає на результат роботи моделі. Ці гіперпараметри впливають на баланс переобладнання та недооблаштування в моделі, тому вибір оптимізованих гіперпараметрів є важливим кроком у навчанні високопродуктивних моделей, які не надто адаптовані до певного набору даних. Для вибору гіперпараметрів у цьому проекті використовувався метод *k*-кратної перехресної перевірки, що на сьогодні є стандартною практикою в ML. Набір даних сегментовано на дві частини у співвідношенні 80%/20% (80% використовуються для процесу перехресної перевірки та навчання моделі, а 20% залишаються осторонь для тестування моделей після визначення оптимальних гіперпараметрів). Це є ще одним кроком для виявлення переобладнання, оскільки, якщо існує велика невідповідність між продуктивністю ваших моделей на даних навчання та тестування, то очевидно, що моделі переобладнані і потрібен інший варіант вибору гіперпараметрів.

Під час п’ятикратної перехресної перевірки навчальні дані сегментуються на п’ять розділів однакового розміру. Для кожного значення гіперпараметра, яке потрібно перевірити, модель навчається на кожному з цих п’яти варіантів розподілу навчальних даних, як це показано на рисунку 4.1. Потім будуються графіки середніх значень продуктивності та

стандартних відхилень і за допомогою цих графіків вибирається оптимальний гіперпараметр. Графіки перехресної перевірки для під експерименту з кількома класами наведено на рисунках 4.2, 4.3, 4.4 і 4.5.

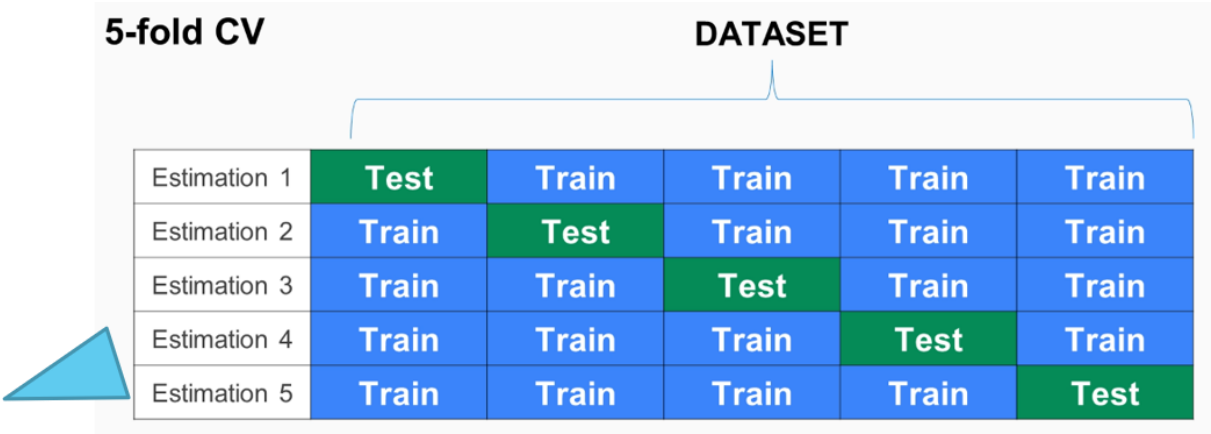

Рисунок 4.1 – Приклад 5-кратної перехресної перевірки

Перш ніж перейти до перехресної перевірки конкретних моделей, розглянемо значення декількох загальних параметрів, які використовувалися для моделей у цьому експерименті, а також деякі загальні конструктивні рішення. Усі моделі, навчені в E1, тренуються лише протягом максимум 10 000 ітерацій, це обмеження зроблено виключно через обмеження обчислень. Кількість сусідів, які розглядаються для моделі KNN становить 2000. Діапазони значень гіперпараметрів, які були протестовані, були обрані згідно з сучасним рівнем техніки (для кожної моделі діапазон гіперпараметрів був обраний таким чином, щоб моделі починалися з недообладнання та закінчувалися переобладнанням. Це робиться для того, щоб можна було знайти оптимальне значення гіперпараметра, навіть якщо воно сталося поблизу граничних значень діапазону).

На рисунку 4.2 показано графік перехресної перевірки для моделі багатокласової логістичної регресії. Гіперпараметром для логістичної регресії є значення *c*, що має обернену залежність від штрафу у функції витрат.

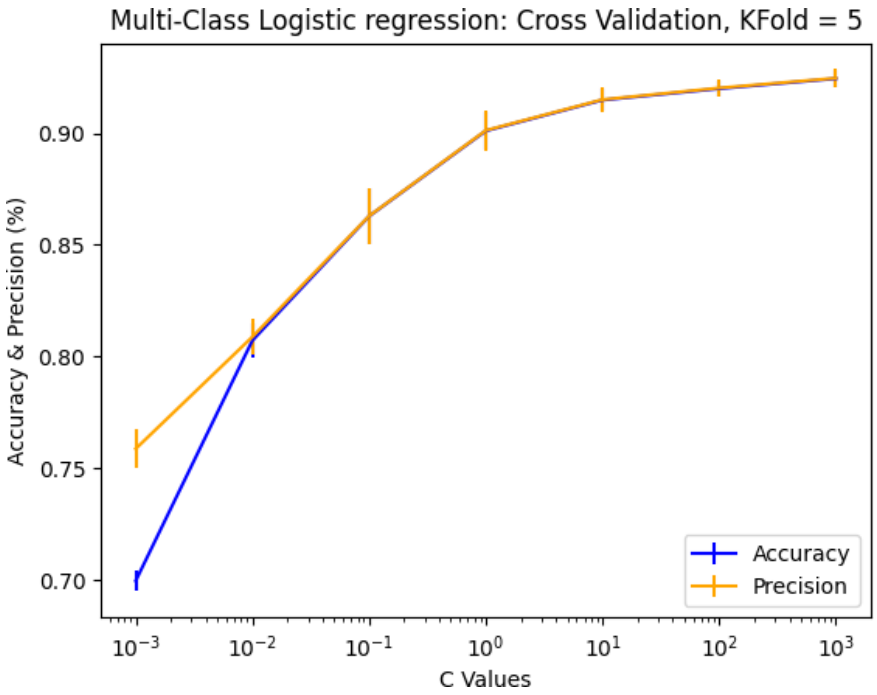

Рисунок 4.2 – Багатокласова перехресна перевірка логістичної регресії для E1

Коли *c* дорівнює 0,001, можна побачити, що і точність і прецизійність є відносно низькими через недостатню підгонку, потім вони швидко збільшуються до значення, коли *c* приблизно дорівнює 10, після чого починають вирівнюватися. Будь-які подальші переваги від більших значень *c* після цього відсутні, а незначні переваги в продуктивності можуть бути досягнуті за рахунок більш і більш точної підгонки моделі.

Для нашої моделі значення *c*, що використовуватиметься як гіперпараметр, дорівнює 10.

На рисунку 4.3 наведено графік перехресної перевірки для багатокласової моделі KNN. Гіперпараметром для KNN є значення γ. Відзначимо, що більші значення γ віддають перевагу ближчим точкам даних, а менші значення γ – меншим. Коли значення γ дорівнює 0, усі точки обробляються однаково. Обидва показники зростають до піку, коли γ досягає значення 100, де обидва показники мають найбільші бали та найменші помилки, після чого будь-які збільшення значення γ призводять

до зниження продуктивності моделі. Значення γ, що дорівнює 100, використовувалося як гіперпараметр для цієї моделі.

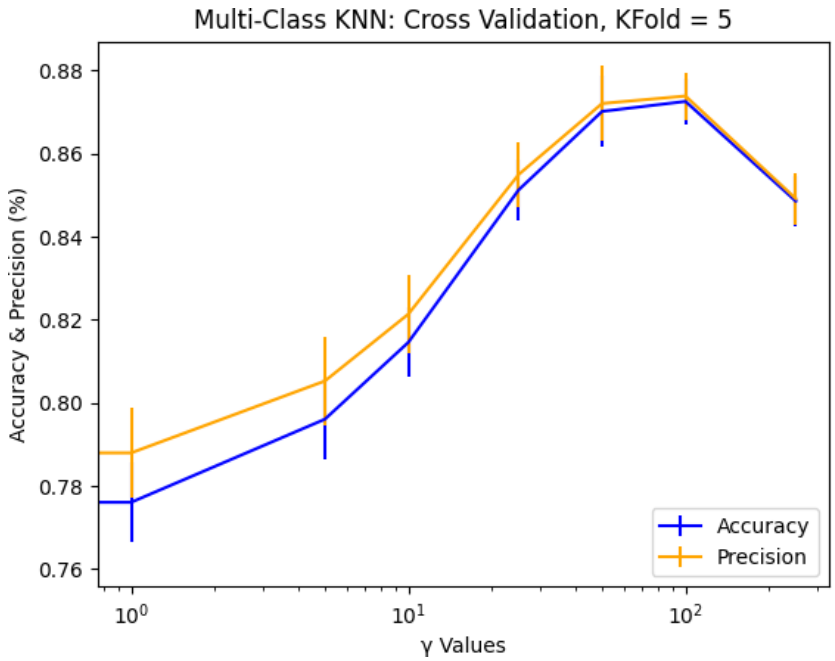

Рисунок 4.3 –Багатокласова перехресна перевірка KNN для E1

На рисунках 4.4 і 4.5 наведено графіки перехресної перевірки для багатокласової моделі SVC. Подібно до логістичної регресії, гіперпараметром для SVC є значення *c* , а його зв'язок із штрафом функції витрат є таким, що спонукає поведінку, подібну до тієї, що обговорювалася вище для перехресної перевірки в разі використання логістичної регресії. Коли *c* дорівнює 0,001 продуктивність моделі є поганою, а точність (ACC) і точність (PPV) значно нижчі за 50%, що вказує на значне недообладнання. Точність (ACC) і точність (PPV) стабільно зростають до значення *c*, що приблизно дорівнює 1, де збільшення вирівнюється вказуючи на наявність переобладнання. Для цієї моделі значення *c* , що дорівнює 1, використовувалося як гіперпараметр.

