НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
НАВЧАЛЬНО-НАУКОВИЙ ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
Кафедра інформаційної безпеки

До захисту допущено
Завідувач кафедри

_____ **Дмитро ЛАНДЕ**
(підпис)

«_____»_____2023 р.

Дипломна робота

на здобуття ступеня бакалавра
за освітньо-професійною програмою
«Системи, технології та математичні методи кібербезпеки»
спеціальності 125 «Кібербезпека»

на тему: «Ідентифікація фейків за допомогою прихованої моделі Маркова»

Виконав (-ла): здобувач вищої освіти **IV** курсу, групи ФБ-91
(шифр групи)

Осьмак Анастасія Антонівна_____ _____
(прізвище, ім'я, по батькові) (підпис)

Керівник д.т.н., професор каф. ІБ, заслужений діяч Качинський А.Б. _____
(посада, науковий ступінь, вчене звання, прізвище, ім'я, по батькові) (підпис)

Рецензент _____ _____
(посада, науковий ступінь, вчене звання, науковий ступінь, прізвище, ім'я, по батькові) (підпис)

Засвідчую, що у цій дипломній роботі немає запозичень з праць інших авторів без відповідних посилань.
Здобувач вищої освіти _____
(підпис)

Київ – 2023 року

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
НАВЧАЛЬНО-НАУКОВИЙ ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
Кафедра інформаційної безпеки

Рівень вищої освіти – перший (бакалаврський)
Спеціальність – 125 «Кібербезпека»
Освітньо-професійна програма «Системи, технології та математичні методи кібербезпеки»

ЗАТВЕРДЖУЮ
Завідувач кафедри

_____ Дмитро ЛАНДЕ
(підпис)

«___»_____2023 р.

ЗАВДАННЯ
на дипломну роботу здобувачу вищої освіти

Осьмак Анастасії Антонівні
(прізвище, ім'я, по батькові)

1. Тема роботи «Ідентифікація фейків за допомогою прихованої моделі Маркова»,
керівник роботи Качинський Анатолій Броніславович, кандидат технічних наук, професор кафедри інформаційної безпеки, заслужений діяч,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від «26» травня 2023 р. № 2028-с

2. Термін подання здобувачем вищої освіти роботи 14 червня 2023 р.

3. Вихідні дані до роботи: класифікація фейкових новин, прихована модель Маркова для ідентифікації фейкових новин, огляд станів прихованої моделі, набори даних для тренування моделі англійською та українською мовами.

4. Зміст роботи: огляд та класифікація фейкових новин, побудова прихованої моделі Маркова для ідентифікації фейків, збір даних для навчання моделі англійською та українською мовами, начання моделів, проведення випробувань на ідентифікацію фейків. Аналіз результатів дослідження.

5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо): презентація.

6. Дата видачі завдання: 24 жовтня 2022 року.

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів дипломної роботи	Примітка
1	Отримання завдання	24.11.2022	виконано
2	Огляд теоретичних відомостей за темою дипломної роботи	25.11.2022 – 30.12.2022	виконано
3	Робота над першим розділом дипломної робготи	08.01.2023 – 25.01.2023	виконано
4	Робота над першим розділом дипломної робготи	25.01.2023 – 15.02.2023	виконано
5	Розробка прихованої моделі Маркова для ідентифікації фейкових новин	16.02.2023 – 01.03.2023	виконано
6	Робота над третім розділом дипломної робготи	01.03.2023 – 10.03.2023	виконано
7	Робота над четвертим розділом дипломної робготи	10.03.2023 – 30.03.2023	виконано
8	Проведення дослідження та аналіз результатів	30.03.2023 – 15.04.2023	виконано
9	Робота над п’ятим розділом дипломної робготи	15.04.2023 – 30.05.2023	виконано

Здобувач вищої освіти _____ Анастасія Осьмак
(підпис) (**Власне ім’я, ПРІЗВИЩЕ**)

Керівник роботи _____ Анатолій Качинський
(підпис) (**Власне ім’я, ПРІЗВИЩЕ**)

РЕФЕРАТ

Робота складається з 64 сторінки, містить 11 ілюстрацій, 3 таблиці, 1 додаток та 8 літературних джерел.

Об’єкт дослідження – ідентифікація фейкових новин.

Предмет дослідження – розробка прихованої моделі Маркова для ідентифікації фейкових новин.

Метою даної роботи є огляд та класифікація сучасних сфальсифікованих новин та розробка прихованої моделі Маркова як потужного засобу для їх виявлення.

Методи дослідження: літературний огляд, аналіз, розробка програмного коду, моделювання, класифікація, експериментальна оцінка, порівняння.

Новизна даної роботи полягає у використанні прихованої моделі Маркова для ідентифікації фейкових нових написаних як англійською так і українською мовами.

Прихована модель Маркова реалізована мовою Python з використанням бібліотеки spaCy, для зображення результатів використані діаграми Cloud Word.

Ключові слова: фейкі, фейкові новини, прихована модель Маркова, обробка природнього тексту, ідентифікація.

ABSTRACT

The paper consists of 64 pages, contains 11 illustrations, 3 tables, 1 appendix and 8 references.

The object of research is the identification of fake news.

The subject of the study is the development of a hidden Markov model for fake news identification.

The purpose of this paper is to review and classify modern fake news and develop a hidden Markov model as a powerful tool for identifying it.

Research methods: literature review, analysis, program code development, modeling, classification, experimental evaluation, comparison.

The novelty of this work is the use of a hidden Markov model to identify fake news articles written in both English and Ukrainian.

The hidden Markov model is implemented in Python using the spaCy library, and Cloud Word diagrams are used to display the results.

Keywords: fake news, fake news, hidden Markov model, natural text processing, identification.

Зміст

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ 7
ВСТУП 8
1 ПОНЯТТЯ ФЕЙК 10
1.1 Історія та розвиток фейків 10
1.2 Класифікація фейків 11
Висновки до розділу 1 14
2 РОЗПІЗНАВАННЯ ТА АНАЛІЗ ФЕЙКОВОЇ ІНФОРМАЦІЇ 15
2.1 Способи розпізнавання та аналізу 15
2.2 Алгоритми машинного навчання для виявлення фейкових новин 26
Висновки до розділу 2 28
3 ПРИХОВАНА МОДЕЛЬ МАРКОВА ДЛЯ ІДЕНТИФІКАЦІЇ ФЕЙКІВ 29
3.1 Прихована модель Маркова. Опис 29
3.2 Прихована модель Маркова в рамках ідентифікації фейкових новин 30
Висновки до розділу 3 37
4 ОТРИМАННЯ ТА ОБРОБКА ДАНИХ ДЛЯ НАВЧАННЯ ПММ 38
4.1 Збір даних 38
4.2 Інструменти для обробки та роботи з даними 43
4.3 Обробка природної мови 45
Висновки до розділу 4 54
5 АНАЛІЗ РЕЗУЛЬТАТІВ 55
Висновки до розділу 5 58
ВИСНОВКИ 59
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ 60
ДОДАТОК А КОД ПРОГРАМНОЇ РЕАЛІЗАЦІЇ ПРИХОВАНОЇ МОДЕЛІ МАРКОВА 61

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

CNN – Convolutional Neural Network

NLTK – Natural Language Toolkit

NLU - Natural Language Understanding

API – Application Programming Interface

ПММ – Прихована Модель Маркова

ВСТУП

На теперішній час, проблема пов’язана з виявленням фейків набула неабиякої популярності, бо фейк є досить гарним способом переконання, маніпуляції або просто впливу на свідомість людей. Залякування, введення в оману широкого загалу надасть непогану перевагу під час ведення гібридної, інформаційної або просто війни у загальному розумінні цього слова, тому у сьогоденні дуже важливо вміти правильно фільтрувати отриману інформацію, а краще мати певні методи для виявлення сфальсифікованих даних. Кожного дня через медіа ми отримуємо багато дуже різної інформації, але чи вся вона правдива і націлена лише на те щоб оповістити нас про якось подію або розширити кругозір? На жаль, не можна довіряти усьому що чуємо або бачимо, потрібно завжди аналізувати, але мозку людини важко виявляти все, що хтось хоче від нього приховати або видати за дійсне, тому тут на допомогу може прийти машинне навчання та штучний інтелект, що може з легкість обробити великі об’єми даних та проаналізувати їх потрібним чином. Тому, на мою думку, важливо шукати способи для застосування штучного інтелекту у задачах з виявлення фейків для їх оптимізації і більшої результативності.

Актуальність роботи. Оскільки, у сьогоденних реаліях, в поширеною проблемої будь якого інформаційноо середовища є використання змінених або цілковито підроблених новин з метою введення в оману загалу людей шляхом проведення інформаційно-психологічних операцій, тому дуже важливо мати інструменти для виявлення сфальсифікованої інформації, а отже тема є надзвичайно актуальною на теперішній час.

Наукова новизна. Новизна даної роботи зумовлена тим, що при широкому застосуванні штучного інтелекту для проблем у сфері кібербезпеки, прихованій моделі Маркова не надавалось особливої уваги, як потужному інструменту машинного навчання для роботи з дезінформацією та фейками, які є наразі надзвичайно поширеною проблемою інформаційного простору.

Мета. Метою даної роботи є огляд та класифікація сучасних сфальсифікованих новин та розробка прихованої моделі Маркова як потужного засобу для їх виявлення.

Об’єкт дослідження. Ідентифікація фейкових новин.

Предмет дослідження. Розробка прихованої моделі Маркова для ідентифікації фейкових новин.

Практичне значення. За допомогою запропонованого у даній роботі метода для ідентифікації фейкових новин можна покращити взаємодію користувача з інформаційними ресурсами, а саме збільшити його довіру до них, а також надати потужний інструмент для перевірки будь якої отриманої інформації на приналежність її до фейкової.

Апробація результатів роботи та публікації. Робота була опублікована у збірці робіт другої Міжнародної науково-практичної конференції «Кібербезпека державних інституцій та подолання кризових станів» на базі Національного технічного університету України КПІ імені Ігоря Сікорського, Вроцлавського Університету та Київського національного економічного університету імені Вадима Гетьмана.

2 ПОНЯТТЯ ФЕЙК

1.1 Історія та розвиток фейків

Саме поняття фейк з’явилось досить недавно, перші згадки про нього можемо знайти приблизно у 70х роках минулого століття, але чи не зустрічались фейки у людському житті раніше, можливо під іншими назвами?

З англійської слово фейк переводиться як підробка або фальшивка, але у нашому контексті – це підробка, фальсифікація, викривлення певної інформації з метою введення широкого загалу в оману. Ще у давнину люди вдавались до підробок певних документів, внесення у них своїх коректив з метою отримання певної вигоди для себе, також фейки активно використовувались під час пропаганди, наприклад під час виборів голови держави, для підтримки авторитету певної особи або політичного режиму, придушення авторитету суперника тощо.

Непоганим прикладом саме з історії України можна привести діяльність КДБ СРСР проти українського мовника Радіо Свобода, проти якого велась активна пропагандистська діяльність з ціллю придушення мовлення та переконанні слухачів у неправдивості поданої їм інформації з цього радіо. КДБ поширювало неправдиву інформацію щодо працівників щодо працівників даного медіа, звинувачувало інші держави у веденні пропаганди та розвідки через дане джерело на свою користь: «Більшість американців, що працюють на станції, є євреями за національністю, вихідцями з території СРСР. Усі керівні посади зайняті, як правило американцями, значний відсоток з яких складають співробітники ЦРУ» [1] йдеться у довідці Комітету держбезпеки за 1969 рік. При цьому це були лише безпідставні звинувачення, доказів яким не було, це і схожі повідомлення були побудовані лише як словесно яскраве затвердження певного факту. Оскільки це медіа джерело об’єднувало під собою багато відділів з різних країн, то для частини населення так заяви виглядали досить правдиво. Також комітет держбезпеки намагався оправдувати й

перекривлювати факти про владу подані на Радіо Свобода, що однозначно можна назвати фейками, так як і плакати зі звинувачуваннями у нацизмі та співпраці з Заходом. Загалом Радянське правління дуже боялось висвітлення правди у медіа, тому й намагалось підроблювати або й зовсім забороняти вільне поширення інформації, натомість людям подавалась викривлена реальність, прикрашена різними пропагандистськими наративами.

Насправді, багато неправдивої інформації можна приховати під гучними заголовками та правильним поданням матеріалу що може зацікавити аудиторію. Прикладом цього може виступати афера втілена у житті американський журналістом Стівеном Глассом, який пропрацювавши у видання всього три роки сфабрикував більше половини своїх статей, опублікованих у журналі, але цікавим фактом є те що до нього весь цей час не виникало жодних питань, а навпаки його популярність зростала. Чоловік навіть розробив сайт компанії про яку писав статтю, щоб підтвердити її правдивість, але саме на ній й викрили його аферу. Отже, з цього можемо зробити висновки, що мало хто перевіряє та аналізує отриману інформацію, а історія що нас зацікавила навіть у виданні до якого ми маємо довіру може виявитись фейком, який з легкістю може вплинути на наше життя та світосприйняття, тож можна уявити якою небезпечною зброєю насправді є фейк.