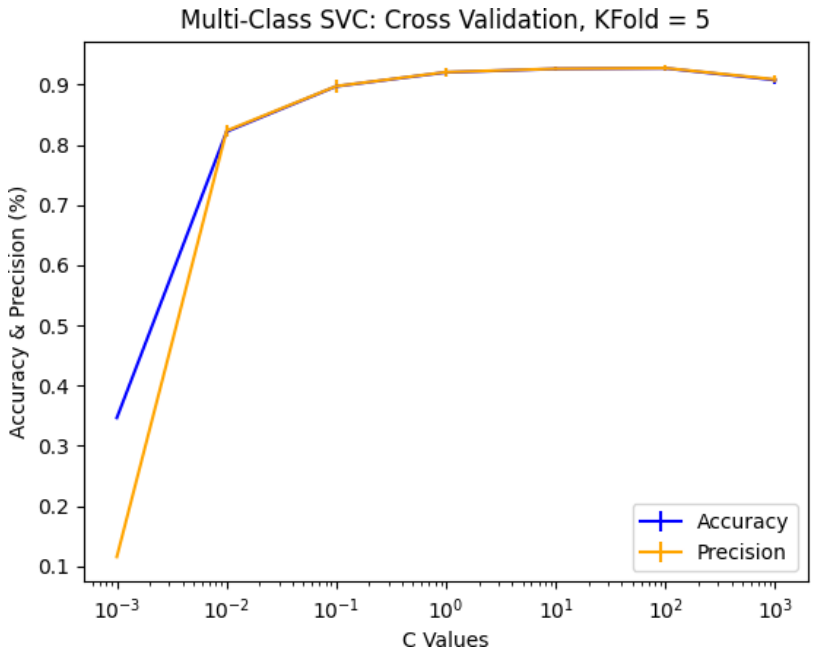

Рисунок 4.4 – Багатокласова перехресна перевірка SVC для E1

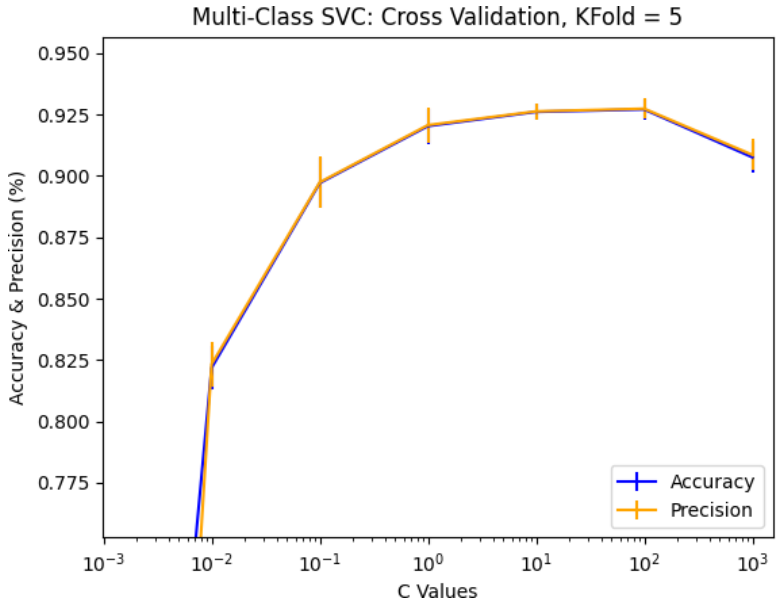

Рисунок 4.5 – Збільшена багатокласова перехресна перевірка SVC для E1

На рисунку 4.6 показано графік перехресної перевірки для багатокласової лінійної моделі SVC. Знову гіперпараметром є значення *c*, яке діє подібно до того, що обговорювалося для моделей логістичної регресії та SVC. Коли значення *c* дорівнює 0,001, точність (ACC) і

точність (PPV) є відносно низькими (близько 80%), що вказує на недостатнє пристосування. При збільшенні значення *c* до 1, ці показники починають вирівнюватися, вказуючи на переобладнання. Для цієї моделі значення *c*, що дорівнює 1, використовувалося як гіперпараметр.

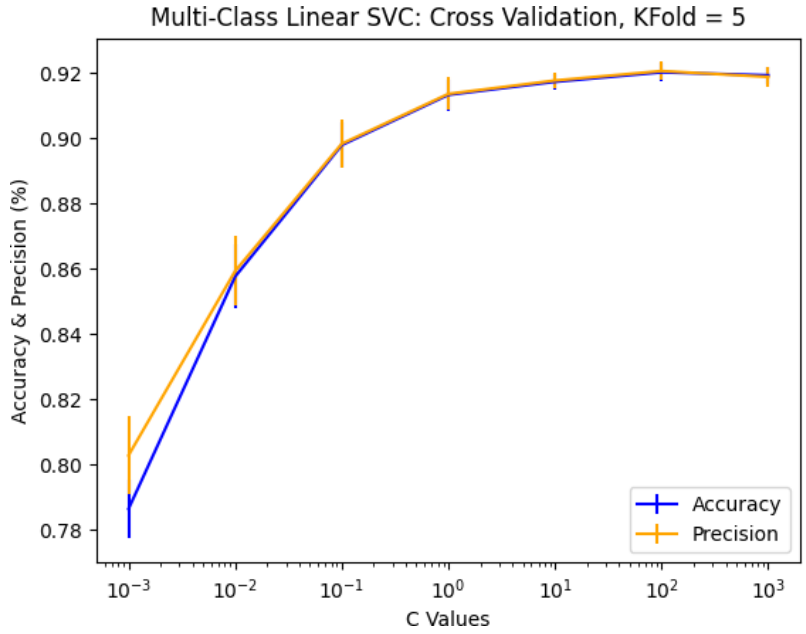

Рисунок 4.6 – Багатокласова перехресна перевірка лінійного SVC для E1

Процес було повторено для вибору гіперпараметрів для інших чотирьох під експериментів в E1. В таблиці 4.1 наведено отримані значення гіперпараметрів.

Таблиця 4.1 – Значення гіперпараметрів після перехресної перевірки для E1

Model	Multi-class	Fake vs Real	Fake vs Satire	Real vs Satire	Real vs All
LR	10	10	10	100	10
KNN	100	100	50	100	50
SVC	1	1	1	10	10
Linear SVC	1	10	100	10	1

Після того, як були обрані оптимальні гіперпараметри для кожної моделі в кожному з під експериментів експерименту E1, моделі були ще раз навчені за допомогою цих гіперпараметрів на 80% даних, збережених для навчання. Потім кожна модель була протестована на 20% даних. Кожна з моделей була оцінена за чотирма ключовими показниками, розглянутими в таблиці 2.1.

Після того, як були створені оптимальні моделі та порівняна ефективність їх класифікації для кожної під проблеми, особливості цих моделей були досліджені. Дослідження ознак проходило в три етапи: на першому етапі досліджувалась важливість кожної ознаки в кожній моделі кожної підпроблеми, на другому етапі було визначено наскільки узагальненою є кожна з цих ознак і завершальний етап – виявити будь-які незрозумілі особливості.

Продуктивність багатокласових моделей за результатами експерименту E1 наведено в таблиці 4.2. Моделі LR і linear SVC показали найкращі результати – обидві моделі отримали результат 90% за всіма чотирма ключовими показниками ефективності. Обидві ці моделі загалом працюють добре, причому жодна з них не віддає перевагу певному класу для досягнення високих результатів.

Таблиця 4.2 – Продуктивність мультикласової моделі для E1

Модель	Точність (ACC)	Точність (PPV)	Повнота (TPR)	Оцінка F1
Rand Baseline	29	29	29	29
Mode Baseline	33	11	33	17
LR	90	90	90	90
KNN	86	86	86	86
SVC	89	89	89	89
Linear SVC	90	90	90	90

З моделей досліджувалася окремо, щоб визначити, які характеристики важливі для кожного класу даних. Це було зроблено для визначення найбільш відмінних рис в моделях у спробі дізнатися більше про те, що

відрізняє справжні статті від фейкових і сатиричних статей. Для цього процесу були побудовані графіки коефіцієнтів моделі для кожної моделі в кожному з під експериментів, приклад яких наведено на рисунку 4.7.

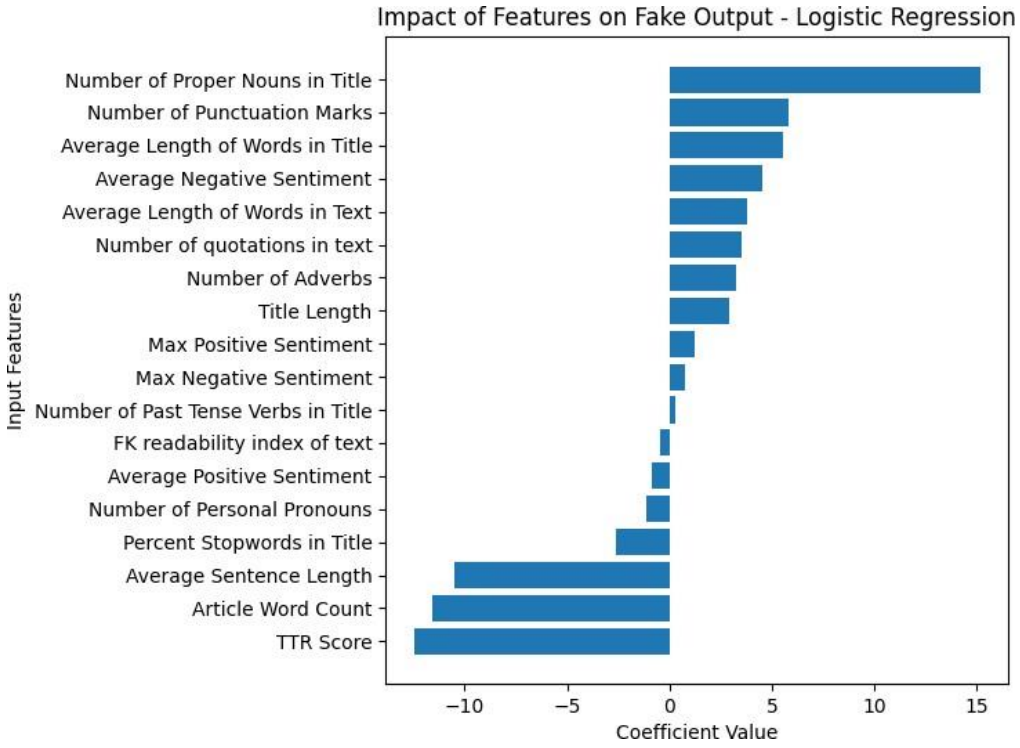

Рисунок 4.7 – Графіки коефіцієнтів для моделі багатокласової логістичної регресії (експеримент E1)

Показники ефективності реальних і підроблених моделей наведено на таблиці 4.3.

Найбільш ефективними моделями є моделі LR і лінійні моделі SVC, які досягли майже ідеального результату (98% за всіма ключовими показниками ефективності). Обидві ці моделі загалом працюють добре, причому жодна з них не віддає перевагу певному класу для досягнення високих результатів.

Таблиця 4.3 – Ефективність справжньої та підробленої моделей в E1

Модель	Точність (ACC)	Точність (PPV)	Повнота (TPR)	Оцінка F1
Rand Baseline	51	51	51	51
Mode Baseline	51	26	50	34
LR	98	98	98	98
KNN	96	96	96	96
SVC	97	97	97	97
Linear SVC	98	98	98	98

Показники ефективності моделей фейку та сатири наведено на таблиці 4.4. Моделі логістичної регресії, SVC і лінійної SVC показали найкращі результати (93% за всіма ключовими показниками ефективності). Кожна з цих моделей загалом працює добре, жодна з моделей не віддає перевагу певному класу для досягнення своїх високих результатів.

Таблиця 4.4 – Ефективність фейкової та сатиричної моделей в E1

Модель	Точність (ACC)	Точність (PPV)	Повнота (TPR)	Оцінка F1
Rand Baseline	53	53	53	53
Mode Baseline	51	25	50	34
LR	93	93	93	93
KNN	89	89	89	89
SVC	93	93	93	93
Linear SVC	93	93	93	93

Показники ефективності реальних і сатиричних моделей наведено в таблиці 4.5. Найкращими моделями є моделі логістичної регресії, SVC і лінійні моделі SVC, які досягли результату 97% за всіма ключовими показниками ефективності. Кожна з цих моделей загалом працює добре, жодна з моделей не віддає перевагу певному класу для досягнення високих результатів.

Таблиця 4.5 – Ефективність реальної та сатиричної моделей в E1

Модель	Точність (ACC)	Точність (PPV)	Повнота (TPR)	Оцінка F1
Rand Baseline	50	50	50	50
Mode Baseline	51	25	50	33
LR	97	97	97	97
KNN	94	94	94	94
SVC	97	97	97	97
Linear SVC	97	97	97	97

Показники ефективності реальних моделей порівняно з усіма моделями наведено в таблиці 4.6. Найкращими моделями є моделі логістичної регресії, SVC і лінійні моделі SVC, які досягли результату 95% за всіма 4 ключовими показниками ефективності. Кожна з цих моделей загалом працює добре, жодна з моделей не віддає перевагу певному класу для досягнення своїх високих результатів.

Таблиця 4.6 – Ефективність реальної та всіх інших моделей в E1

Модель	Точність (ACC)	Точність (PPV)	Повнота (TPR)	Оцінка F1
Rand Baseline	49	49	49	49
Mode Baseline	50	25	50	34
LR	95	95	95	95
KNN	93	93	93	93
SVC	95	95	95	95
Linear SVC	95	95	95	95

З цих результатів випливає, що в контексті набору даних 1 можна створити точні моделі для розрізнення кожного класу новин. Найскладніше було розрізнити сатиру та фейкові новини, причому моделі під час експерименту «фейк проти сатири» показали найгірші результати. Усі моделі досягли результатів 90% або більше за кожним з показників і в результаті виникла деяка стурбованість щодо надмірного оснащення, тому це було ретельно досліджено. Однак завдяки подібним результатам, досягнутим як на даних навчання так і на тестуванні, модель явно вивчає риси набору даних у цілому, а не лише дані, на яких вона навчалася.

Обмеження щодо різноманітності статей у наборі даних можуть створити додаткову проблему.

У цьому розділі будуть розглянуті результати дослідження відмінних рис для кожного класу. Такі графіки, як на рисунках 4.8, 4.9 і 4.10 були створені для моделей логістичної регресії, SVC і лінійної SVC, як для багатокласових, так і для півекспериментів «реальна проти всіх». Це дослідження не проводилося для моделей KNN, оскільки вони не підлягають поясненню і тому неможливо визначити важливість функції для цієї моделі.

Для будь-якого даного класу в задачі класифікації існують позитивні та негативні показники, найбільші з яких вважаються відмінними ознаками класу. Позитивні показники вказують на те, що стаття, ймовірно, належить до класу, а негативні вказують на те, що стаття, ймовірно, не належить до класу.

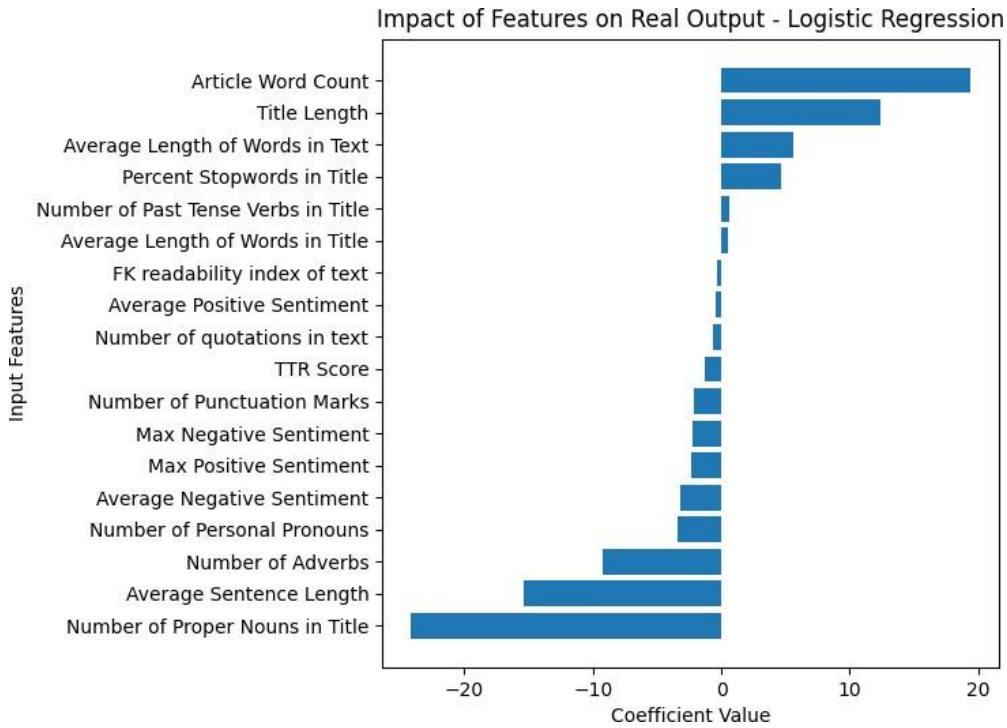

Рисунок 4.8 – Важливість функції для багатокласової моделі LR (E1)

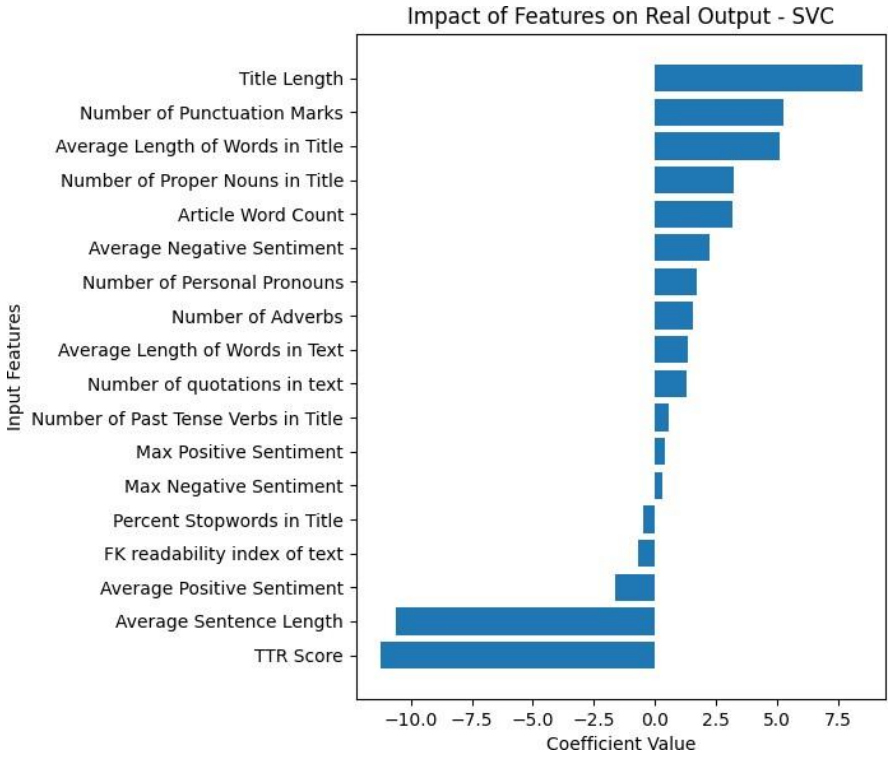

Рисунок 4.9 – Важливість функції для багатокласової моделі SVC (E1)