1.2 Класифікація фейків

Одним із важливих кроків перед розробкою способу для ідентифікації фейкових новин є створення їх класифікації, оскільки вона допоможе отримати чітке розуміння різних типів фейків та їхніх характеристик. Основні переваги класифікації фейків полягають у тому, що вона, по-перше, дозволить ідентифікувати та групувати фейкові новини за спільними характеристиками, такими як спосіб їх створення, типи маніпуляцій, використані медіа формати тощо. По-друге, розробивши правильну класифікацію зусилля будуть

зосереджені саме на розробці ефективних методів ідентифікації для кожного типу фейків. По-третє, допоможе визначити основні сфери, де вони найбільш поширені або мають найбільший вплив. Це може бути корисним при розробці пріоритетів і розподілі ресурсів для боротьби з фейковими новинами. Також вона може бути корисною для встановлення основних закономірностей та особливостей, які використовуються для створення певного типу фейків, що сприятиме розробці спеціалізованих моделей і алгоритмів, які ефективно виявляють визначені типи фейків.

Загалом, класифікація фейків є важливим кроком у боротьбі зі спалахами фейкових новин. Вона допомагає зрозуміти актуальне розмаїття фейків і допомогти у розробці практичних методів їхньої ідентифікації та запобігання поширенню.

1.2.1 Класифікація фейкових новин запропонована університетом Онтаріо

На жаль, немає єдиної класифікації фейків, але різні наукові структури та медіа намагались певним чином їх класифікувати. У публікації науковців з університету Онтаріо фейки поділяють на три категорій

1. Серйозні фабрикації (фейкові новини, які створені спеціально):

до цього типу можна віднести шахрайські репортажі, націлені лише на обман і привернення уваги, жовта преса, яка використовує яскраві заголовки та цікаві теми, які приваблюють аудиторію та тим самим сприяють збагаченню таблоїда, збільшення популярності. Найчастішими темами, які можна побачити серед жовтої преси є кримінальні історії (вбивства, пограбування тощо), плітки про зірок або політиків, сенсаційні історії тощо.

2. Масштабні містифікації: до цього типу відносяться вигадки поширені в соціальних мережах, наприклад інформація розміщена на сторінці деякого популярного блогу, яка може бути замаскована під новину з ціллю обманути широкий загал людей.

3. Гумористичні фейки: жартівливі новини, сатира, за основу може бути взята реальна новина і до неї додана певна частка жартівливого контенту. Зазвичай читачі не сприймають серйозно такі джерела, бо вони несуть у собі розважальну мету, але у кожному жарті може бути доля правди, тому хтось може сприйняти гумористичну новину за правду.

Прикладом цього може виступати джерело The Onion – американський розробник іронічних новин, який представлений у вигляді газети, а також веб сайти які містять у собі як і місцеві так і жартівливі новини представлені у розважальній формі. Оскільки ця агенція дуже користується популярністю серед читачів, які люблять ділитись цікавими й веселими новинами з друзями й близькими то вона може нести неабияку загрозу, бо по суті вся начинена фейками, які за допомогою аудиторії безупинно поширюються і вводять багатьох в оману.

1.2.2 Класифікація фейковий новин запропонована BBC

BBC пропонує схожу, але дещо іншу класифікацію фейків:

- Першим видом вони виділяють спеціально сфальсифіковані новини, поширення яких відбувалось навмисно, тобто новини були розміщенні у популярних паперових та інтернет виданнях, медіа спеціально, за сприянням певної особи або групи осіб, які за допомогою цього намагаються отримати для себе матеріальну, моральну або іншу вигоду.
- Другим видом за BBC є не абсолютно точні новин, тобто це ті що містять у собі більшу частину правдивої інформації, але деякий відсоток новини, факт, дата, інформація про учасників події може бути гіперболізована або викривлена, або зовсім сфальсифікована, але при цьому наприклад подія описана у новині може бути цілковито реальною.

Висновки до розділу 1

Отже, у цьому розділі було цілковито описано суть поняття фейк, розглянуто певні моменти з історії його формування та чим саме фейкові новини можуть бути небезпечні для суспільства. Також, розглянувши деякі класифікації фейків, ми вже можемо зробити для себе певні припущення, шукаючи і переглядаючи різні джерела інформації, на можливість наявності у них фейкової частини, але видавництва намагаються як найкраще приховати і замаскувати неправдиву інформацію, щоб у читача не виникало жодної підозри, тож далі треба розібратися зі способами виявлення фейків та як їх віднайти у широкому потоці інформації, яку ми щоденно черпаємо з різних джерел.

2 РОЗПІЗНАВАННЯ ТА АНАЛІЗ ФЕЙКОВОЇ ІНФОРМАЦІЇ

2.1 Способи розпізнавання та аналізу

На сьогоднішній день існує багато способів розпізнавання фейкової інформації, особливо фейкових зображень, які використовуються дуже часто в різних каналах у соціальних мережах та веб виданнях іноді через дефіцит зображень для описання певної новини, адже наявність фото та відео, постерів розширюють спектр емоцій, що виникають у читачів під час перегляду інформації та впливають на її сприйняття.

Але іншою стороною медалі є те що старі фото та відео часто модифікуються і видаються за нові з метою гіперболізації описаних подій, таке часто відбувається з зображеннями з місць жахливих катастроф, катаклізмів, воєн тощо, адже деякі такі фото можна з легкістю доповнити та переробити для видання однієї ситуації за іншу або зміні учасників та місця події. Прикладом цього можуть виступати модернізовані фото війни з Сирії певними медіа виданнями та розміщення їх у Інтернет просторі як фото звітування з бойових дій на сході України 2014 року.

На даний момент використання старих фото та видання їх за нові є дуже поширеною проблемою через не дуже стабільний стан подій у світі та емоційної складової у цьому, тож допомога звичайним громадянам у виявленні сфальсифікованої проти них інформації є дуже актуальною проблемою. На разі вже існують деякі варіанти пошукових програм, що здійснюють пошук за фото та дуже спрощують життя людям, найпростіші з них це плагіни, які просто встановлюються на пошуковий браузер і при обранні фото або відео, яке у нас підпало під сумнів, за допомогою плагіну ми можемо знайти дати і видання у яких воно фігурувало раніше і основуючись на цьому зробити певні висновки. Прикладом такого плагіну є RevEye, який використовується для Google Chrome і функціонує за зазначеним раніше сценарієм. Також ще одним прикладом, який вже можна використовувати у декількох браузерах є плагін Who stole my

pictures, що був початково створений з метою перевірки власних публікацій на те чи використовуються вони ще якимись користувачами або веб ресурсами, але загалом він має такий самий принцип дії як і попередній плагін для Google Chrome. Також існують певні додатки що здійснюють пошук за геолокацією фото та датою його створення, що може допомогти для проведення певного слідчого експерименту, пошуку свідків певної події та загалом огляду певної локації для отримання потрібної нам інформації.

Проаналізувавши низку таких додатків, можна сказати що вони використовують доволі різні способи для припущенні правдивості чи неправдивості інформації, тож далі розглянемо які саме складові є важливими для ідентифікації фейків та які взагалі існують аспекти на які варто звернути увагу при аналізі новин.

2.1.1 Візуальна складова у розпізнаванні фейків

Візуальна складова будь якої новини є одним з основних пунктів на які ми маємо звертати увагу під час аналізу інформації. Вона багато говорить як і про саму інформації так і її приналежність до фейків. Дослідниками у сфері розпізнавання фейків за візуальною складовою розділили фейки за деякими ознаками за якими можна проводити аналіз: ознаки криміналістики, семантичні, смислові та статистичні.

Під криміналістичними мається на увазі такий собі пошук доказів фальсифікування зображення або відео, які автор залишив по собі, бо безслідно змінити щось є неможливим. Це можуть бути проміжки залишені під час суміщення декількох відео, перехід між якістю зображення на різних моментах у відео, кольорова гамма, інтерполяція, шуми тощо. Для виявлення таких маніпуляцій використовуються бінарні паттерни на зображенні, які змінюються відповідно до проведеної над ними маніпуляції. Широке застосування нейронних мереж також додало роботи слідчим по фейковим новинам, адже штучний інтелект активно використовується для генерування

нового медіа з високою долею правдоподібності. Загалом слідча експертиза у таких випадках базується на аналізі сигналів, спектральних характеристиках та кореляції, але оскільки більшість підробок базується на підробленні відео контенту з участю людей то припущення щодо фейковості робились і на основі візуальної складової, а саме на недостатньо природній поведінці персонажу у кадрі, деформації його обличчя.

2.1.2 Семантична складова у розпізнаванні фейків

Не менш важливою ознакою яка допоможе відрізнити фейкову новину від звичайною є семантична складова. Як вже було вказано раніше, автори фейковий новин часто зображують їх у такому вигляді, який викличе якнайбільше емоцій у людини, тому тут візуальна складова неймовірно важлива, бо вона у цьому випадку керує почуттями читача. Одним із найпотужніших інструментів для дослідження семантики зображень є згорткова нейромережа VGG(візуальна геометрична група) від CNN. Вона використовується для різного роду зображень, але також може бути використана для дослідження фейків. Ця нейромережа складається з 13 згорткових шарів та ще 3 основних та об’єднання між собою шарів. Тож, розглянемо трошки глибше архітектуру:

- Ввід, на який модель отримує зображення чітко визначеного розміру, а саме 244×244
- Згорткові шари, які вилучають основні ознаки зображення і всі трансформації що на нього накладені для подальшого аналізу.
- Об’єднання шарів, для зменшення кількості ознак для аналізу.
- З’єднані шари, які використовуються для вирішення задач класифікації.

Крім цієї моделі у розробках CNN містяться це деякі схожі моделі, але більш спрямовані на пошук фейковий новин, які працюють на основі

об’єднанні інформації яка базується на аналізі пікселів та частот, шар, який відповідає за аналіз піксельної складової й аналізує семантику. Також для виявлення деяких емоційних провокацій, які є частою складовою фейкової новини і є складовою семантичної частини використовуються розгорнуті мережі.

2.1.3 Статистичні риси та їх роль у роспізаванні фейків

Слідуючою ознакою для аналізу фейковий новин є статистичні риси. Простими словами статистична риса відповідає за кількість та варіативність зображень що відповідають тій чи іншій новині. На мою думку, може видатись досить підозрілим коли ми гортаємо нашу стрічку новин і на якусь шокуючу чи провокативну подію маємо лише одне зображення, навряд журналісти, які вели репортаж з місця подій, зробили лише одну або дві фотографії на підтвердження певної новини, або взагалі на всіх ресурсах маємо одні й ті ж фото з одного ракурсу. Тому проводити певну статистику, а саме можемо ввести статистичні функції що оброблюють аналіз розподілу правдивих новин і фейків. Розглянемо основні характеристики на яких базуються роботи з проведення статистичного аналізу:

- Кількісні відношення: прикладом цього при аналізу новини є зіставлення кількості зображень, використаних для її опису, з загальною кількістю додатків до подій, які описуються у новинах.
- Відношення популярності: аналіз базується на популярних картинках взятих зі стрічки, вони можуть мати багато реакцій, вподобайок, збережень, перенадсилань тощо і співставленні їх зі звичайними дописами.
- Аналіз типів зображень: загалом, читаючі різні портали з новинами ми бачимо зображення різного формату, іноді то чи інше джерело намагається слідкувати за збереженням єдиного формату, але й іноді зустрічаємо зовсім різні, тож аналізуючи й статистично виводячи найбільш популярний формат

серед потоку, основуючись на ньому ми можемо аналізувати й інші зображення.

Ми розглянули основні статистичні прийоми, але також маємо й більш глибокі й наукові:

- Оцінка чіткості: метод є розбіжності Куллбака-Лейблера між певними наборами зображень – колекцією(збірка зображень з певних подій) та подією(зображення з аналізуємої події), аналіз проводиться за формулою:

$$VCS = DKL(p(w|c)||p(w|k)), \quad (2.1)$$

де p(w|c) та p(w|k) оцінюють частоту появлення візуального зображення певного слова w, у колекції або у самій події відповідно.

- Оцінка когерентності: функція подібності застосовується для порівняння зображень з колекції події, яка оцінюється, на схожість, когерентність. Формулою для оцінки когерентності виступає:

$$VCoS = \frac{1}{|N*(N-1)|} * \sum_{i,j=1,\dots,N,j\neq i} sim\ (xi, xj\ ), \quad (2.2)$$

де N – кількість зображень у наборі, функція sim (xi, xj) описує схожість між двома деякими зображеннями зображеннями.

- Оцінка через гістограму розподілу: оцінюючи події цільової новини, складається матриця подібності зображень, на основі неї отримуємо певний вектор подібності зображень і основуючись на цих дослідженнях можемо побудувати гістограму розподілу подібності зображень на рівні мілкої деталізації.

- Оцінка багатогранності візуальної складової: оцінюється певна візуальна різниця між зображеннями з колекції, за основу береться найпопулярніше або зображення, що описує новину більш досконало, найкраще

по якості і на його основі проводиться ранжування низки інших зображень колекції. Маємо формулу:

$$VDS = \sum_{i=1}^{N} \frac{1}{i} \sum_{j=1}^{i} (1 - sim(xi, xj)) \quad (2.3)$$

- Оцінка через кластеризацію: застосовується метод кластеризації для оцінки колекції зображень з певної новини, для проведення цього використовується алгоритм ієрархічної агломеративної кластеризації.