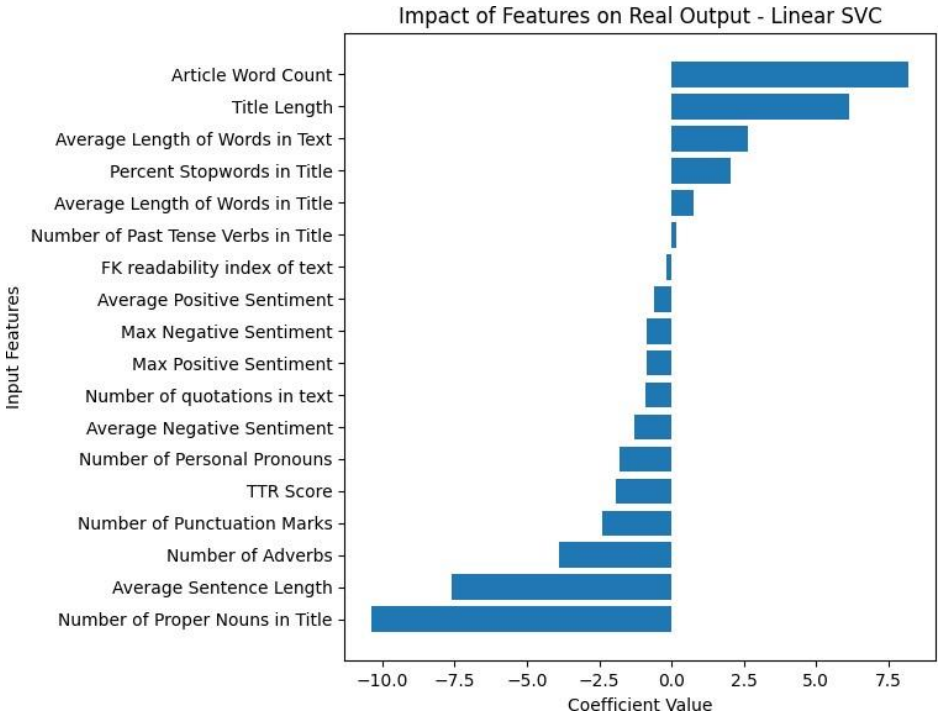

Рисунок 4.10 – Важливість функції для багатокласової лінійної моделі SVC (E1)

Більшість відмінних рис, визначених для кожного класу, є передбачуваними, однак виявлені також і деякі цікаві особливості, які не очікувалися (наприклад, той факт, що статті фейкових новин використовують більше слів у заголовках). Результати для моделі «реальне проти всіх» показують, що позитивні показники для іншого класу такі ж, як позитивні показники для класу сатири в багатокласовому під експерименті, тоді як негативні показники для іншого класу є такі ж, як і негативні показники для підробленого класу в багатокласовому під експерименті. Це свідчить про те, що фейкові новини більше схожі на новини сатири, ніж на справжні новини, підкріплюючи ідею про те, що визначення різниці між фейковими та сатиричними новинами є одним із головних завдань у кваліфікаційній роботі.

Існує обмеження в наборі даних 1: як реальні, так і сатиричні статті беруться з одного джерела, що робить можливим те, що ці моделі брали до уваги особливості публікацій, а не особливості реальних і сатиричних новин загалом. Це буде досліджено далі під час аналізу характеристик моделей, розроблених із набором даних 2.

4.2 Другий експеримент

Метою експерименту E2 є створення набору зрозумілих і остаточно пояснюваних моделей. Це має бути досягнуто шляхом навчання моделей з використанням набору функцій 2. Усі моделі, розглянуті в цьому підрозділі, навчаються та перевіряються на наборі даних 1.

Знову набір даних сегментується на дві частини у співвідношенні 80%/20% (80% використовуються для процесу перехресної перевірки, а 20% – для тестування моделей після вибору оптимальних гіперпараметрів.

Ті самі загальні параметри моделі, що описані в E1, були знову реалізовані для E2.

В цьому експерименті використовуються ті ж самі моделі, тому гіперпараметри та процес їх вибору такі самі, як описано у реалізації E1 вище. Щоб уникнути зайвих операцій, процес відбору не буде повторюватися, а замість цього виберемо гіперпараметри для цього експерименту, значення яких наведено в таблиці 4.7.

Таблиця 4.7 – Значення гіперпараметрів, знайдених шляхом перехресної перевірки для E2

Model	Multi-class	Fake vs Real	Fake vs Satire	Real vs Satire	Real vs All
LR	100	10	10	100	10
KNN	100	100	100	100	100
SVC	1	1	1	10	1
Linear SVC	10	10	10	10	10

Після вибору оптимальних гіперпараметрів для кожної моделі в кожній під проблеми E2 моделі знову навчалися за допомогою вибраних гіперпараметрів на 80% даних збережених для навчання. Потім кожна модель була протестована на 20% даних.

Показники продуктивності багатокласових моделей наведено в таблиці 4.8.

Найкращою моделлю є модель логістичної регресії, яка досягла результату 89% за всіма чотрима ключовими показниками ефективності. Ця модель загалом працює добре, однак у неї виникають певні проблеми з розрізненням реальних і сатиричних статей, що видно з матриці плутанини де більшість неправильних класифікацій відбувається через те, що справжні новинні статті класифікуються як фейкові і навпаки.

Таблиця 4.8 – Багатокласова продуктивність моделі для E2

Модель	Точність (ACC)	Точність (PPV)	Повнота (TPR)	Оцінка F1
Rand Baseline	33	33	33	33
Mode Baseline	33	11	33	17
LR	89	89	89	89
KNN	85	85	85	85
SVC	88	88	88	88
Linear SVC	88	88	88	88

Показники продуктивності реальних і фейкових моделей можна побачити в таблиці 4.9. Найкращою моделлю є модель Linear SVC, яка отримала результат 98% за всіма ключовими показниками ефективності.

Ця модель в цілому добре працює, не віддаючи переваги жодному класу під час прогнозування.

Таблиця 4.9 – Фейкова та справжня продуктивність моделі для E2

Модель	Точність (ACC)	Точність (PPV)	Повнота (TPR)	Оцінка F1
Rand Baseline	51	51	51	51
Mode Baseline	51	26	50	34
LR	97	97	97	97
KNN	96	96	96	96
SVC	97	97	97	97
Linear SVC	98	98	98	98

Показники продуктивності моделей фейку та сатири можна побачити в таблиці 4.10. Найкращою моделлю є лінійна модель SVC, яка досягла результату 92% за всіма 4 ключовими показниками ефективності.

Ця модель в цілому працює добре, не віддаючи переваги жодному класу під час прогнозування.

Таблиця 4.10 – Ефективність моделей фейку та сатири в E2

Модель	Точність (ACC)	Точність (PPV)	Повнота (TPR)	Оцінка F1
Rand Baseline	50	50	50	50
Mode Baseline	51	25	50	34
LR	91	91	91	91
KNN	88	88	88	88
SVC	91	91	91	91
Linear SVC	92	92	92	92

Показники ефективності реальних і сатиричних моделей можна побачити в таблиці 4.11. Найкращими моделями є моделі логістичної регресії, SVC і лінійні моделі SVC, які досягли результату 96% за всіма ключовими показниками ефективності. Кожна з цих моделей загалом працює добре, жодна з моделей не віддає перевагу певному класу для досягнення своїх високих результатів.

Таблиця 4.11 – Ефективність реальної та сатиричної моделі для E2

Модель	Точність (ACC)	Точність (PPV)	Повнота (TPR)	Оцінка F1
Rand Baseline	51	51	51	51
Mode Baseline	51	25	50	33
LR	96	96	96	96
KNN	94	94	94	94
SVC	96	96	96	96
Linear SVC	96	96	96	96

Показники ефективності реальних моделей порівняно з усіма моделями наведено в таблиці 4.12. Найкращою моделлю є лінійна модель SVC, яка досягла результату 96% за всіма ключовими показниками ефективності. Ця модель загалом працює добре, не надаючи переваги жодному класу під час прогнозування.

Таблиця 4.12 – Реальна продуктивність порівняно з усіма моделями в E2

Модель	Точність (ACC)	Точність (PPV)	Повнота (TPR)	Оцінка F1
Rand Baseline	51	51	51	51
Mode Baseline	50	25	50	34
LR	95	95	95	95
KNN	93	93	93	93
SVC	94	94	94	94
Linear SVC	96	96	96	96

З цих результатів випливає, що в контексті набору даних 1 можна створити точні моделі для розрізнення кожного класу новин за допомогою набору функцій 2. Найважче розрізнити сатиру та фейкові новини, де моделі мали найнижчу точність. У під експерименті 5 досліджувалася здатність моделей відрізнити справжні статті від класу загальної дезінформації (поєднання сатири та фейкових статей). Ці моделі досягли таких самих балів, як і найкращі моделі під експериментів 2 і 4, де моделі намагалися відокремити справжні статті від фейкових, а також справжні статті від сатиричних. Однак багатокласова модель все ще відставала від усіх бінарних класифікаторів з точки зору продуктивності, що вказує на те, що спроба відокремити різні типи дезінформації може знизити продуктивність моделі. Коли моделі досягли понад 90% за кожним із показників, надмірне оснащення викликало занепокоєння, але це виявилося не так з тих самих причин, які були зазначені під час загального обговорення результатів моделі E1.

4.2 Третій експеримент

Мета експерименту E3 – визначити, наскільки добре моделі, що були згенеровані в E2, узагальнені до проблеми розрізнення справжніх, фейкових і сатиричних новин. Це дуже важко виміряти в рамках одного набору даних, тому для цього експерименту оптимізовані моделі, навчені в E2 з

використанням набору даних 1, були перевірені на наборі даних 2. Це дає можливість визначити, наскільки добре моделі підігнані до загальної проблеми виявлення дезінформації.

Моделі, навчені для кожного з п’яти під експериментів, беруться з E2 і перевіряються на 100% даних з набору даних 2. Зазвичай набір даних сегментується на дані для навчання та тестування, однак, оскільки навчання проводилося на повністю окремому наборі даних, весь набір даних 2 можна використовувати як набір для тестування. Ті самі загальні параметри моделі, що описані в E2, були знову реалізовані для E3. Кожна з моделей була оцінена за чотирма ключовими показниками, розглянутими в таблиці 2.1.