2.1.4 Контекст та метадані та їх роль у аналізі інформації

Ще одним невід'ємним пунктом оцінки фейковості новини є її контекст. На жаль, перевірки на збіжність тексту з зображеннями не дають того результату на який ми очікуємо, бо творці фейкових новин ставлять собі за мету узгоджувати підписи і інформацію у своїх твореннях, тож цей метод більше стосується оцінки більш скритих факторів, таких як вміст вед сторінки на якому опубліковано новину, її візуальна складова і метадані зображень.

По-перше, розглянемо метадані. Це таки собі інформація про медіа файли, дата, час та місце створення, розмір, розширення тощо. Насправді, отримавши їх ми можемо дізнатися про будь яке відео чи зображення, мабуть, всю потрібну нам інформацію для подальшої оцінки, але, на жаль, дуже часто ці дані затираються після публікації і вже не можуть нам дати бажаного, але якщо вони присутні, то це може виступати потужним доказом щодо фейковості.

По-друге, крім метаданих, також активно використовуються контекстуальні характеристики, отримані із зовнішніх знань за допомогою зворотнього пошуку зображень. Зворотній пошук зображень відрізняється від традиційного пошуку тим, що він приймає зображення як вхідні дані та повертає список релевантних веб-сторінок, які містять відповідне зображення, разом із його назвою, описом і часом. Цю процедуру можна легко

автоматизувати та застосувати до величезної кількості зображень за допомогою API пошукових систем, таких як зворотній пошук зображень Google. Отже, розглянемо три функції контексту:

- Часові проміжки: визначаються як проміжки у часі між моментом публікації новини та моментом першої публікації візуального вмісту. Ця функція може бути використана для перевірки оригінальності візуального контенту. Якщо часовий інтервал перевищує певне порогове значення, то ймовірність того, що візуальний вміст походить від нерелевантної події, збільшується.
- Схожість між новинами: подібність між новинами визначається як подібність між заявою у новині та текстовим вмістом просканованих веб-сайтів. Враховуючи, що текстова інформація веб-сайтів є корисною для розуміння оригінальної події зображення, ця функція використовується для перевірки узгодженості події між текстовою претензією та відповідним візуальним вмістом.
- Надійність платформи: означає довіру до вихідної платформи, на якій було опубліковано візуальний контент. За допомогою набору даних, веб-сайту, який надає достовірну інформацію про понад 2700 медіа-джерел, кожну веб-сторінку, яку повертає зворотний пошук зображень, було класифіковано за такими категоріями: висока реальність, низька фактичність і змішана фактичність. Відсоток веб-сторінок з кожної категорії, отриманих зворотним пошуком зображень, був визначений як характеристика надійності платформи.

2.1.5 Значення візуального вмісту у аналізі новин

Раніше вже було розглянуто чотири типи візуальних функцій з різних точок зору, наприклад, криміналістичні, семантичні, статистичні та контекстні, для виявлення мультимедійних фейкових новин. Ці функції відображають характеристики візуального вмісту та зазвичай поєднуються на практиці для

охоплення більшої кількості ситуацій, які можуть бути сфальсифіковано. На разі також важливо розглянути й низку підходів, які використовують візуальний контент для виявлення фейкових новин, які можна в цілому класифікувати на підходи, засновані на контенті, і підходи, засновані на знаннях.

Підходи, засновані на вмісті, зосереджені на захопленні та поєднанні сигналів із вмісту різних модальностей для виявлення фейкових новин без використання будь-яких довідкових наборів даних.

Підходи, засновані на знаннях, спрямовані на використання зовнішніх джерел для перевірки фактів вхідних тверджень. Вони припускають існування відносно великого набору еталонних даних і оцінюють цілісність допису новин, порівнюючи його з одним або декількома дописами, отриманими з еталонного набору даних.

Загалом новина складається з текстового та візуального вмісту, обидві ці складові надають відмінні підказки або докази за допомогою яких ми можемо стверджувати про фейковість. Тому останні дослідження з цієї проблеми зосереджувались на використанні та ефективному поєднанні інформації з багатьох модальностей. Здебільшого вони просто використовували загальну рекурентну нейронну мережу (RNN) і попередньо навчену CNN для отримання текстових і візуальних семантичних характеристик, подальшого дослідження.

Для найсучасніших підходів властиве об’єднання зібраної інформації з усіх площин для виявлення фейкових новин, тобто створення мультимодальної інформації. Вперше включили мультимодальний контент через глибокі нейронні мережі, щоб вирішити проблему виявлення фейкових новин. Він запропонував інноваційний RNN із механізмом уваги для ефективного поєднання функцій текстового, візуального та соціального контексту. Для певної новини текст і соціальний контекст спочатку об’єднуються з LSTM для спільного представлення. Це уявлення потім поєднується з візуальними характеристиками, отриманими з попередньо навченого глибокого CNN.

Вихідні дані LSTM на кожному кроці часу використовуються як увага на рівні нейронів для координації візуальних характеристик під час злиття.

Також була запропонована наскрізна нейронна мережа подій для виявлення нових фейкових новинних подій на основі мультимодальних функцій, інваріантних до однієї події. Вона складається з трьох основних компонентів: мультимодального екстрактора ознак, детектора фейкових новин і дискримінатора подій. Мультимодальний екстрактор функцій відповідає за вилучення текстових і візуальних функцій із публікацій. Він співпрацює з детектором фейкових новин, щоб навчитися дискримінаційному представленню для виявлення фейкових новин. Роль дискримінатору полягає у видаленні специфічних для події функцій і збереженні спільних особливостей серед подій.

Таким чином відбувається аналіз новини на фейковість за допомогою контекстних даних.

Підходи, що ґрунтуються на знаннях базуються на тому, що мультимедійні новини часто складаються з кількох частин, як-от зображення чи відео з пов’язаним текстом і метаданими, де інформація про подію не повністю фіксується кожною з частин окремо, тобто кожна з них може нести нам певні частини інформації про подію. Такі пакети мультимедійних даних, тобто кортежі мультимодальної інформації про публікації, схильні до маніпуляцій, коли їх підмножина може бути модифікована для представлення або перепрофілювання мультимедійного пакета. Однак деталі, якими маніпулюють, є досить тонкими та часто переплітаються з правдою, через що підходи, що базуються на вмісті, навряд чи можуть виявити ці маніпуляції.

Зіштовхнувшись із цією проблемою, підходи, засновані на знаннях, вже використовують зовнішні джерела, набір довідкових даних необроблених пакетів як джерело світових знань, щоб допомогти перевірити семантичну цілісність мультимедійних новин.

Jaiswal вперше формально визначили проблему оцінки семантичної цілісності мультимедіа та об’єднали вивчення глибокого мультимодального

представлення без додаткових методів виявлення, щоб оцінити, чи відповідає підпис зображенню у своєму пакеті.

Пакети даних у еталонному наборі використовувалися для навчання моделі навчання глибокого мультимодального представлення, яка потім використовувалася для оцінки цілісності пакетів запитів шляхом обчислення балів узгодженості підписів із зображеннями та використання моделей виявлення викидів для визначення їхньої відповідності відносно еталонного набору даних.

Однією з головних проблем для розробки методів оцінки семантичної цілісності мультимедіа є відсутність даних для навчання та оцінки. У світлі цього було запропоновано нову структуру, Adversarial Image Repurposing Detection, для виявлення зміни призначення зображень, яку можна навчити за відсутності навчальних даних, що містять маніпульовані метадані. AIRD має моделювати реальну взаємодію між певним виконавцем, який використовує зображення з підробленими метаданими, і аналізатором, який перевіряє цю семантичну узгодженість між зображеннями та метаданими, що їх супроводжують. Зокрема, AIRD складається з двох моделей: фальсифіувтора та детектора, які навчені змагальним способом. Поки детектор збирає докази з набору посилань, фальсифікатор використовує їх, щоб створити переконливо оманливі фейкові метадані для даного пакета запиту.

2.1.6 На що потрібно звернути увагу під час озайомлення з новою інформацією

На жаль, переглядаючи інформацію у нас не завжди є час і бажання глибоко перевіряти побачене та прочитане, але все таки ми маємо право знати правду і огороджувати себе від брехні, тому маємо чітко знати правила за якими можемо зробити припущення щодо правдивості певної отриманої нами інформації, можливо з метою подальшої перевірки. По-перше, слід розглянути способи виявлення фейків у друкованих медіа, адже вони не піддаються одразу

перевірці за допомогою певних програм та інших автоматизованих способів аналізу.

Переглядаючи видання перше на що ми звертаємо нашу увагу це заголовок і саме він створює у нас перше враження і також є першочерговим способом виявлення або підозри на фейковість новини. Багато журналів та газет наповнені сенсаційними заголовками, що описують неймовірний факт або на чомусь наголошують та чи співпадає інформація у заголовку з тим що описано у статті? Це перший маркер який вказує на те що певна стаття може містити підроблені факти, бо що заважає авторам гучного заголовку написати ще більш гучні факти у статті.

Друге на що маємо звернути увагу це ставлення автора до події, яка описана, якщо автор погоджується з однією стороною конфлікту в описаній події і засуджує іншу це може бути маніпуляцією у наш бік, бо читаючи статтю ми маємо певну довіру до видання і автора тож маємо й підтримувати його думку і точку зору, а приймаючи його сторону ми піддаємось ймовірно свідомій маніпуляції. Такі прийоми часто можуть використовуватись під час проведення виборчої кампанії, з метою отримання прихильності аудиторії через широковідомих, впливових людей, які й можуть виступати авторами певних статей у медіа.

Й останній, але не менш важливий пункт на який ми маємо звернути увагу це джерела з яких взяті певні частини інформації, наприклад автори фото, імена людей що займались проведенням дослідження дати. Якщо ми не маємо жодних джерел то має одразу виникати підозра у фальсифікації цієї інформації, бо автори одразу обрізають нам шлях до її перевірки. Такі самі правила розповсюджуються й на ту інформацію, яку ми отримуємо з телевізора або Інтернету, але додаються й інші способи перевірки. Наприклад тональність репортажу. Репортаж не врівноважений і агресивний спрямований не лише на те щоб викликати емоції у глядача, але й також на те щоб висвітлити будь яку корисну інформацію для того чи іншого медіа ресурсу під певним потрібним світлом. Сприймати інформацію треба спокійно й врівноважено, контролювати

себе й намагатися не вестися на психологічні прийоми досвідчених експертів, з чистим розумом буде значно легше відрізнити правду від брехні і перебільшень.

2.2 Алгоритми машинного навчання для виявлення фейкових новин

Машинне навчання значно спрощує багато сфер людського життя, серед них і процес відокремлення справжніх новин від підроблених, тому ми маємо розглянути 5 найкращих алгоритмів машинного навчання для виявлення фейкових новин.

1. GNN

Нещодавні роботи, такі як Graph Neural Networks with Continuous Learning for Fake News Detection from Social Media та User Preference-aware Fake News Detection, використовують складну комбінацію представлення речень на основі уваги та навчання на них графової нейронної мережі. Ця система використовує текст, а також метадані користувача (наприклад, деталі твітів або історію таких зловмисних публікацій) не лише для виявлення неправдивих новин, але й для відстеження користувачів або акаунтів, які, ймовірно, генерують неправдиві новини, виходячи з їхньої нещодавньої поведінки, та запобігання подальшому поширенню таких публікацій.

2. BiLSTM + увага

Цей метод посів друге місце в рейтингу FNC-1 (Fake News Challenge Stage 1) із середньозваженим показником точності 84,60. У статті "Поєднання ознак схожості та глибокого навчання репрезентації для виявлення позиції в контексті перевірки фейкових новин" двонаправлені рекурентні нейронні мережі (RNN) використовуються разом із максимальним об'єднанням часового/послідовного виміру та нейронною увагою для представлення заголовка, перших двох речень новини та всієї статті. Ці репрезентації об'єднуються і передаються на кінцевий рівень, який прогнозує позицію новинних ЗМІ.

3. CNN+DNN

Цей відносно простий метод векторизує текст, використовуючи термін частота-обернена частота документа (TF-IDF). Він обчислює схожість між заголовками та основним текстом. Ця матриця подається на поверхневу ШНМ для вилучення ознак, а потім пересилається на глибоку нейронну мережу з прямим поширенням, яка закінчується шаром softmax. Цей метод добре працює для набору даних FNC-1, оскільки відмінності між правдивими і неправдивими новинами настільки очевидні.

Однак термін "частота, обернена до частоти документа" (TF-IDF) і поєднання CNN і DNN допомагають розпізнати

4. CNN+Boosted Trees

Алгоритм CNN+Boosted Trees - це підхід машинного навчання, який дозволяє виявляти неправдиві новини. Він поєднує в собі можливості згорткових нейронних мереж (CNN) і дерев, які виділяють ознаки з вхідного тексту і класифікують його як фейковий або справжній. У цьому алгоритмі спочатку потрібно попередньо обробити вхідний текст і перетворити його в матрицю вкладених слів. Потім потрібно застосувати CNN до матриці, щоб виділити суттєві ознаки з тексту. Результатом CNN є вектор ознак, який ви повинні ввести в модель Boosted Trees. Модель Boosted Trees - це дерево рішень, яке об'єднує декілька слабких класифікаторів для формування сильного класифікатора. На основі виокремлених ознак модель Boosted Trees навчається класифікувати новини як фейкові чи справжні.

Алгоритм CNN+Boosted Trees має кілька переваг для виявлення неправдивих новин. Він може обробляти короткі та довгі тексти, фіксувати їхній контекст і зміст. Цей підхід також менш схильний до перенавчання, ніж інші підходи машинного навчання. Поєднання CNNs і Boosted Trees дозволяє більш надійно і точно виявляти неправдиві новини.

5. MLP

Алгоритм MLP (Multilayer Perceptron) - це тип нейронної мережі, який можна використовувати для виявлення неправдивих новин за допомогою

машинного навчання. Він складається з декількох шарів взаємопов'язаних вузлів і може навчатися на маркованих даних, щоб класифікувати новини як фейкові чи справжні. Щоб виявити неправдиві новини, спочатку потрібно попередньо обробити вхідні текстові дані, щоб перетворити їх на числове представлення. Потім ви повинні використовувати це представлення як вхідні дані для MLP, який навчається класифікувати новини як фейкові або правдиві на основі ознак, витягнутих із вхідних текстових даних. Процес навчання передбачає коригування ваг вузлів мережі, щоб мінімізувати похибку між передбачуваними і фактичними мітками.

Алгоритм MLP має перевагу в управлінні складними, нелінійними зв'язками між вхідними та вихідними даними. Якщо навчальний набір даних великий і різноманітний, він також може виконувати завдання класифікації з високою точністю. Популярні інструменти машинного навчання, такі як TensorFlow і Keras, також можуть бути використані для ефективної реалізації алгоритму MLP.

Висновки до розділу 2

Отже, у цьому розділі було розглянуто деякі способи для розпізнавання та аналізу фейкових новин їх особливості, певні переваги та недоліки. Для подальшого дослідження буде використовуватись метод що базується на семантиці новини в поєднанні з машинним навчанням, бо він є найзручнішим для аналізу саме текстової інформації, яка і буде аналізуватися у межах даної роботи.

Також у цьому розділі було виділено основні особливості за якими звичайний користувач інформаційного середовища зможе відрізнити сфальсифіковані новини від правдивих, що є корисним у рамках збільшення довіри суспільства до інформаційних джерел.

3 ПРИХОВАНА МОДЕЛЬ МАРКОВА ДЛЯ ІДЕНТИФІКАЦІЇ ФЕЙКІВ

3.1 Прихована модель Маркова. Опис

Випадковий процес, який є найпростішим узагальненням послідовності незалежних випадкових величин, є дискретним ланцюгом Маркова. Майбутній випадковий характер процесу залежить тільки від поточного стану процесу і не залежить від його минулої історії, це називається властивістю Маркова. В кожному стані ланцюг Маркова може бути прямо спостережений. Але іноді є послідовність станів, яку потрібно знати, але не можна спостерігати безпосередньо, а тільки через спостережувані стани, це називається прихованим ланцюгом Маркова. Для того щоб вдало застосовувати приховану модель Маркова розберемося як працюють приховані ланцюги, розглянемо ймовірності виникнення фейкових новин і основні проблеми у прихованих ланцюгах Маркова: проблему оцінки, проблему декодування та проблему навчання.

Приховані моделі Маркова складаються з марківських ланцюгів, які в свою чергу містять кінцеву кількість станів, матрицю ймовірних переходів між станами, матрицю розподілу ймовірностей для початкового стану. Самі стани є прихованими, кожен з яких вимальовується через ймовірності розподілу. Розглянемо основні елементи прихованої моделі Маркова:

1. Набір з N прихованих станів, S = $\{S_1, S_2, \dots S_N\}$, зі станами у час t, позначаємо як q_t є S;
2. Початковий розподіл ймовірностей

$$\Pi = \{\pi i\}, \tag{3.1}$$

де $\pi i = P(Q1 = si), 1 \leq i \leq N$;

3. M, кількість чітких символів спостереження на кожен прихований стан, позначимо V = $\{v_1, v_2, \dots v_M\}$;
4. Матриця ймовірностей переходів,

$$[A]ij = \{aij\}, \text{де } aij = P(Qt + 1 = sj|Qt = sj), \quad (3.2)$$

де $1 \leq i \leq N$;

5. Матриця розподілу ймовірності спостереження

$$[B]jk = \{bj(vk)\}, \text{де } bj(vk) = P(Qt = vk|Qt = sj), \quad (3.3)$$

де $1 \leq i \leq N, 1 \leq k \leq M$

Дані значення N, M, A, B та π, можуть використовуватися в прихованій моделі Маркова як генератори для послідовності спостереження:

$O = \{O_1, O_2 \dots O_T\}$, де T це номер спостереження у послідовності.

Побудуємо можливий приклад прихованої моделі Маркова для ідентифікації фейків.

3.2 Прихована модель Маркова в рамках ідентифікації фейкових новин

Прихована модель Маркова (ПММ) — це статистична модель, яку можна використовувати для різноманітних програм, зокрема для розпізнавання мовлення, обробки природної мови та обробки сигналів. ПММ особливо корисні для проблем, де базові дані генеруються послідовністю прихованих станів, які неможливо безпосередньо спостерігати.

У контексті виявлення підроблених даних ПММ можна використовувати для моделювання послідовності ознак, які спостерігаються в даних. Наприклад, якщо підроблені дані генерує бот, який запрограмований на імітацію поведінки людини, то послідовність дій, які він виконує, може бути змодельована за допомогою прихованої моделі Маркова. НММ матиме набір прихованих станів, які представляють різні поведінки бота, а спостережувані дані будуть послідовністю дій, які виконує бот.

Щоб використовувати ПММ для підробленої ідентифікації, вам спочатку потрібно буде навчити модель на наборі даних із відомими підробленими та справжніми даними. Вона буде вивчати статистичні властивості кожного типу даних, дозволяючи розрізняти їх.

Після навчання ПММ також може використовуватися для класифікації нових даних як підроблених або справжніх. Для користувача залишається лише ввести послідовність спостережуваних функцій у модель, а ПММ виведе розподіл ймовірностей за можливими прихованими станами. На основі цього розподілу й можна визначити, чи є дані більш імовірно фейковими чи справжніми.

Одним з обмежень використання ПММ для підробленої ідентифікації є те, що вона може не вміти охопити всі нюанси даних. Наприклад, якщо підроблені дані генеруються складним алгоритмом, розробленим для імітації людської поведінки, може бути важко відрізнити їх від справжніх даних за допомогою лише прихованої моделі Маркова. У таких випадках можуть знадобитися додаткові методи, такі як глибоке навчання або виявлення аномалій.

Прихована модель Маркова також може бути використана для виявлення фейкових новинних статей. У цьому контексті ПММ буде використовуватися для моделювання послідовності слів у статті та матиме на меті відрізнити статті, які ймовірно є фейковими, від тих, які ймовірно є справжніми.

Обмеженням використання ПММ для ідентифікації фейкових новин є те, що вона може не вловити всі нюанси мови, яка використовується в статті. Наприклад, якщо фейкова стаття новин написана таким чином, що її неможливо відрізнити від справжньої статті, ПММ може бути важко її виявити. У таких випадках можуть знадобитися додаткові методи, такі як обробка природної мови та глибоке навчання.

Крім того, для ПММ припускають, що ймовірність спостереження певного слова залежить лише від поточного стану моделі. Це може бути нереалістичним припущенням для природної мови, оскільки ймовірність

спостереження слова може залежати від контексту слів навколо нього. Таким чином, більш складні моделі, такі як приховані напівмарковські моделі або рекурентні нейронні мережі, можуть бути більш ефективними для ідентифікації фейкових новин.

Переваги використання прихованої маркової моделі (ПММ) для ідентифікації фейків можуть включати наступні:

1. Моделювання послідовностей: ПММ можуть моделювати послідовності даних, такі як послідовність слів у новинах, тому вони можуть допомогти виявити певні шаблони та структури, які можуть бути пов'язані з фейками.

2. Підтримка статистичного аналізу: приховані моделі Маркова є статистичними моделями, які можуть робити передбачення на основі статистичних закономірностей у даних. Це може бути корисно для ідентифікації фейків, оскільки фейки можуть мати характеристики, які відрізняються від статистичних характеристик реальних даних.

3. Різноманітність застосування: ПММ можуть бути використані для моделювання різних видів даних, таких як текстові дані, відео, аудіо тощо. Це робить їх більш універсальними для використання в різних сферах, включаючи медіа, соціальні мережі тощо.

4. Моделювання залежностей: прихована модель Маркова може враховувати залежності між даними та їх взаємодії, що може допомогти виявити фейки, які використовують певні техніки, щоб обійти стандартні методи виявлення фейків.

5. Використання для аналізу масштабних даних: ПММ можуть бути ефективними для аналізу великого обсягу даних, що є цінним для виявлення фейків у масштабах, які можуть бути складними для аналізу вручну.

3.2.1 Рівні прихованої моделі Маркова в рамках завдання ідентифікації фейків

Приховані стани, які можуть використовуватися при ідентифікації фейків за допомогою прихованої маркової моделі (ПММ), можуть включати наступні:

1. Стани, що відображають різні типи новин: ПММ можуть мати приховані стани, які відображають різні типи новин, наприклад, політичні новини, наукові новини, спортивні новини тощо. Це допомагає ПММ відрізняти різні типи фейків та реальних новин.
2. Стани, що відображають рівні достовірності новин: ПММ можуть мати стани, які відображають різні рівні достовірності новин, такі як високі, середні та низькі. Це допомагає ПММ відрізняти реальні новини від фейків.
3. Стани, що відображають джерела новин: ПММ можуть мати стани, що відображають джерела новин, такі як різні видання новин, сайти, соціальні мережі тощо. Це допомагає ПММ відрізняти реальні новини від фейків, що можуть бути розповсюджені через певні джерела.
4. Стани, що відображають різні мови: ПММ можуть мати стани, що відображають різні мови, що використовуються у новинах. Це допомагає ПММ відрізняти реальні новини від фейків, що можуть бути написані на інших мовах або містити помилки перекладу.
5. Стани, що відображають різні теми: ПММ можуть мати стани, що відображають різні теми, які порушуються у новинах, такі як політика, спорт, наука, культура тощо. Це допомагає ПММ відрізняти реальні новини від фейкових.

Набором з прихованих станів для ідентифікації фейків можемо вважати набір з наступних значень множини станів {“інформація є правдою”, “інформація фейкова”, “невизначеність”}.

Для стану “інформація є правдивою” вважається, що інформація, яку перевіряється, є правдивою.

Для стану "інформація фейкова" інформація, яка перевіряється, є неправдивою, тобто фейковою.

Для стану “невизначеність” вважається, що інформацію не є можливо точно визначити як правдиву або фейкову з високим рівнем впевненості.

Окрім вже представлених станів, можуть бути ще й деякі додаткові стани, що відображають різні ступені впевненості в тому, що інформація є правдивою або неправдивлю. Наприклад, можуть бути такі стани як “висока ймовірність правдивості”, “низька ймовірність правдивості”, “висока ймовірність фейковості” і “низька ймовірність фейковості”.

У моделі Маркова для розпізнавання фейків можуть також бути використані різні параметри, такі як ймовірності переходів між станами, а також емісійні ймовірності, які відображають ймовірність вірного визначення фейку чи правди в кожному стані.

Тепер при наявності множини станів можемо розглянути й початкові ймовірності виникнення цих станів:

1. $\pi(1)$ = P(“інформація є правдою”);
2. $\pi(2)$ = P(“інформація фейкова”);
3. $\pi(3)$ = P(“невизначенісь”);

$\pi = [0.6, 0.3, 0.1]$

Запропонований початковий розподіл ймовірностей відображає попередні очікування щодо того, який стан є найбільш імовірним на початку аналізу інформації. В даному випадку, ймовірність перебування в стані "правдива інформація" є найбільшою (0.6), чим зображено те, що більша частина інформації, яка надійшла на перевірку є правдивою.

Для параметра кількості однозначно спостережуваних символів у станах виявлення фейкових новин не можна навести чіткого прикладу, оскільки він може варіюватися залежно від конкретного застосування та ознак, що використовуються для виявлення фейкових новин. Загалом, спостережувані символи в ПММ для виявлення фейкових новин можуть містити різні типи інформації, які дають змогу зробити висновки про те, чи є інформація

справжньою або фейковою, наприклад, опис новини, контент, зображення, аудіо, відео або метадані.

Наприклад, у системі виявлення фейкових новин, що базується на текстових даних, виявлені символи можуть бути окремими словами, фразами або реченнями з аналізованої статті. На відміну від цього, в системі виявлення фейкових новин на основі зображень виявлені символи можуть включати ознаки, властиві зображенню, такі як кольорові гістограми, особливості текстури або результати розпізнавання об'єктів, або інші особливості зображення, які дозволяють зробити висновок про те, чи було воно раніше змінено.