Оцінки продуктивності багатокласових моделей наведено в таблиці 4.13. Найкращою моделлю є модель SVC, яка отримала найвищий бал за трьома з чотирьох ключових показників ефективності. Однак ця модель все ще працює погано з точністю лише на 20% кращою, ніж випадкова модель. Модель SVC також досягла такої точності завдяки практичному ігноруванню класу даних сатири, прогнозуючи сатиру лише 7 разів у 225 статтях. Усі багатокласові моделі в цьому експерименті стикаються з цією проблемою, як показано в таблиці 4.13. Це підкреслює відмінності між набором даних 1 і набором даних 2, особливо коли йдеться про сатиричний клас даних.

Таблиця 4.13 – Продуктивність багатокласової моделі в E3

Модель	Точність (ACC)	Точність (PPV)	Повнота (TPR)	Оцінка F1
Rand Baseline	37	37	37	37
Mode Baseline	33	11	33	17
LR	48	65	48	38
KNN	52	53	52	50
SVC	54	67	54	47
Linear SVC	45	64	45	35

Оцінки ефективності реальних і фейкових моделей наведено в таблиці 4.14. Найкращою моделлю є модель SVC, яка отримала найвищий

бал за всіма ключовими показниками ефективності. Модель SVC трохи надає перевагу фейковому класу, приблизно 64% прогнозів моделі стосуються фейкових новин. Цю тенденцію поділяють усі інші моделі, крім моделі Linear SVC, яка значною мірою надає перевагу реальному класу.

Таблиця 4.14 – Фейкова та справжня продуктивності моделі для E3

Модель	Точність (ACC)	Точність (PPV)	Повнота (TPR)	Оцінка F1
Rand Baseline	47	47	47	47
Mode Baseline	50	25	50	33
LR	78	78	78	78
KNN	75	75	75	75
SVC	78	80	78	78
Linear SVC	68	75	68	66

Оцінки ефективності моделей фейку та сатири наведено в таблиці 4.15. Найкращою моделлю є модель KNN, яка є єдиною моделлю, яка має кращий результат ніж випадкова модель за всіма ключовими показниками ефективності. Однак, оскільки це не пояснювана модель, її не можна використовувати для цього проекту, підкреслюючи деякі обмеження, які створює розробка пояснюваних моделей.

Усі моделі в цьому під експерименті віддають перевагу класу фейків і майже ніколи не передбачають клас сатири, що видно з таблиці 4.15. Це ще раз підкреслює потенційні відмінності між набором даних 1 і набором даних 2, оскільки моделі, навчені на наборі даних 1, явно мало або зовсім не використовують інформацію про різницю між фейковими та сатиричними статтями в наборі даних 2.

Таблиця 4.15 – Ефективність моделі фейку та сатири в E3

Модель	Точність (ACC)	Точність (PPV)	Повнота (TPR)	Оцінка F1
Rand Baseline	53	53	53	53
Mode Baseline	50	25	50	33
LR	52	66	52	39
KNN	59	60	59	58
SVC	53	63	53	41
Linear SVC	51	75	51	35

Ефективність реальних і сатиричних моделей можна побачити в таблиці 4.16. Найкращою моделлю є модель KNN, яка набрала 74% за всіма ключовими показниками ефективності. Знову ж таки, оскільки це не піддається поясненню, його не можна використовувати, тому наступною найкращою моделлю є модель SVC. Модель SVC працює досить добре, хоча вона надає перевагу класу реальних даних (приблизно 76% прогнозів стосуються класу реальних даних). Це тенденція для всіх моделей у цьому під експерименті з моделями логістичної регресії та лінійними моделями SVC, які майже ніколи не прогнозують. Це ще раз підкреслює, що позитивні показники для сатиричних статей можуть бути відносно слабшими або потенційно іншими в наборі даних 2.

Таблиця 4.16 – Ефективність реальної та сатиричної моделей в E3

Модель	Точність (ACC)	Точність (PPV)	Повнота (TPR)	Оцінка F1
Rand Baseline	49	49	49	49
Mode Baseline	50	25	50	33
LR	55	76	55	43
KNN	74	74	74	74
SVC	73	81	73	71
Linear SVC	51	75	51	36

Показники продуктивності реальних моделей порівняно з усіма моделями наведено в таблиці 4.17. Найкращою моделлю є модель логістичної регресії, яка отримала найвищий бал за всіма ключовими

показниками ефективності. Ця модель загалом працює добре, не віддаючи переваги жодному класу під час прогнозування.

Таблиця 4.17 – Реальна продуктивність порівняно з усіма моделями в E3

Модель	Точність (ACC)	Точність (PPV)	Повнота (TPR)	Оцінка F1
Rand Baseline	55	55	55	55
Mode Baseline	50	25	50	33
LR	77	78	77	77
KNN	74	74	74	74
SVC	74	75	74	74
Linear SVC	66	78	66	62

Моделі в цьому експерименті показали значно гірші результати ніж ті, що були навчені в E1 і E2. Це певною мірою очікувалося, оскільки тестування більшості моделей на новому наборі даних призведе до деякого зниження продуктивності. Однак дослідження зниження продуктивності моделей, створених для під експериментів з кількома класами та під експериментами «фейк проти сатири», виявило слабкість цих моделей щодо визначення сатиричних новин. Це вказує на те, що сатиричні новинні статті в наборах даних 1 і 2 більше відрізняються від інших класів статей.

Усі моделі з інших під експериментів також зазнали значного зниження продуктивності (від 15 до 20%), що вказує на наявність суттєвих відмінностей між даними в наборі даних 1 і наборі даних 2. Однак режими для інших під експериментів все ще працювали достатньо добре (найкращі моделі для кожного під експерименту мали точність від 70% до 80%). Це вказує на те, що моделі, що навчені на наборі даних 1, мають непогану продуктивність і здатні працювати зі зниженою потужністю на абсолютно новому наборі даних. Це свідчить про те, що вони можуть бути залучені до вирішення загальної проблеми дезінформації.

4.2 Четвертий експеримент

Мета експерименту E4 – визначити, наскільки ефективним є набір ознак 2 у вирішенні загальної проблеми класифікації дезінформації. Це важливо, оскільки багато вибраних функцій було вибрано порівняно з іншими такими ж дійсними функціями через їх ефективність у наборі даних 1. Шляхом навчання та тестування моделей на наборі даних 2 з використанням набору функцій 2 можна дослідити ефективність цих функцій на іншому наборі даних, що допоможе визначити, чи функції ефективні лише на наборі даних 1, або вони можуть бути застосовні до загальної проблема виявлення дезінформації.

Знову набір даних сегментувався на 2 частини у співвідношенні 80%/20% (80% використовуються для процесу перехресної перевірки, а 20% залишаються осторонь для тестування моделей після вибору оптимальних гіперпараметрів).

При цьому кілька загальних параметрів, які використовувалися для моделей в E4, відрізняються від параметрів у попередніх експериментах (зокрема, кількість сусідів), що розглядаються для моделі KNN, становить лише 90.

Це є зменшенням порівняно з 2000 сусідів в інших трьох експериментах, що є результатом меншого розміру набору даних 2). Інші загальні параметри залишаються незмінними. Діапазони значень гіперпараметрів і метрики перехресної перевірки вибираються з використанням тих самих критеріїв, які обговорювалися вище при реалізації E1. Оскільки в розглянутих експериментах використовуються однакові моделі, то гіперпараметри та процес їх вибору залишаються такими ж, як у реалізації E1. Значення вибраних гіперпараметрів для цього експерименту наведено в таблиці 4.18.

Таблиця 4.18 – Значення гіперпараметрів, визначених шляхом перехресної перевірки

Model	Multi-class	Fake vs Real	Fake vs Satire	Real vs Satire	Real vs All
LR	100	100	100	100	100
KNN	10	5	25	25	10
SVC	10	1	10	10	10
Linear SVC	100	10	100	10	10

Після того, як були обрані оптимальні гіперпараметри для кожної моделі в кожній підпроблемі E4, моделі були ще раз навчені за допомогою цих гіперпараметрів на 80% даних, збережених для навчання. Потім кожну модель перевіряли на 20% даних, а кожна з моделей була оцінена за чотирма ключовими показниками, розглянутими в таблиці 2.1.

Моделі були досліджені, щоб визначити, які характеристики важливі для кожного класу. Важливі характеристики моделей, навчених на наборі даних 1, потім порівнювали з важливими характеристиками моделей, навчених на наборі даних 2, щоб побачити, наскільки узгодженими є відмінні риси в наборах даних.

Нарешті, результати класифікації моделей, створених в E4, порівнювалися безпосередньо з результатами для набору даних 2, наведеними в роботі [8]. Субекспериментi «фейк проти реальності», «фейк проти сатири» та «реальна проти сатири» є важливими для цього, оскільки вони дозволяють пряме порівняння результатів. Це справедливе порівняння, оскільки обидва набори моделей навчаються та тестуються на одному наборі даних і використовують подібні методи (такі як перехресна перевірка для моделей навчання та вибір гіперпараметрів). Основна відмінність у підходах полягає у використовуваних наборах функцій. Їхня робота може мати невелику перевагу в цій області, оскільки вони оптимізували свій вибір функцій для набору даних, який використовується для порівняння, тоді як в кваліфікаційній роботі функції були вибрані та оптимізовані для набору даних 1.

Оцінки продуктивніості багатокласових моделей наведено в таблиці 4.19. Найкращою моделлю є лінійна модель SVC, яка отримала найвищий бал за всіма ключовими показниками ефективності. Ця модель добре розрізняє фейкові та справжні новинні статті, але має проблеми з виявленням сатиричних статей. Лінійна модель SVC правильно класифікувала лише близько половини сатиричних статей, а усі моделі в цьому під експерименті мають схожі проблеми з відрізненням сатиричних новин від інших класів новин.