Загалом, вибір чітких символів залежатиме лише від конкретного застосування, типу даних, що аналізуються, та ознак, які є найбільш важливими для розрізнення справжньої та фальшивої інформації.

Тепер можемо розглянути приклад матриці ймовірностей переходів між станами.

За стани візьмемо: {правда, фейк, невизначеність}. Побудуємо таблицю 3.1 для кращого візуального уявлення:

Таблиця 3.1 - Матриця ймовірностей переходів між станами

	правда	фейк	невизначеність
правда	0.7	0.2	0.1
фейк	0.2	0.6	0.2
невизначеність	0.3	0.3	0.4

Дана матриця при проведені дослідження може бути використана як основа для пошуку найбільш ймовірного переходу між станами за певною послідовність чітких спостережуваних символів, що як вже описано раніше можуть змінюватися в залежності від поставленої задачі.

Матрицю розподілу ймовірності спостереження можна побудувати лише в рамках певної поставленої задачі. Вона буде побудована для вже відомою

матриці ймовірностей переходів між станами, але вже беручи до уваги символи виявлені при аналізі певної інформації і основуючись на ній ми можемо побачити найвірогіднішу послідовність станів. Тобто можемо виявити що правдива новина враховуючи певні символи й ознаки знайдені при її аналізі була модифікована й перейшла до стану фейковості.

Розглянувши всі рівні моделі прихованої моделі маркова можемо підсумувати всю отриману інформацію за допомогою таблиці 3.2 рівнів моделі:

Таблиця 3.2 - Опис рівнів моделі

Назва рівня	Опис для системи ідентифікації фейків
Початковий стан	Рівень представляє ймовірність початку в певному прихованому стані. У випадку розпізнавання підробки, ймовірності початкового стану можуть представляти ймовірність того, що зразок є справжнім або підробленим ще до того, як були зроблені будь-які спостереження.
Стан емісії	Рівень представляє ймовірність спостереження певної ознаки відповідно до прихованого стану. У випадку розпізнавання підробок, ймовірності емісії можуть представляти ймовірність спостереження певного набору рис обличчя за умови, що зразок є або справжнім, або підробленим або ймовірності зустрічі певних слів або висловлювань для текстових даних.
Перехідний стан	Даний рівень представляє ймовірності переходу від одного прихованого стану до іншого. У випадку розпізнавання підробки, ймовірності переходу можуть представляти перехід від справжнього стану до підробленого і навпаки.

Кінець таблиці 3.2

Прихований стан	Представляє певні базові стани, які генерують спостережувані ознаки, за допомогою яких може здійснюватись перехід між станами. У випадку розпізнавання підробок ці стани можуть відповідати тому, чи є зразок справжнім або підробленим.
Стан спостереження	Представляє спостережувані ознаки або характеристики даних, які використовуються для розрізнення справжніх і підроблених зразків.

Висновки до розділу 3

У даному розділі було розглянуто приховану модель Маркова, її переваги та недоліки. Також було описано план побудови моделі для задачі ідентифікації фейків, розроблено опис станів моделі на кожному з рівнів, що й було основною задачею у цьому розділі. Було роглянуто певні проблеми що можуть виникати під час використання прихованої моделі Маркова для ідентифікації фейків.

4 ОТРИМАННЯ ТА ОБРОБКА ДАНИХ ДЛЯ НАВЧАННЯ ПММ

4.1 Збір даних

Одним з найважливіших моментів під час розробки засобу для ефективної ідентифікації фейків за допомогою машинного навчання, а в даній роботі прихованої моделі Маркова є збір даних. Від правильно підібраної та розмежованої збірки даних для навчання моделі залежить її майбутня робота, а саме її правильність та результативність, тому треба приділити неабияку увагу до цього процесу.

В межах даної роботи опишемо процес отримання даних для навчання прихованої моделі Маркова.

Збір даних може включати наступні етапи: першочергово потрібно зібрати набір даних, який містить вибірку як справжніх, так і фейкові новини. Це можуть бути тексти новин, статті, пости в соціальних мережах тощо. Важливо забезпечити різноманітність даних, щоб модель могла навчитися розпізнавати різні типи фейкових новин.

Для навчання нашої моделі будуть надаватись варіанти правдивих новин і відповідних сфальсифікованих новин, щоб модель базуючись на одному контексті новини вдало могла розрізнити правду від брехні. Наступним етапом можемо виділити маркування даних, бо для того щоб побудувати вдалу приховану модель Маркова, необхідно мати марковані дані, тобто позначити кожну новину як справжню або фейкову. Маркування може здійснюватися вручну експертами, які аналізують кожну новину та визначають її правдивість. Цей процес може бути тривалим і трудомістким, але він необхідний для навчання моделі. Завершальним та не менш важливим етапом є попередня обробка даних: перш ніж навчати модель машинного навчання, потрібно попередньо обробити дані. Це може включати видалення непотрібних символів, нормалізацію тексту, токенізацію, видалення стоп-слів тощо. Це допомагає поліпшити якість даних і продуктивність моделі. Отже ми описали процеси

збору інформації покроково, але що саме мають містити в собі як і правдиві так і фейкові статті, чим вони мають відрізнятися і як полегшити розпізнавання даних моделлю?

Фейкові новини досить часто є результатом вдалої гри слів, спрямованої на те щоб привернути увагу читача і зобразити новину досить таки правдоподібної. Частіше за все алгоритми для виявлення фейкових новин вміють розпізнають кожне слово як ознаку, але для цих методів є й певні недоліки, а саме те, що вони не здатні розпізнати гру слів притаманну певній мові, бо знаючи мову й її структуру можна будувати певні речення, дослівний перевід яких не нестиме того самого сенсу що несе речення. Також вони не надають жодного значення й структурі речення, а різний порядок слів й розділових знаків може надавати новині й різного смислового забарвлення.

Розглянемо для прикладу наступні реальну та неправдиву новини:

S_1: "Ціни на оренду житла у Києві зросли на 7% з настанням літнього сезону, хоча при цьому не спостерігається великого припливу людей до столиці".

S_2: "Ціни на оренду житла у Києві знизились на 7% з настанням літнього сезону, хоча при цьому спостерігається великий приплив людей до столиці".

Новина S_1 є правдивою, натомість S_2 є фейком, але розглядаючи ці два речення з застосуванням TF-IDF-представлення, а саме числової статистики, яка використовується в обробці природної мови та пошуку інформації для визначення важливості певного терміна в документі або в колекції документів, речень S_1 і S_2 будуть абсолютно ідентичними. Керуючись цим фактом, можемо стверджувати, що алгоритм машинного навчання, який використовує такі представлення речень для навчання моделі, не зможе розрізнити їх. Такі випадки зумовлюють потребу в моделі машинного навчання, яка не тільки має високу точність, але й враховує вагу структури речення.

Саме ланцюги Маркова надають великого значення структурі речення. А сама прихована модель Маркова працює на припущенні, що майбутній стан залежить лише від поточного стану, а не від будь-якого стану до нього. Кожен

вузол ланцюга Маркова направляє до набору слів, які зустрічалися після нього в навчальному наборі даних. Така побудова моделі полегшує виявлення можливої гри слів.

Традиційно вважається, що схеми класифікації на основі прихованих моделей Маркова функціонують шляхом навчання однієї прихованої моделі для кожного із класів окремо, а потім використовуються для обчислення щільності умовних класів за допомогою парадигми класифікації Байєса. Потім послідовність класифікується на тому, яка з моделей Маркова, що базується на класах, найімовірніше, згенерувала б цю послідовність. Таке правило відоме як правило класифікації з максимальною правдоподібністю.

Головної проблемою такої класифікації є обмежена здатність моделі враховувати лише частоти слів, а не фактичний порядок слів, тому у цій роботі буде запропонована схема що враховує структуру речення, а не лише частоти слів.

Експериментальна частина цієї роботи є виконаною на наборі даних FakeNewsNet, який містить у собі новини та фейки написані англійською та наборі даних фейкових новин зібраних та оброблених самостійно, а саме новин які написані українською мовою і включають певні особливості мови. Це було зроблено з метою того щоб порівняти ефективність навчання моделі на різних даних та виявити головні відмінності навчання моделі на датасетах, які складаються з даних написаних різними мовами.

Отже, як вже зазначалось раніше, для отримання гарної віддачі від моделі після машинного навчання, дані, на яких модель навчається мають бути грамотно розділені та оброблені, тож розглянемо детальніше набори даних, використані в межах даної роботи.

По-перше, слід відмітити що як англійський так і український датасети використовуються у форматі csv, цей формат є одним із найзручніших та найпоширеніших форматів збереження наборів даних для машинного навчання.

Розглянемо які ще переваги є у csv формату:

- Легко зберігати та читати: csv-файл - це текстовий файл, в якому дані розділені комами або деякими іншими розділювачами, такими як крапка з комою або табуляція. Це дуже спрощує зберігання та обробку даних, тому більша кількість програм для обробки даних і машинного навчання можуть без проблем читати і записувати дані.
- Універсальність: csv можна використовувати в будь-якій операційній системі і в різних програмах для обробки даних і машинного навчання. Це дає більше гнучкості при роботі з наборами даних.
- Підтримка структурованих даних: дозволяє легко представляти структуровані дані у вигляді таблиці з рядками і стовпчиками. Кожен рядок представляє окремий запис або вибірку, а стовпці відповідають різним атрибутам або характеристикам даних. Це зручно для роботи з табличними даними, що часто зустрічається в задачах машинного навчання. Саме ця властивість і буде використана у роботі для представлення нашого набору з фейковими та реальними новинами.
- Сумісність з іншими програмами: Багато програм для роботи з даними та машинного навчання підтримують імпорт та експорт даних у форматі csv. Це означає, що є можливість легко обмінюватися даними між різними інструментами та бібліотеками без проблем із сумісністю.
- Компактність: csv зазвичай має компактний розмір файлу в порівнянні з іншими форматами, такими як Excel або JSON. Це може бути важливим фактором, особливо якщо ви працюєте з великими наборами даних.

Загалом, формат CSV простий у використанні, універсальний, зручний для роботи зі структурованими даними та сумісний з багатьма іншими програмами. Це робить його популярним вибором для зберігання наборів даних у машинному навчанні, тому саме його було обрано для представлення датасету у межах цієї роботі.

4.1.1 Вибір набору даних для навчання

За останні кілька років, коли поширення фейкових новин в Інтернеті стало дедалі більшою проблемою, з'явилася інформація про те, що багато загальнодоступних наборів даних містять новинні статті, твіти або дописи з відповідною позначкою про їхню легітимність.

1. FakeNewsNet: Набір даних політичних твітів і пліток

Цей набір даних використовує офіційний API Twitter для отримання твітів від користувачів Twitter, включаючи метадані та соціальний контекст. Але для початку вони надають значну кількість чистих даних, які можна легко використовувати для тестування моделей, що і буде зроблено в межах даної дипломної роботи.

2. Корпус фейкових новин

Цей набір даних з відкритим вихідним кодом містить мільйони новинних статей, витягнутих з курованого списку зі 1001 домену. Набір даних включає понад 9 400 000 статей з більш ніж 700 доменів, вилучених з декількох доменів, таких як NYTimes і WebHose English News Articles.

3. FakeHealth

FakeHealth збирається для вирішення проблем виявлення фейкових новин про здоров'я з таких джерел, як новинний контент, огляди та соціальні мережі. Як пропонується в статті "Імбир не може вилікувати рак: Боротьба з фейковими новинами про здоров'я за допомогою комплексного сховища даних", дані можуть бути отримані з Twitter за допомогою Twitter API та ідентифікаторів, наданих у репозиторії GitHub - EnyanDai/FakeHealth. Зрештою, набір даних містить дані з 500 тис. твітів, 29 тис. відповідей, 14 тис. ретвітів і 27 тис. профілів користувачів з хронологією та списками друзів. Огляди містять пояснення щодо десяти критеріїв оцінки новин про здоров'я.

4. Набір даних про фейкові новини, спричинені COVID-19

Цей набір даних містить пости в соціальних мережах, пов'язані з COVID-19 і вакцинацією, з різних популярних платформ. Пости анотовані як справжні

та фейкові новини, і набір даних можна краще зрозуміти в документі "Боротьба з інфодемією: COVID-19 Fake News Dataset", а дані, а також запропонований базовий бенчмарк з використанням моделей уваги та евристичної постобробки можна знайти в репозиторії GitHub - COVID Fake News.

5. FNC-1 (Fake News Challenge Stage 1)

FNC-1 був розроблений як набір даних для виявлення позиції, що містить 75 385 маркованих пар заголовків і статей. Ці пари є або згодними, або незгодними, або дискусійними, або не пов'язаними між собою. Крім того, базова лінія оцінює позицію тексту з новинної статті щодо заголовка. Зокрема, текст може погоджуватися, не погоджуватися, обговорюватися або бути не пов'язаним із заголовком. Набір даних FNC-1 і пропонована базова лінія використовують ці переваги.

4.2 Інструменти для обробки та роботи з даними

Python: під час реалізації алгоритмів машинного навчання моделі Маркова, Python виявився кращою мовою у порівнянні зі своїми аналогами. У цій дипломній роботі ми розглянемо причини домінування Python у машинному навчанні ланцюгів Маркова та підкреслимо його переваги над іншими мовами.

Бібліотеки та пакети Python: розгалужена екосистема бібліотек та пакетів для машинного навчання є однією з головних переваг Python. Мова пропонує надійні бібліотеки, такі як TensorFlow, Keras та PyTorch, які надають потужні інструменти для реалізації алгоритмів ланцюгів Маркова. Ці бібліотеки мають вбудовані функції для генерації та маніпулювання моделями ланцюгів Маркова, що спрощує процес розробки. Крім того, екосистема бібліотек Python заохочує співпрацю та обмін знаннями, оскільки користувачі можуть легко отримати доступ до існуючих реалізацій з відкритим вихідним кодом та зробити свій внесок у них.

Простота та читабельність: Python славиться своєю простотою та зручністю для читання, що робить її ідеальною мовою для розробки моделей машинного навчання на основі ланцюгів Маркова. Її чистий синтаксис і проста структура коду дозволяють розробникам стисло висловлювати складні ідеї. Ця простота не тільки скорочує час розробки, але й робить код легшим для розуміння та підтримки. Алгоритми ланцюгів Маркова часто передбачають складні обчислення та переходи, а інтуїтивно зрозумілий синтаксис Python полегшує реалізацію таких моделей, підвищуючи продуктивність та зменшуючи ймовірність помилок.

Легка інтеграція з обробкою та візуалізацією даних: Ще однією значною перевагою Python є його безшовна інтеграція з інструментами обробки та візуалізації даних. Python надає потужні бібліотеки, такі як NumPy та Pandas, які полегшують маніпуляції з даними та їх аналіз, дозволяючи фахівцям без особливих зусиль попередньо обробляти та трансформувати свої набори даних. Крім того, бібліотеки візуалізації Python, такі як Matplotlib та Seaborn, дозволяють користувачам створювати наочні візуальні представлення ланцюгів Маркова та їх результатів. Можливість безперешкодної інтеграції обробки даних і візуалізації спрощує аналіз моделей ланцюгів Маркова, допомагаючи в інтерпретації та передачі результатів.

Підтримка: Python може похвалитися активною та підтримуючою спільнотою, яка відіграє важливу роль у його домінуванні в галузі машинного навчання. Спільнота пропонує широкий спектр ресурсів, включаючи навчальні посібники, документацію та форуми, що полегшує новачкам розуміння концепцій машинного навчання за допомогою ланцюга Маркова та усунення будь-яких проблем, з якими вони стикаються. Крім того, колективні знання спільноти постійно сприяють вдосконаленню та розвитку бібліотек та інструментів Python, гарантуючи користувачам доступ до найсучасніших функцій та методів.

Крос-доменна застосовність: універсальність Python та широке застосування в різних галузях роблять його чудовим вибором для машинного

навчання за допомогою ланцюгів Маркова, оскільки ці алгоритми часто знаходять застосування в різноманітних сферах. Будь то фінанси, охорона здоров'я, обробка природної мови або соціальні науки, універсальність Python дозволяє дослідникам і практикам легко впроваджувати моделі ланцюгів Маркова у своїх галузях. Ця міждоменна застосовність ще більше зміцнює позицію Python як найкращої мови для машинного навчання за допомогою ланцюгів Маркова.

Враховуючи широкий спектр бібліотек Python, простоту та зручність для читання, безперешкодну інтеграцію з обробкою та візуалізацією даних, потужну підтримку спільноти та крос-доменну застосовність, стає очевидним, що Python перевершує інші мови, коли мова йде про реалізацію алгоритмів машинного навчання ланцюга Маркова. Переваги Python з точки зору продуктивності, читабельності коду та доступу до ресурсів роблять його мовою вибору для дослідників та практиків у цій галузі. Таким чином, для тих, хто прагне досягти успіху в машинному навчанні за методом ланцюга Маркова, Python є безперечно найкращим варіантом.

4.3 Обробка природної мови

Розглянемо бібліотеки Python, що використовуються для задач (NLP).

NLTK (Natural Language Toolkit): це комплексна бібліотека для завдань NLP в Python. Вона надає широкий спектр функціональних можливостей та інструментів для таких завдань, як токенізація, стеммінг, лематизація, тегування частин мови, розпізнавання іменованих сутностей, аналіз настроїв тощо. NLTK також включає різні корпуси, лексичні ресурси та попередньо навчені моделі, які можуть бути корисними для навчання та оцінювання моделей NLP. Він широко використовується дослідниками та практиками у спільноті NLP завдяки своїй універсальності та обширній документації.

spaCy: одна потужна та ефективна бібліотека для НЛП на Python. Вона зосереджена на створенні швидких і готових до використання конвеєрів NLP. spaCy пропонує такі функції, як токенізація, тегування частин мови, синтаксичний аналіз залежностей, розпізнавання іменованих сутностей, сегментація речень та багато іншого. Він також включає в себе попередньо навчені моделі для декількох мов, що дозволяє користувачам виконувати завдання NLP "з коробки". spaCy відомий своєю швидкістю та продуктивністю, що робить його чудовим вибором для великомасштабних додатків NLP.

Порівняємо ці бібліотеки:

Функціональність:

NLTK: надає повний набір інструментів і ресурсів для завдань NLP. Він охоплює широкий спектр функцій, включаючи токенізацію, стеммінг, лематизацію, тегування частин мови, синтаксичний розбір, розпізнавання іменованих сутностей, аналіз настроїв та багато іншого. NLTK також включає різні лексичні ресурси.

spaCy: пропонує спрощений та ефективний підхід до NLP. Він надає такі функції, як токенізація, тегування частин мови, синтаксичний аналіз залежностей, розпізнавання іменованих об'єктів, сегментація речень тощо. spaCy фокусується на високопродуктивних та оптимізованих конвеєрах обробки.

Продуктивність:

NLTK: не завжди може бути найпродуктивнішим варіантом, особливо для великомасштабних додатків або додатків у режимі реального часу. Вона може мати обмеження щодо швидкості та використання пам'яті.

spaCy: відома своєю швидкістю та ефективністю. Він призначений для створення швидких і готових до виробництва конвеєрів NLP, що робить його підходящим вибором для роботи в режимі реального часу або великомасштабних додатків. Оптимізована обробка spaCy може значно підвищити продуктивність порівняно з NLTK в певних сценаріях.

Попередньо навчені моделі:

NLTK: включає в себе різні попередньо навчені моделі, але діапазон і покриття може бути не таким широким, як у spaCy. Проте NLTK надає широкий вибір корпусів, лексичних ресурсів і наборів даних, які можуть бути корисними для дослідницьких і навчальних цілей.

spaCy: постачається з попередньо навченими моделями для кількох мов, що охоплюють такі завдання, як тегування частин мови, розпізнавання іменованих сутностей та синтаксичний аналіз залежностей. Попередньо навчені моделі в spaCy часто хвалять за їхню продуктивність і точність.

Простота використання:

NLTK: має м'яку криву навчання і надає обширну документацію, що робить його зручним для початківців. Він пропонує ряд навчальних посібників та прикладів, які допоможуть користувачам розпочати роботу та зрозуміти концепції НЛП.

spaCy: spaCy також зручна для користувача, з добре задокументованими API та корисними навчальними посібниками. Він має більш впорядкований та інтуїтивно зрозумілий дизайн API, що може спростити процес впровадження для розробників.

Підтримка:

NLTK: має велику та активну спільноту з багатою екосистемою. Вона існує вже довгий час, і, як наслідок, існує безліч ресурсів, навчальних посібників та розширень, створених спільнотою.

spaCy: набув значної популярності в останні роки і має зростаючу спільноту. Хоча екосистема може бути не такою розгалуженою, як у NLTK, spaCy виграє завдяки своїй орієнтованості на ефективність та продуктивність, що приваблює віддану базу користувачів.

В данному дипломі подалі будемо використовути spaCy – в неї значна простота використання. Але головним фактором буде саме підтримка української мови в порівнянні з NTLK.

Для обробки тексту будуть використані такі кроки рис 4.1., подалі кожен крок буде описано детально.

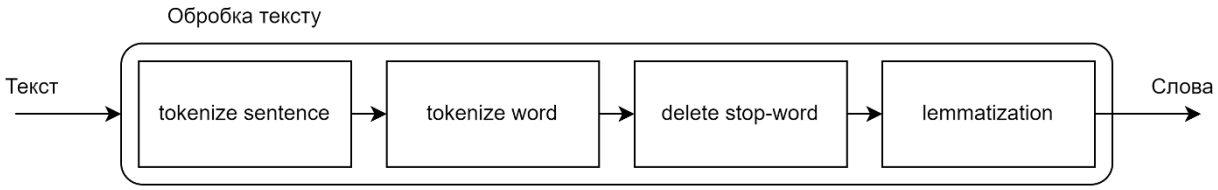


Рисунок 4.1 – Алгоритм обробки тексту

4.3.1 Токенезація за реченням

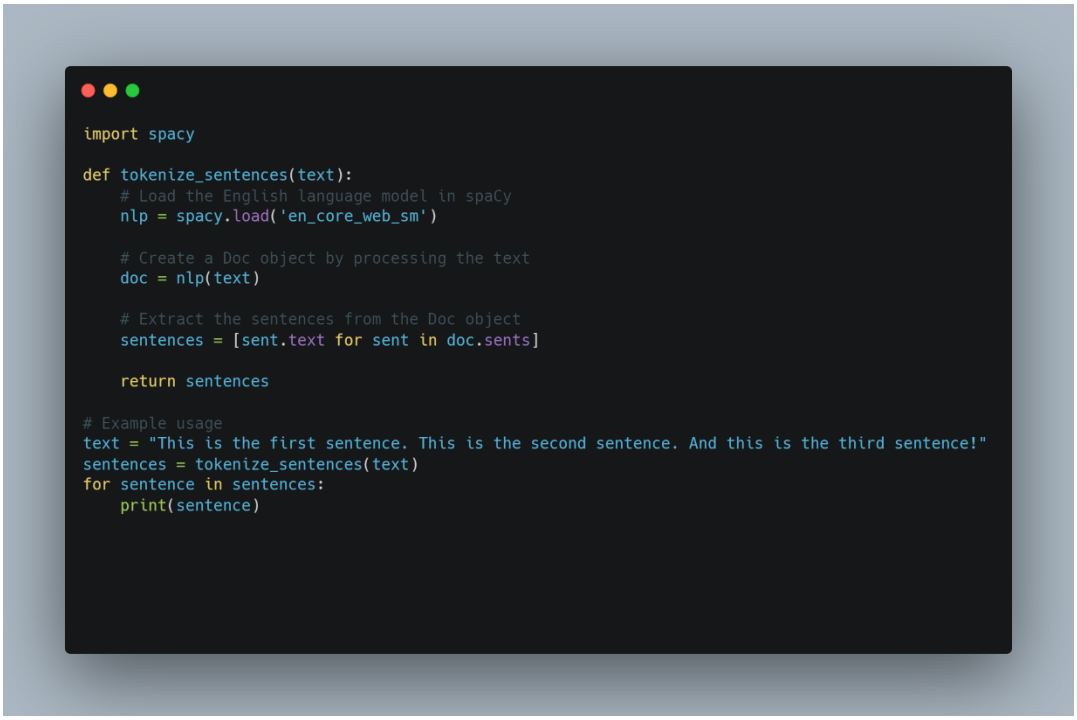
У галузі обробки природної мови (NLP) токенізація речень відіграє життєво важливу роль у розбитті текстових даних на окремі речення. У цьому дипломному описі основна увага приділяється важливості токенізації речень і надається огляд методів та інструментів, які зазвичай використовуються для виконання цього завдання.

Визначення та значення:

Токенізація речень, також відома як сегментація речень або виявлення меж речень, - це процес поділу тексту на окремі речення. Це фундаментальний крок у таких завданнях NLP, як аналіз тексту, вилучення інформації, машинний переклад та аналіз настроїв. Токенізація речень допомагає витягувати значення на рівні речень, що дозволяє проводити більш детальний аналіз і полегшує подальші завдання NLP.

Токенізація на основі правил:

Одним із поширених підходів до токенізації речень є токенізація на основі правил рис. 4.1. Ця методика спирається на заздалегідь визначені правила, шаблони та евристики для визначення меж речень. Ці правила часто розглядають розділові знаки (наприклад, крапки, знаки питання та знаки оклику) як індикатори кінця речення. Однак обробка винятків, таких як абревіатури або вкладені речення, може бути складним завданням за допомогою лише методів, заснованих на правилах.



```python
import spacy

def tokenize_sentences(text):
    # Load the English language model in spaCy
    nlp = spacy.load('en_core_web_sm')

    # Create a Doc object by processing the text
    doc = nlp(text)

    # Extract the sentences from the Doc object
    sentences = [sent.text for sent in doc.sents]

    return sentences

# Example usage
text = "This is the first sentence. This is the second sentence. And this is the third sentence!"
sentences = tokenize_sentences(text)
for sentence in sentences:
    print(sentence)
```

Рисунок 4.2 – Токенізація за реченням. Приклад коду

Він розбиває текст на окремі речення:

This is the first sentence.