Таблиця 4.19 – Показники продуктивності багатокласової моделі для експерименту E4

Модель	Точність (ACC)	Точність (PPV)	Повнота (TPR)	Оцінка F1
Rand Baseline	35	35	35	35
Mode Baseline	35	33	33	17
LR	64	64	65	64
KNN	47	46	47	47
SVC	47	48	47	48
Linear SVC	71	70	72	70

Оцінки ефективності реальних і підроблених моделей наведено в таблиці 4.20. Найкращими моделями є логістична регресія та лінійна модель SVC, обидві з яких набрали 87% за всіма ключовими показниками ефективності. Ці моделі не віддають перевагу жодному класу (такої тенденції дотримуються всі моделі в цій підпроблемі). Це вказує на те, що створення моделей для відмінності фейкових новин від справжніх у наборі даних 2 є відносно простим за допомогою набору функцій 2.

Таблиця 4.20 – Фейкова та справжня продуктивність моделі для E4

Модель	Точність (ACC)	Точність (PPV)	Повнота (TPR)	Оцінка F1
Rand Baseline	50	50	50	50
Mode Baseline	50	25	50	33
LR	87	87	87	87
KNN	80	80	80	80
SVC	87	87	87	87
Linear SVC	87	87	87	87

Оцінки ефективності моделей фейку та сатири наведено в таблиці 4.21. Найкращою моделлю є модель SVC, що має дуже подібні показники до моделей логістичної регресії та лінійної моделі SVC за всіма ключовими показниками ефективності, однак її можна вважати більш збалансованим класифікатором.

Класифікатор SVC досягає оцінок лише 63% за кожним ключовим показником ефективності, що лише трохи краще, ніж випадковий класифікатор. Це ще раз підкреслює складність розрізнення фейкових і сатиричних новин і пояснює, чому багатокласові моделі надають перевагу справжнім і фейковим класам.

Таблиця 4.21 – Ефективність фейкової та сатиричної моделі для E4

Модель	Точність (ACC)	Точність (PPV)	Повнота (TPR)	Оцінка F1
Rand Baseline	40	40	40	40
Mode Baseline	50	25	50	33
LR	63	64	63	63
KNN	60	60	60	60
SVC	63	63	63	63
Linear SVC	63	64	63	63

Ефективність реальних і сатиричних моделей можна побачити в таблиці 4.22. Найкращими моделями є логістична регресія та лінійні моделі SVC, які досягли спільного найвищого результату за всіма ключовими показниками ефективності. Обидві ці моделі віддають перевагу класу

реальних даних, причому трохи менше 66% прогнозів стосуються реального класу та цієї тенденції дотримуються всі моделі в цьому під експерименті.

Таблиця 4.22 – Ефективність реальної та сатиричної моделі для E4

Модель	Точність (ACC)	Точність (PPV)	Повнота (TPR)	Оцінка F1
Rand Baseline	50	50	50	49
Mode Baseline	50	25	50	33
LR	80	82	80	80
KNN	77	80	77	76
SVC	77	78	77	76
Linear SVC	80	82	80	80

Оцінку ефективності реальних моделей порівняно з усіма моделями наведено в таблиці 4.23. Найкращою моделлю є модель SVC, яка набрала 86% за всіма ключовими показниками ефективності. Ця модель загалом добре працює, не віддаючи переваги жодному класу під час прогнозування (це тенденція, якої дотримуються інші моделі в цьому під експерименті). Це вказує на те, що створення моделей для відмінності підроблених статей від справжніх і сатиричних статей у наборі даних 2 відносно легко можна здійснити за допомогою набору функцій 2.

Таблиця 4.23 – Оцінка ефективності реальних моделей порівняно з усіма моделями для E4

Модель	Точність (ACC)	Точність (PPV)	Повнота (TPR)	Оцінка F1
Rand Baseline	45	47	47	45
Mode Baseline	58	29	50	36
LR	84	84	83	83
KNN	81	81	79	80
SVC	87	87	87	87
Linear SVC	84	84	83	83

Моделі в цьому експерименті показали кращі результати, ніж моделі, навчені в E3, що було до певної міри очікуваним, оскільки моделі, навчені та протестовані з використанням даних з того самого набору даних, мають

тенденцію працювати краще. Знову головна складність для цих моделей полягає в розрізненні фейкових і сатиричних статей. Моделі під експерименту «фейк проти сатири» показали значно гірші результати, ніж усі інші бінарні класифікатори, і навіть гірші, ніж багатокласові моделі. Це вказує на те, що набір функцій 2 не є ефективним для всіх типів сатиричних новин і щоб створити ефективну загальну модель, може бути потрібним ввести більше функцій, які відрізнятимуть сатиричні статті від справжніх і фальшивих статей. Крім того, можливо, що характеристики сатиричних новин можуть бути більш різноманітними і тому може знадобитися більший набір даних для навчання ефективних моделей за допомогою набору функцій 2.

Найкращі моделі, навчені згідно з під експериментам «справжнє проти всіх», «справжнє проти сатири» та «справжнє проти фейку» показали лише в середньому на 5% кращі результати, ніж моделі, які були навчені на наборі даних 1 із використанням набору функцій 2. Це ще раз підкреслює працездатність моделей, навчених в процесі експерименту E2, та їх здатність вирішувати загальну проблему фейкових новин, а не лише досягати високої продуктивності на даних, на яких вони навчалися. Загалом навчання високоефективних моделей на наборі даних 2 виявилося значно складнішим, ніж навчання моделей для набору даних 1. Однією з головних причин цього, ймовірно, є розмір набору даних, оскільки навчання ефективних моделей із багатьма функціями часто є складним на менших наборах даних.

На рисунках 4.11, 4.12 і 4.13 наведено графіки, що були створені для моделей логістичної регресії, SVC і лінійної SVC, як для багатокласових, так і для під експериментів «реальне проти всіх». Цей процес не проводився для моделей KNN, оскільки вони не піддаються поясненню, через що неможливо визначити важливість функції для цієї моделі.

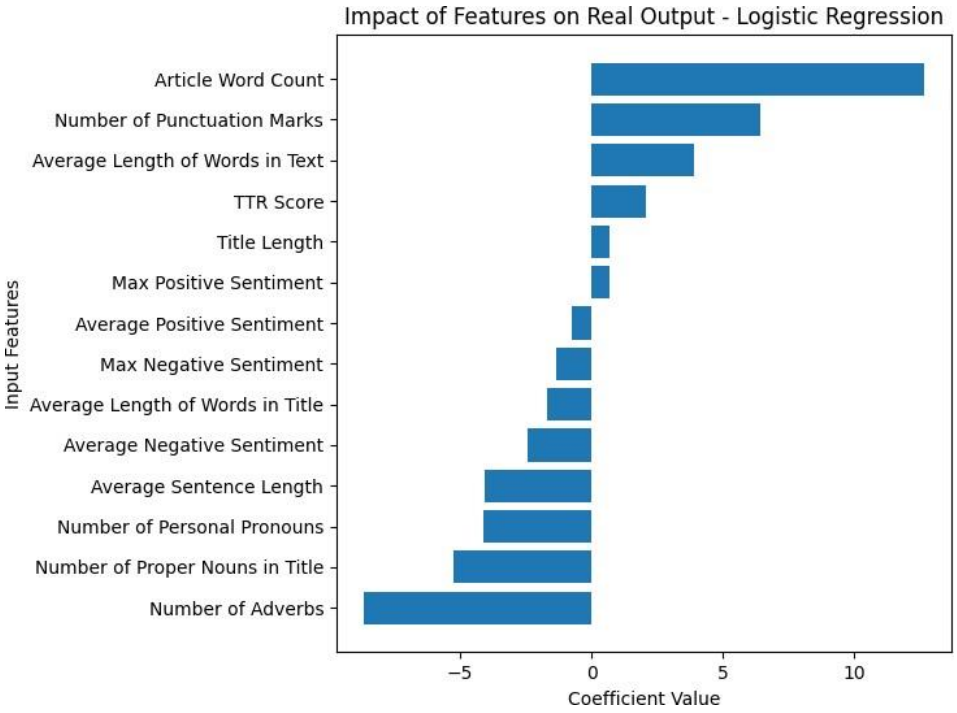

Рисунок 4.11 – Важливість функції для реального класу (багатокласова модель LR для E4)

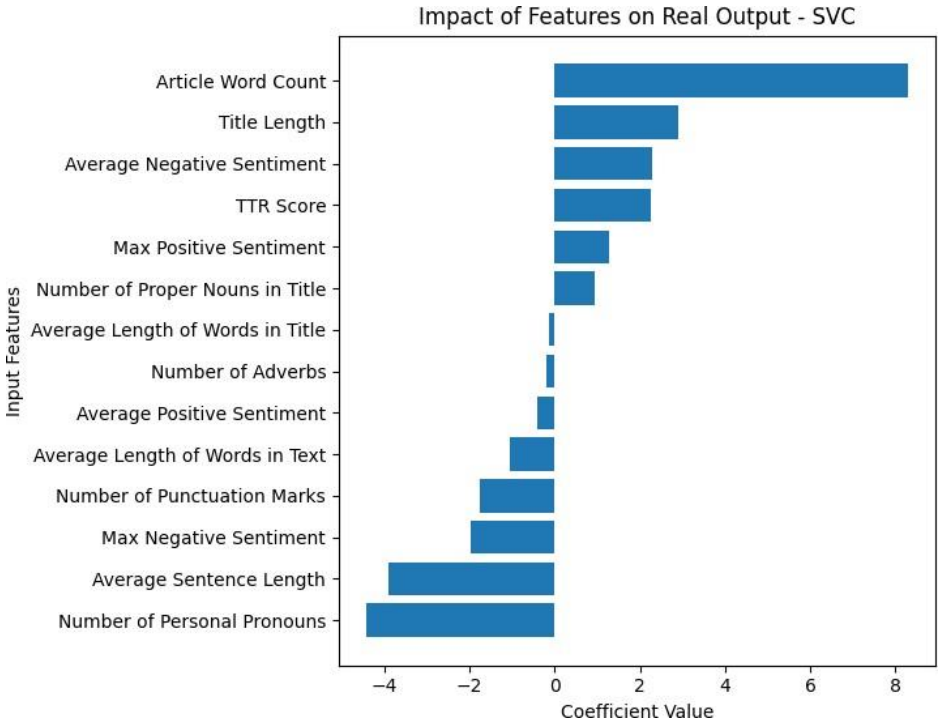

Рисунок 4.12 – Важливість функції для реального класу (модель SVC для E4)

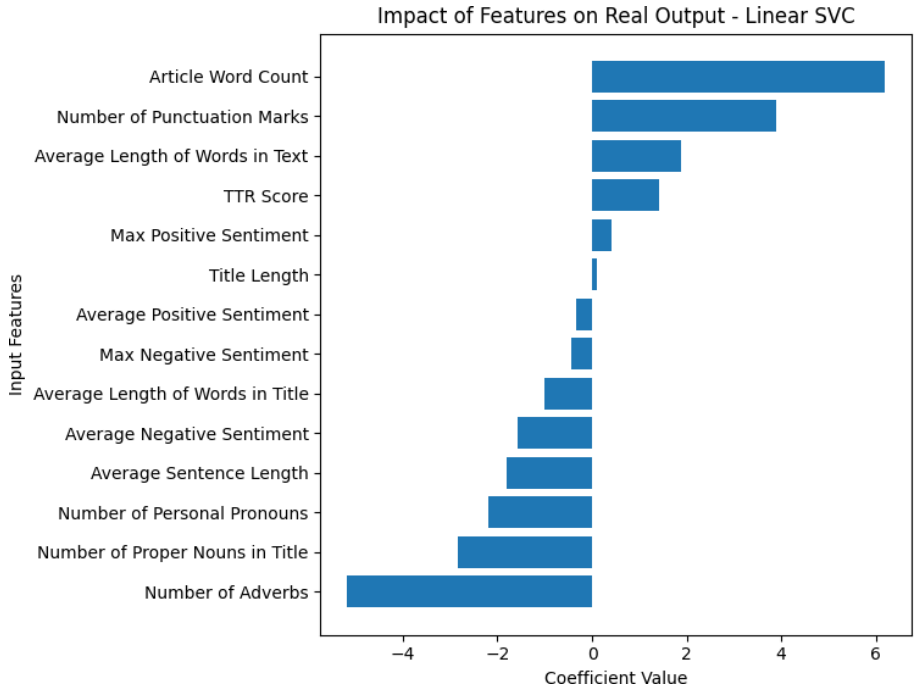

Рисунок 4.13 – Важливість функції для реального класу (багатокласова лінійна модель SVC для E4)

Більшість відмінних рис, визначених для кожного класу, є передбачуваними за винятком деяких неочікуваних особливостей. Наприклад, у сатиричних новинах використовуються довші слова, що свідчить про складнішу мову. Більшість показників для реального класу в під експерименті «Реальний проти всіх» такі ж, як і для реального класу в під експерименті з кількома класами, однак функція «Середній негативний настрій» змінилася з позитивного показника у під експерименті з кількома класами до негативного в експерименті «Реальний проти всіх». Це ще раз вказує на додаткову складність вирішення багатокласової проблеми, оскільки аналіз функцій багатокласової моделі не обов'язково показує найбільш помітні ознаки кожного класу, натомість показує найбільш відмінні риси між ними.

Існує обмеження щодо набору даних 2, адже набір даних містить лише 225 статей (75 на клас). Це є недостатнім для навчання складних моделей з великою кількістю функцій і з цієї причини може бути певний

ступінь переобладнання в моделях і в аналізі функцій. Однак багато показників були такими ж, як ті, що використовувалися для набору даних 1, що є багатообіцяючим з точки зору переобладнання.

4.5 Результати експериментів

Експеримент E1 було проведено для створення набору базових моделей за допомогою набору функцій 1, щоб визначити верхню межу продуктивності моделей. Цей експеримент був успішним, оскільки він дозволив створити моделі, які показали хороші результати в кожному з п’яти під експериментів. Ці моделі не піддаються поясненню, оскільки набір функцій 1 містить деякі функції, незрозумілі для пересічного споживача новин.

Експеримент E2 було проведено для створення набору моделей з використанням набору функцій 2, щоб спостерігати за падінням продуктивності після видалення деяких потенційно проблемних функцій. Цей експеримент знову виявився успішним, оскільки, незважаючи на зменшення розміру набору функцій з 18 до 14 функцій, продуктивність моделі майже не впала, причому моделі з найкращими показниками в кожному під експерименті E2 втратили приблизно 1% у кожному показнику оцінки відносно відповідних показників в E1. Це зниження продуктивності є відносно невеликим і прийнятним, оскільки дозволяє створювати моделі, навчені лише зрозумілими функціями.

Експеримент E3 був реалізований, щоб перевірити, наскільки ефективними були моделі, згенеровані в E2 під час тестування на новому наборі даних. Це мало показати, наскільки модель узагальнена для загальної проблеми або наскільки модель переобладнана для конкретного набору даних. Цей експеримент показав, що деякі моделі з E2 сильно переобладнують дані з набору даних 1 і в результаті під час їх тестування на

наборі даних 2 вони зазнали серйозного зниження продуктивності що показано на таблиці 4.24.

Таблиця 4.24 – Порівняння продуктивності моделей E2 і E3

Під експеримент	Точність (ACC)		Точність (PPV)		Повнота (TPR)		Оцінка F1	
	E2	E3	E2	E3	E2	E3	E2	E3
Multi-class	89	54	89	67	89	54	89	47
Real vs Fake	98	78	98	80	98	78	98	78
Fake vs Satire	92	53	92	63	92	53	92	41
Real vs Satire	96	73	96	81	96	73	96	71
Real vs All	96	77	96	78	96	77	96	77

Моделі під експериментів «справжнє проти фейку», «справжнє проти сатири» та «реальне проти всіх» достатньо добре працюють, причому найкращі моделі для цих трьох під експериментів мають показники точності понад 70%. Втім, це є значним зниженням (в середньому приблизно на 20% у кожному показнику) в порівнянні з бінарними моделями, точність яких в E2 досягла 90%. Це вказує на те, що підроблені та справжні дані в наборі даних 1 і наборі даних 2 мають багато подібностей. Модель фейку та сатири виконана майже випадково, що вказує на те, що було дуже мало подібностей між сатиричними новинними статтями в наборах даних 1 і 2. Це сильно вплинуло на багатокласові моделі, оскільки вони не змогли ефективно ідентифікувати сатиричні новини, що призвело до суттєвого скорочення випадків вдалої класифікації. Проблема дуже різних сатиричних статей у наборах даних 1 і 2, ймовірно, є результатом обмежень набору даних 1.

Загалом, виключаючи клас сатири, можна побачити, що моделі в E2 мають цінність у більш узагальненому середовищі фейкових новин. Зокрема, модель «реальний проти всіх» здатна відрізнити клас справжніх новин від інших трьох класів з точністю 77 % на новому наборі даних без суттєвої переваги будь-якого класу.

Експеримент E4 було проведено, щоб визначити, чи набір функцій 2, який було вибрано та оптимізовано для набору даних 1, має цінність під час

створення моделей на нових наборах даних. В Е4 моделі створювалися з використанням набору даних 2 і набору функцій 2. Ці моделі показують покращення в порівнянні з найкращими моделями з Е3, причому найбільші покращення спостерігаються в під експериментах із кількома класами та «фейк проти сатири» (покращення складає в середньому приблизно 15% за один показник). Інші три під експерименти в Е4 також показали покращення приблизно на 5% на один показник в порівнянні з аналогами Е3. Це слід було очікувати, адже моделі, створені в Е4, тренувалися на даних, більш схожих на набір даних тестування, і через це досягли вищих балів.

Моделі в Е4 не досягли такого ж рівня продуктивності, як моделі в Е1 і Е2, які також були навчені та перевірені на даних із того самого набору даних. Для цього існує кілька потенційних причин. Перша з них – розмір набору даних 2 обмежений лише 225 статтями, тобто лише 75 прикладами з кожного класу. Створення складних моделей із відносно великим набором функцій часто потребує великих наборів даних. Порівняно з майже 15 000 статей у наборі даних 1, моделі, навчені на наборі даних 2, мали набагато менше прикладів, на яких можна було вчитися, тому не досягли такого ж рівня продуктивності. Другою причиною є потенційна слабкість у наборі даних 1. Як реальні так і сатиричні дані були взяті лише з 1 публікації, тому відповідна модель могла не лише дізнатися характеристики справжніх, фейкових і сатиричних новин, але й врахувати особливості цих конкретних публікацій, штучно підвищуючи продуктивність моделей у Е1 та Е2. Третя причина полягає в тому, що функції були певною мірою оптимізовані для набору даних 1, що могло дати моделям, навченим на наборі даних 1, невелику перевагу над моделями, навченими на наборі даних 2.

Найкращі моделі створені для трьох під експериментів в Е4, порівнювалися безпосередньо з моделями для набору даних 2, наведеними в [8]. Це справедливе порівняння, оскільки обидві моделі використовують однакові вхідні дані для навчання та тестування. Відмінностями між

підходом у цій кваліфікаційній роботі та підходом у роботі [8] є вибрані функції та гіперпараметри.

У роботі [8] не повідомляється про точностей, повноту або оцінку F1 отриманих моделей. Через це неможливо визначити, чи такі моделі надають перевагу певному класу. В кваліфікаційній роботі доведено більшу ефективність у відрізненні справжніх новин від фейкових завдяки збільшенню точності реальних і фейкових моделей цієї роботи на 9% порівняно з еквівалентною моделлю в [8] . Однак модель в роботі [8] має перевагу в ідентифікації сатиричних новин, маючи перевагу 4% для категорії фейків проти сатири та перевагу в 11% для категорії справжніх новин проти сатири. Це свідчить про недостатню ефективність моделей виявлення сатиричних новин за межами набору даних 1. В категорії «реальні проти всіх» моделі, розроблені в E4, досягають результатів 87% за всіма ключовими показниками продуктивності, випереджаючи в категорії «реальні проти фальшивої» та незначно програючи в категорії «реальні чи сатиричні» в порівнянні з моделлю, розробленою в [8].