This is the second sentence.

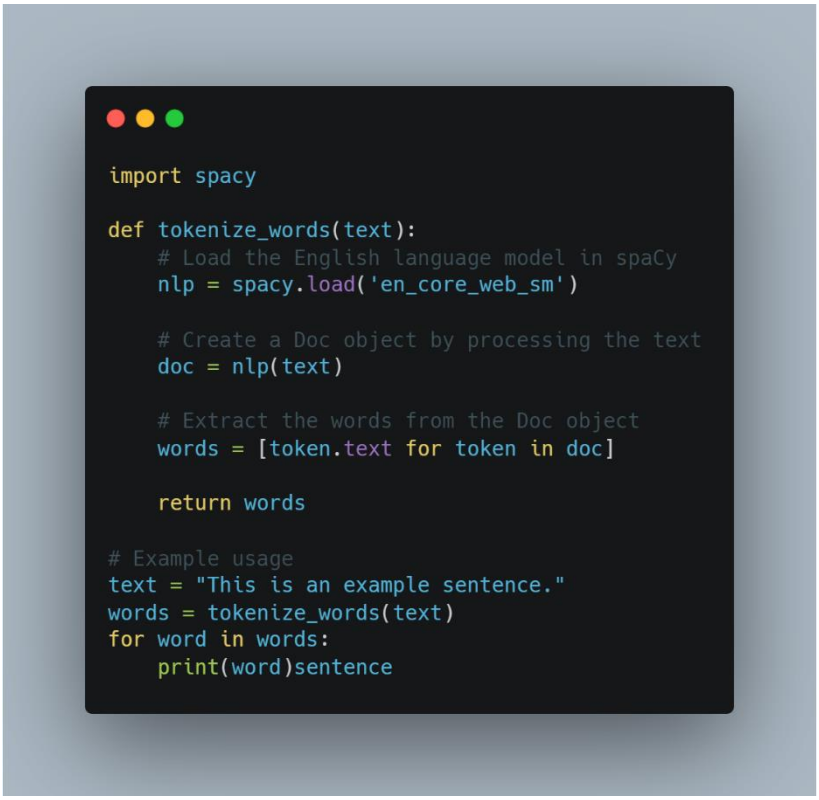
And this is the third sentence!

4.3.2 Токенезація слів

Токенізація слів, також відома як сегментація слів або виявлення меж слів, - це процес поділу тексту або документа на окремі слова. Він відіграє важливу роль у різних завданнях NLP, таких як попередня обробка тексту, пошук інформації, маркування частин мови, аналіз настрою та машинний переклад. Токенізація слів створює основу для подальшого аналізу, розбиваючи текст на значущі одиниці.

Одним із поширених підходів до токенізації слів є токенізація на основі правил рис. 4.2. Ця техніка спирається на заздалегідь визначені правила і шаблони для визначення меж слів. Як індикатори меж слів часто

використовуються звичайні роздільники, такі як пробіли, розділові знаки та символи. Токенізація на основі правил може бути ефективною в багатьох випадках, але може стикатися з певними складнощами, такими як обробка скорочень, складних слів або слів з дефісами.



```python
import spacy

def tokenize_words(text):
    # Load the English language model in spaCy
    nlp = spacy.load('en_core_web_sm')

    # Create a Doc object by processing the text
    doc = nlp(text)

    # Extract the words from the Doc object
    words = [token.text for token in doc]

    return words

# Example usage
text = "This is an example sentence."
words = tokenize_words(text)
for word in words:
    print(word)sentence
```

Рисунок 4.3 – Токенізація слів. Приклад коду

4.3.3 Лематизація та стеммінг

Лематизація та стиммінг є важливими техніками для нормалізації тексту.

Лематизація - це процес приведення слів до їхньої базової або словникової форми, відомої як лема рис. 4.3. Він має на меті перетворити різні відмінювані форми слова на загальну базову форму, що полегшує точний аналіз і розуміння тексту. При лематизації враховується контекст і частина мови (ЧМ) слова, щоб визначити відповідну лему. Наприклад, лемою слова "біг" є "бігти", а лемою слова "кращий" - "хороший".

```python
import spacy

def lemmatize_text(text):
    # Load the English language model in spaCy
    nlp = spacy.load('en_core_web_sm')

    # Create a Doc object by processing the text
    doc = nlp(text)

    # Extract the lemmas from the Doc object
    lemmas = [token.lemma_ for token in doc]

    return lemmas

# Example usage
text = "The quick brown foxes jumped over the lazy dogs"
lemmas = lemmatize_text(text)
print(lemmas)
```

Рисунок 4.4 – Лематизація та стеммінг. Приклад коду

Вивід:

['the', 'quick', 'brown', 'fox', 'jump', 'over', 'the', 'lazy', 'dog']

Стеммінг - це простіший і більш евристичний підхід до нормалізації тексту. Він передбачає приведення слів до їхньої кореневої або основної форми шляхом видалення суфіксів або префіксів. Отримана основа може не завжди бути правильним словом, але вона зберігає основне значення вихідного слова. Методи усічення, як правило, дотримуються правил та евристик для виконання усічення і можуть не враховувати контекст або місцезнаходження слова. Наприклад, усічення слова "біг" призведе до утворення "біг", а усічення слова "краще" - до утворення "добре".

Відмінності:

Точність: лексифікація створює основну форму слова, забезпечуючи більш точне представлення. Основоскладання, з іншого боку, може призвести до утворення основ, які не є повнозначними словами або можуть мати інше значення.

Контекст і місцезнаходження: при лематизації враховується контекст і місцезнаходження слова, що забезпечує отримання правильної леми. Стеммінг, навпаки, зазвичай застосовує правила і шаблони без урахування цих факторів.

Результат: Лематизація створює леми, які є справжніми словами, що робить її більш придатною для завдань, які вимагають осмисленого і граматично правильного тексту. Стеммінг генерує основи, які не завжди можуть бути повноцінними словами.

Складність: лематизація передбачає складніший лінгвістичний аналіз, включно з POS-тегуванням і морфологічним аналізом, що робить її більш трудомісткою порівняно зі стеммінгом.

Використання:

Пошук інформації: обидва методи можна використовувати для скорочення слів до їхніх базових форм, покращуючи точність пошуку та вилучення інформації.

Класифікація тексту: лемматизація та стеммінг можуть допомогти зменшити варіації слів та покращити вилучення ознак для задач класифікації тексту.

Сентиментальний аналіз: нормалізація слів до їхніх базових форм може покращити аналіз настрою, вловлюючи суть слів і зменшуючи при цьому рівень шуму.

Машинний переклад: лематизація та стеммінг можуть допомогти спростити словниковий запас і зменшити кількість словоформ під час перекладу.

4.3.4 Стоп слова

Стоп-слова - це загальновживані слова, які часто зустрічаються в мові, але не несуть жодного або майже жодного суттєвого значення рис 4.4. Такі слова, як "the", "and", "is" і "of", не роблять значного внеску в загальну семантику або розуміння тексту. У багатьох завданнях NLP видалення стоп-

слів може підвищити ефективність алгоритмів, зменшити шум і підвищити точність подальшого аналізу.

Мета та переваги видалення стоп-слів

Покращена обчислювальна ефективність: видаляючи слова, що часто зустрічаються, але не мають значення, можна значно зменшити час обробки та вимоги до пам'яті для алгоритмів NLP.

Зменшення шуму: стоп-слова часто додають шуму до аналізу тексту, вносячи непотрібні варіації та відволікаючі фактори. Видалення їх може покращити якість і чіткість оброблених даних.

Посилений фокус на змістовному контенті: видалення стоп-слів дозволяє алгоритмам НЛП сконцентруватися на більш інформативних і релевантних словах, що призводить до кращого вилучення та аналізу ознак.

Зменшення обсягу словника: Видалення стоп-слів зменшує розмір словника, що може бути особливо корисним у таких завданнях, як класифікація текстів, пошук інформації та тематичне моделювання.

```python
import spacy

def remove_stop_words(text):
    # Load the English language model in spaCy
    nlp = spacy.load('en_core_web_sm')

    # Process the text using spaCy
    doc = nlp(text)

    # Remove stop words from the text
    filtered_words = [token.text for token in doc if not token.is_stop]

    # Join the filtered words back into a single string
    filtered_text = ' '.join(filtered_words)

    return filtered_text

# Example usage
text = "This is an example sentence demonstrating the removal of stop words."
filtered_text = remove_stop_words(text)
print(filtered_text)
```

Рисунок 4.5 – Стоп-слова. Приклад коду

Висновки до розділу 4

У цьому розділі розглянуто та обгрунтовано обрання даних для навчання моделі, описаний формат зберігання цих даних. Також описано процес обробки природньої мови для підготовки її для навчання моделі та мову програмування за домогою якої оброблювались дані та створювала прихована модель Маркова, також зроблений опис використаних для цього бібліотек.

5 АНАЛІЗ РЕЗУЛЬТАТІВ

Після обробки тексту і створення матриці переходів для заданого тексту, матриця складається зі слів з ймовірностями яке може бути наступне слово. Основним результатом є найбільше значення по індексу вхідного тексту.

Результати

Для представлення результатів були використані діаграми Cloud Word, які показують які і наскільки часто зустрічаються слова у сфальсифікованих (рис 5.1) та справжніх (рис 5.2) новинах для датасету українською мовою.

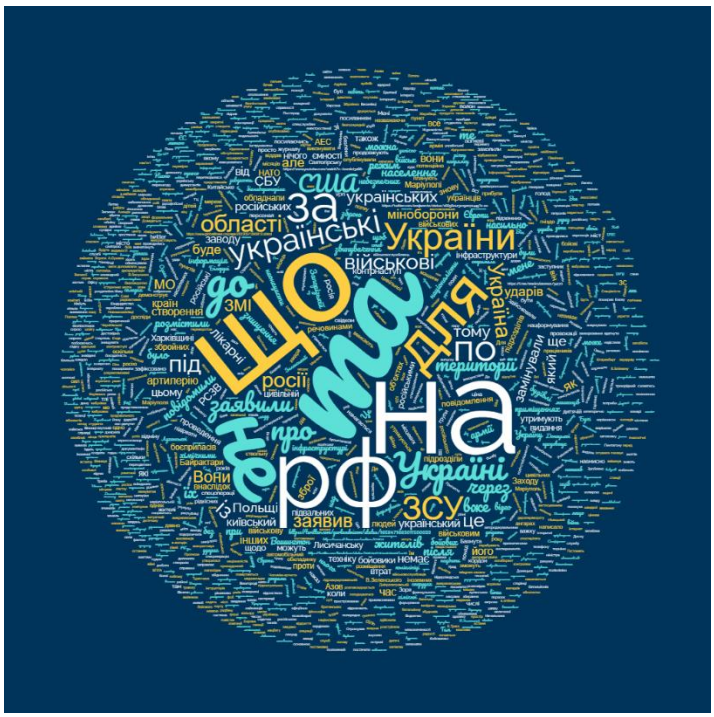


Рисунок 5.1 – Частота зустрічі слів у фейкових новинах (український датасет)

Для перевірки візьмемо новину, що не була в данних для тренування: *«у Кременчуці не було жодного теракту ... армія рф не б'є ні по жодних цивільних об'єктах під час спецоперації».*

Рисунок 5.2 - Частота зустрічі слів у справжніх новинах (український датасет)

Модель початково було натренована на 200 хибних та фейкових даних. Вивід показав, що дана новина є фейком, що є вірно. Для тестування було взято 40 (20 правильних, 20 хибних), результати тестування можна побачити на графіку рис 5.3.

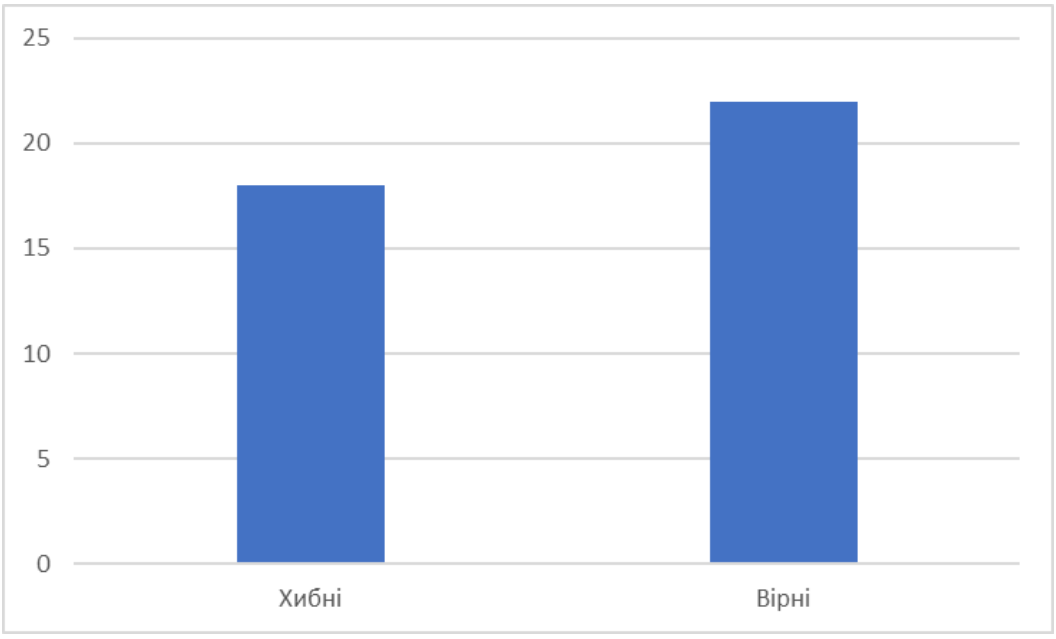


Рисунок 5.3 – Результати визначення фейкової новини (український датасет)

Аналогічно для FakeNewsNet датасету були використані діаграми Cloud Word, які тепер показали які і наскільки часто зустрічаються слова вже у сфальсифікованих (рис 5.4) та справжніх (рис 5.5) новинах для датасету англійською мовою, що був також й більш обширним. Тестовий сет складався приблизно з 4000 новин.

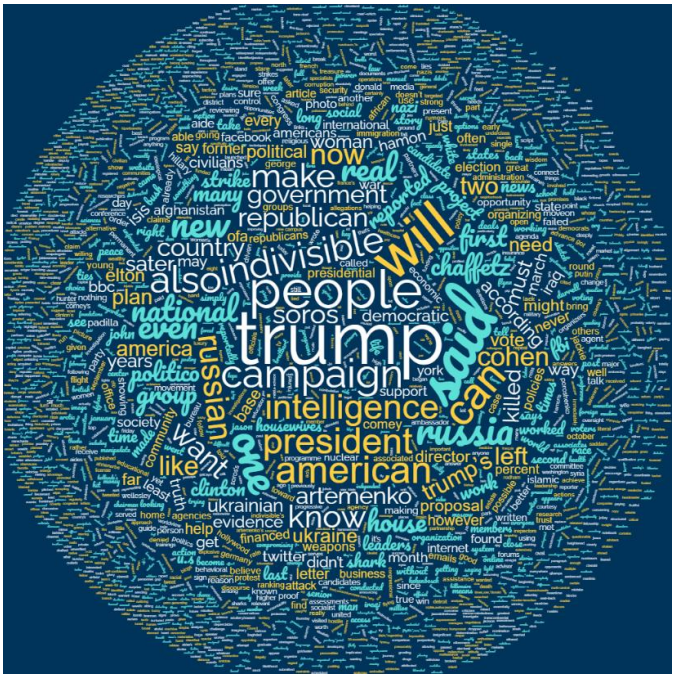


Рисунок 5.4 – Частота зустрічі слів у фейкових новинах (англійський датасет)

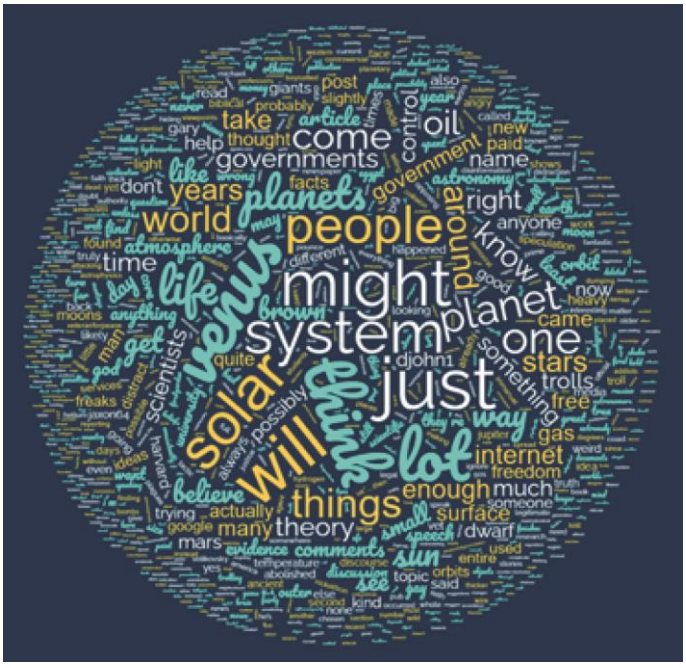


Рисунок 5.5 – Частота зустрічі слів у справжніх новинах (англійський датасет)

Датасет англійською мовою для перевірки складався з більше ніж з 1000 новин. І були отримані наступні результати рис 5.6.

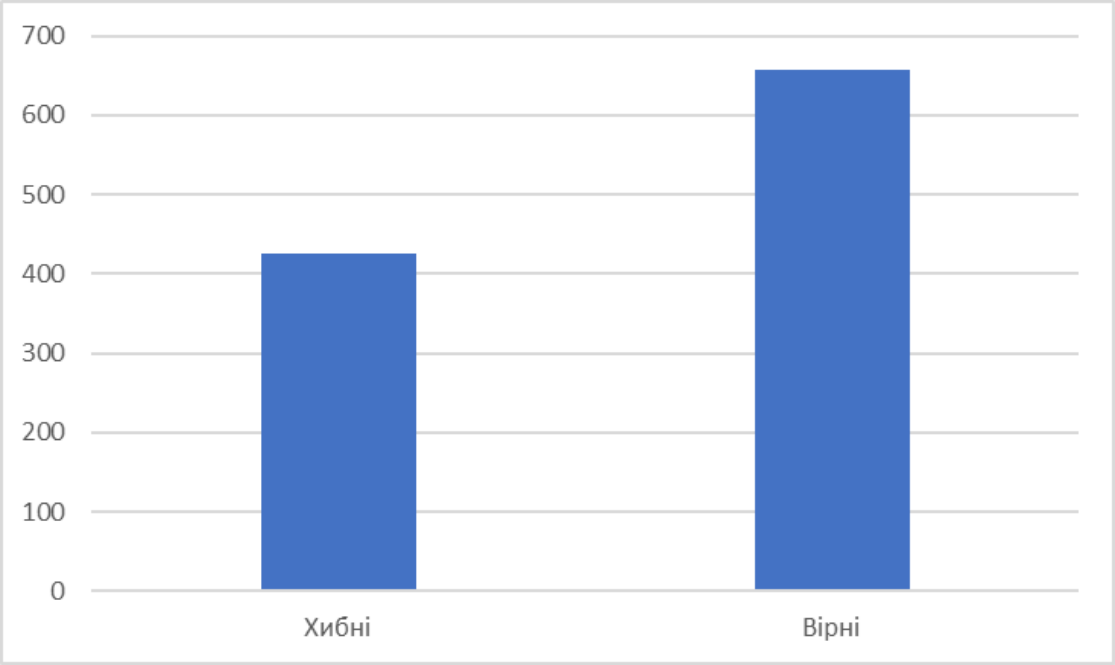


Рисунок 5.6 – Результати визначення фейкової новини (англійський датасет)

Висновки до розділу 5

У цьому розділі були описані результати проведеного дослідження для англійського та українського датасету з фейковими та справжніми новинами, побудовані графіки.

В результаті дослідження маємо, що для малої вибірки даних модель видає успішний результат лише в половині випадків, відношення вірних до хибних складає 54%. В той же час модель яка натренована на великій виборці із фейкових нових відношення складає вже 68%. Отже, маємо досить непоганий результат для вибірки даних уранською мовою що є значно меншою.

ВИСНОВКИ

В результаті виконання даної дипломної роботи було розглянуто поняття фейкових новин та проведенна їх класифікація. Також було описано методи виявлення сфальсифікованої інформації та проведення аналізу отриманих даних взагалому.

Було розглянути приховану модель Маркова як інструмент для ідентифікації фейкових нових, його переваги та недоліки у межах вирішення даної задачі. Також дану модель було реалізовано за допомогою мови програмування Python з використанням бібліотеки spaCy, описано всі рівні моделі.

Для навчання прихованої моделі маркова було розроблено набір даних, який складався з приблизно 200 правдивих та фейкових новин українською мовою. Також окремо була взята готова вибірка новин англійською мовою (приблизно 4000 правдивих та сфальсифікованих новин) для порівняння результатів.

Проведено начання моделі на тренувальних даних та дослідження, яке показало що експеримент проведений на українських даних надавав у 54% вірний результат, що є досить непогано для малого набору, бо для порівняння набір на даних англійською мовою надав результат у 68% правильних відповідей, хоча й вибірка було у 20 разів більшою.

Отже, розглядаючи отримані результати, рекомендується продовжувати дослідження в цьому напрямку для поліпшення результатів дослідження. У подальшому слід розширювати тренувальний набір даних українською мовою для збільшення відсотку правильної ідентифікації новин як фейкових або правдивих.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. Шпигунство, фейки, провокації і ЦРУ [Електронний ресурс] – Режим доступу: https://www.radiosvoboda.org/a/29548546.html (дата звернення 10.01.2022)
2. Афера Стивена Гласса [Електронний ресурс] – Режим доступу: https://cripo.com.ua/processes/?p=184105/ (дата звернення 20.01.2022) The ONION [Електронний ресурс] – Режим доступу: https://www.theonion.com/ (дата звернення 25.01.2022)
3. Spotting 'fake news' among the real stories [Електронний ресурс] – Режим доступу: https://www.bbc.com/news/av/education-43404254 (дата звернення 30.01.2022)
4. David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Penny-cook, David Rothschild, et al. The science of fake news. Science. 2018. – сторінки 1094–1096.
5. "Is 'fake news' a fake problem?". [Електронний ресурс] – Режим доступу: https://www.cjr.org/analysis/fake-news-facebook-audience-drudge-breitbart-study.php (дата звернення 07.04.2023)
6. Universita Pendidikan Indonesia. Hidden Markov Model. [Електронний ресурс] – Режим доступу: areasearch.upi.edu/operator/upload/ta_mtk_0706664_chapter3.pdf. (дата звернення 25.04.2023)
7. Miroslav Goljan, Jessica Fridrich, and Mo Chen. Defending against fingerprint-copy attack in sensor-based camera identification. IEEE Transactions on Information [Текст]. 2010. – сторінки 227-236
8. Jang, S. M., & Kim, J. K. (2018). Third person effects of fake news: Fake news regulation and media literacy interventions. Computers in Human Behavior [Текст]. 2018. – сторінки 295-302.

ДОДАТОК А

КОД ПРОГРАМНОЇ РЕАЛІЗАЦІЇ ПРИХОВАНОЇ МОДЕЛІ МАРКОВА

```
import spacy
import random
from collections import defaultdict
import pickle
import pandas as pd

language = 'en_core_web_sm'

def preprocess_text(text):
    nlp = spacy.load(language)

    doc = nlp(text)

    processed_words = []
    for sentence in doc.sents:
        sentence_words = []
        for token in sentence:
            if not token.is_stop:
                lemma = token.lemma_
                sentence_words.append(lemma)
        processed_words.extend(sentence_words)

    return processed_words

def build_markov_chain(words):
    chain = defaultdict(list)
    prev_word = None
```

```
    word_count = defaultdict(int)

    for word in words:
        if prev_word is not None:
            chain[prev_word].append(word)
            word_count[prev_word] += 1
        prev_word = word

    for key in chain:
        total_count = word_count[key]
        transition_words = chain[key]
        probabilities = [word_count[word] / total_count for word in
transition_words]
        chain[key] = list(zip(transition_words, probabilities))

    return chain

def evaluate_sentence(chain, sentence):
    words = sentence.split()
    likelihood = 1.0
    prev_word = None

    for word in words:
        if prev_word is not None:
            if prev_word in chain:
                transition_words, probabilities = zip(*chain[prev_word])
                if word in transition_words:
                    index = transition_words.index(word)
                    likelihood *= probabilities[index]
                else:
```

```
                likelihood = 0.0
            else:
                likelihood = 0.0
        prev_word = word

    return likelihood

def save_markov_chain(chain, filename):
    with open(filename, 'wb') as file:
        pickle.dump(chain, file)

def load_markov_chain(filename):
    with open(filename, 'rb') as file:
        chain = pickle.load(file)
    return chain

data = pd.read_csv('data.csv')

# Preprocess text and title columns
fake_news = []
good_news = []

for index, row in data.iterrows():
    processed_text = preprocess_text(row['text'] + " " + row['title'])

    if row['label'] == 1:
        fake_news.extend(processed_text)
    else:
        good_news.extend(processed_title)
```

```
label1_markov_chain = build_markov_chain(fake_news)
label2_markov_chain = build_markov_chain(good_news)

save_markov_chain(label1_markov_chain, 'markov_model1.pkl')
save_markov_chain(label2_markov_chain, 'markov_model2.pkl')

loaded_label1_chain = load_markov_chain('markov_model1.pkl')
loaded_label2_chain = load_markov_chain('markov_model2.pkl')

sentence = "у Кременчуці не було жодного теракту … армія рф не б’є ні
по жодних цивільних об’єктах під час спецоперації"
likelihood1 = evaluate_sentence(loaded_label1_chain, sentence)
likelihood2 = evaluate_sentence(loaded_label2_chain, sentence)

if likelihood1 > likelihood2:
    print("The sentence is more suitable for first")
else:
    print("The sentence is more suitable for second")
```