Моделі, розроблені в [8], протестували понад 60 функцій, перш ніж зупинитися на двох моделях, кожна з яких використовує чотири функції. На відміну від набору функцій 2, ці функції були вибрані спеціально для їх продуктивності на наборі даних 2 без урахування пояснюваності, що дає моделям в [8] невелику перевагу в порівнянні. Незважаючи на це, отримані результати показують, що моделі, розроблені з використанням зрозумілих функцій, можуть досягати рівня продуктивності, порівнянного з моделями, розробленими з використанням незрозумілих функцій.

Відзначимо, що якість загальнодоступних наборів даних у просторі виявлення дезінформації загалом досить низька для використання в моделях класифікації статей. Наприклад, такі веб-сайти, як Kaggle, містять набори даних, на яких легко досягти майже ідеальної продуктивності класифікації з мінімальною роботою, однак, коли ці моделі тестуються на інших наборах

даних, вони, як правило, працюють майже випадково. Існують інші набори даних для класифікації тексту, але вони зазвичай зосереджені на аналізі настроїв менших частин тексту. З цієї причини створення узагальнених моделей, які можуть добре працювати в різних наборах даних, є серйозною проблемою. Також слід відзначити труднощі в розрізненні фейкових і сатиричних новин, оскільки різні публікації сатиричних статей мають дуже різні характеристики, часто поєднуючи риси та стилі написання фейкових і справжніх новинних статей. Результати, отримані в кваліфікаційній роботі, підтверджують в цілому можливість створення ефективних узагальнених моделей для виявлення дезінформації. Зокрема, ця робота демонструє високу ефективність отриманих моделей в категоріях реальних новин проти фейкових, справжніх новин проти сатири та реальних новин проти поєднання фейкових новин і сатири.

4.6 Обмеження дослідження та перспективи покращення моделей класифікації

У цьому дослідженні існують три обмеження. Першим обмеженням є недосконаліть наборів даних, які використовуються. Набір даних 1 містив лише реальні та сатиричні статті з однією публікацією в кожній, що призводить до деяких проблем із переналаштуванням функцій публікацій, а не характеристик реальних і сатиричних новин загалом. Набір даних 2 містив лише 225 статей, що є недостатнім для ефективного навчання великих і складних моделей.

Друге обмеження полягає в тому, що створені моделі не піддаються поясненню. Мета цієї роботи полягала у створенні моделей, які можна пояснити. Ця мета була частково досягнута завдяки моделям, які використовують функції і методи, що можна пояснити, але отримані моделі не можна оцінювати як моделі, що повністю пояснюються. Через це їх

можна порівнювати безпосередньо на основі продуктивності з іншими моделями, які не піддаються поясненню.

Третє обмеження полягає в тому, що всі моделі були створені з використанням базових моделей машинного навчання, таких як логістична регресія, та оптимізованих функцій. Більш складні моделі, такі як згорточні нейронні мережі, не розглядалися, тому неможливо передбачити, як повністю непояснена модель працюватиме на тих самих даних, якщо вона не обмежена у виборі функцій.

Розглянемо можливі перспективи покращення результатів, отриманих в кваліфікаційній роботі. Вони, насамперед, пов’язані з подоланням відзначених вище обмежень. Набір даних 1 можна покращити, зібравши більше даних із різних публікацій, можливо, з джерел за межами України та Європи, оскільки це дозволить моделям, навченим на цьому наборі даних, краще відповідати загальній проблемі детектування фейкових новин. Цей процес може виявитися складним, оскільки важко ідентифікувати публікації як такі, що створюють лише справжні новини. Вирішенням цього може бути налагодження контактів з організаціями, що перевіряють факти, з метою отримання доступу до корпусів вже класифікованих статей, оскільки це сприятиме підвищенню якості формування наборів даних для навчання та тестування моделей.

Створення зрозумілої структури для пояснення класифікацій з використанням згорткової нейронної мережі також є перспективним розвитком досліджень, проведених в кваліфікаційній роботі.

ВИСНОВКИ

В кваліфікаційній роботі було досліджено можливість створювати високоефективні узагальнені моделі для розрізнення справжніх і фальшивих статей, справжніх і сатиричних статей та справжніх проти сатиричних і фейкових статей разом. Відзначено складність розрізнення фейкових і сатиричних статей через те, що характеристики фейкових і сатиричних новин мають багато подібностей. Проведено детальний аналіз характеристик класів даних у двох наборах даних, використаних у цій роботі.

Проведені дослідження показали, що ефективні моделі машинного навчання можна розробити для вирішення різних завдань проблеми виявлення дезінформації в новинних статтях. Було показано, що цього можна досягти за допомогою функцій, зрозумілих користувачам. Відзначено доцільність проведення додаткових досліджень для створення високоефективних та повністю зрозумілих моделей машинного навчання, які можуть вирішити загальну проблему виявлення фейкових новин в статтях.

Таким чином було проведено аналіз відмінних рис справжніх, фейкових і сатиричних новин, що може бути цінним не лише у сфері машинного навчання, але й потенційно в проектах, де розглядаються різні аспекти класифікації новин. Моделі, отримані в кваліфікаційній роботі, можуть бути корисними для подальшої побудови систем автоматизованого виявлення дезінформації в новинних статтях в мережі Інтернет.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Як розпізнати фейкову новину. Інфографіка Міжнародної федерації бібліотек. Детектор медіа. URL: https://ms.detector.media/mediadoslidzhennya/post/22563/2019-03-08-yak-rozpiznaty-feykovu-novynu-infografika-mizhnarodnoi-federatsii-bibliotek/ (дата звернення: 24.04.2023).

2. A year in fake news, and what to look forward to (or how to tune out) in 2018. NiemanLab. URL: https://www.niemanlab.org/2017/12/a-year-in-fake-news-and-what-to-look-forward-to-or-how-to-tune-out-in-2018/ (дата звернення: 24.04.2023).

3. Staff E. Fake News: True or False Quiz Book. Egmont Books, Limited, 2020. 96 с.

4. Nautiyal D. ML | Underfitting and Overfitting. GeeksforGeeks. URL: https://www.geeksforgeeks.org/underfitting-and-overfitting-in-machine-learning/ (дата звернення: 25.04.2023).

5. Ahmed H., Traore I., Saad S. Detecting opinion spams and fake news using text classification. Security and Privacy. 2017. Т. 1, № 1. С. e9. URL: https://doi.org/10.1002/spy2.9 (дата звернення: 25.04.2023).

6. Twitter as a sentinel tool to monitor public opinion on vaccination: an opinion mining analysis from September 2016 to August 2017 in Italy / L. Tavoschi та ін. Human Vaccines & Immunotherapeutics. 2020. Т. 16, № 5. С. 1062–1069. URL: https://doi.org/10.1080/21645515.2020.1714311 (дата звернення: 25.04.2023).

7. How to Spot Fake News. FactCheck. URL: https://www.factcheck.org/2016/11/how-to-spot-fake-news/ (дата звернення: 24.04.2023).

8. Sharma D. K., Garg S. IFND: a benchmark dataset for fake news detection. Complex & Intelligent Systems. 2021. URL: https://doi.org/10.1007/s40747-021-00552-1 (дата звернення: 25.04.2023).

9. Conroy N. K., Rubin V. L., Chen Y. Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology. 2015. Т. 52, № 1. С. 1-4. URL: https://doi.org/10.1002/pra2.2015.145052010082 (дата звернення: 24.04.2023).

10. Horne B., Adali S. This Just In: Fake News Packs A Lot In Title, Uses Simpler, Repetitive Content in Text Body, More Similar To Satire Than Real News. Proceedings of the International AAAI Conference on Web and Social Media. 2017. С. 759–766. URL: https://doi.org/10.1609/icwsm.v11i1.14976 (дата звернення: 25.04.2023).

11. Kai Shu, Limeng Cui, Suhang Wang, Dongwon Lee, Huan Liu. Defend: Explainable fake news detection. International Conference on Knowledge Discovery & Data Mining, 2019, 395-405 с.

12. Vosoughi S., Roy D., Aral S. The spread of true and false news online. Science. 2018. Т. 359, № 6380. С. 1146–1151. URL: https://doi.org/10.1126/science.aap9559 (дата звернення: 24.04.2023).

13. Automatic Detection of Fake News / B. Kleinberg та ін. New York, 2017. URL: https://arxiv.org/abs/1708.07104 (дата звернення: 24.04.2023).

14. Tony Harcup, Deirdre O'neill. What is news? news values revisited (again). Journalism studies, 2017, 1470–1488 с.

15. Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social. URL: https://doi.org/10.1007/s11109-019-09533-0 (дата звернення: 25.04.2023).

16. Anandpara R. Spam E-Mail Filtering: A Review of Techniques. International Journal for Research in Applied Science and Engineering Technology. 2021. Т. 9, № VI. С. 5098–5101. URL: https://doi.org/10.22214/ijraset.2021.35992 (дата звернення: 25.04.2023).

17. Faustini P., Covoes T. Fake News Detection Using One-Class Classification. 2019 8th Brazilian Conference on Intelligent Systems (BRACIS), URL: https://doi.org/10.1109/bracis.2019.00109 (дата звернення: 25.04.2023).

ДОДАТОК А

Відомість кваліфікаційної роботи

Позначення	Найменування	Дод. відомості
	Текстові документи	
1.	Пояснювальна записка	_86_ с.
	Інші документи	
2.	Презентаційні матеріали	_15_ плакатів

.	.	Прізвище та ініціали.	Підп.	Дата	
Розробив		Панкратов Є.О.			Модель машинного навчання для виявлення фейкової складової у новинних текстах
Перевірив		Чала Л. Е.			
Н.контр.		Малєєва І.А.			
Затв.		Філатов В.О.			

Шифр групи	Код напр./спец.
СШІм-21-2	122
ХНУРЕ кафедра ШІ	