КВАЛІФІКАЦІЙНА РОБОТА
БАКАЛАВРА

на тему *Метод виявлення кібербулінгу в дописах соціальних інтернет-мереж нейромережевими засобами*

Галузь знань *12 – Інформаційні технології*
Шифр і назва галузі знань

Спеціальність *122 – Комп'ютерні науки*
Шифр і назва спеціальності

Освітня програма *Комп'ютерні науки*
Назва освітньої програми

Виконав:	*студент групи КНс-21-1*		*Владислав АНДРОЩУК*
	Курс, група виконавця	Підпис	Ім'я, ПРІЗВИЩЕ
Керівник:	*викладач каф. КН*		*Марина МОЛЧАНОВА*
	Науковий ступінь, посада	Підпис	Ім'я, ПРІЗВИЩЕ
Нормоконтроль:	*к.т.н., доц. каф. КН*		*Руслан БАГРІЙ*
	Науковий ступінь, посада	Підпис	Ім'я, ПРІЗВИЩЕ

До захисту допускаю:
зав. кафедри КН, д.т.н., професор — *Олександр БАРМАК*
Підпис — Ім'я, ПРІЗВИЩЕ

21 червня 2024 р.

Хмельницький 2024

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет *інформаційних технологій*
Кафедра *комп'ютерних наук*
Освітній ступінь *бакалавр*
Галузь знань *12 – Інформаційні технології*
Спеціальність *122 – Комп'ютерні науки*

ЗАТВЕРДЖУЮ
Завідувач кафедри *комп'ютерних наук*

_____
(підпис)
д.т.н., професор Олександр БАРМАК
«*16*» *02* *2024* року

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ БАКАЛАВРА

1. Тема кваліфікаційної роботи бакалавра: *«Метод виявлення кібербулінгу в дописах соціальних інтернет-мереж нейромережевими засобами»*

2. Завдання видано студенту *Владиславу АНДРОЩУКУ*
(Ім'я, прізвище)

3. Керівник роботи *викладач кафедри КН Марина МОЛЧАНОВА*
(посада, ім'я, прізвище)

4. Затверджено наказом університету від «*15*» *02* *2024* р. № *8*

5. Дата видачі завдання студенту: «*16*» *02* *2024* р.

6. Зміст пояснювальної записки (перелік задач) та вихідні дані:

Мета роботи – спрощення експертизи виявлення кібербулінгу за рахунок автоматизованого виявлення кібербулінгу в дописах соціальних інтернет-мереж. Задачі дослідження: виконати аналіз пердметної області, обрати теоретичні підходи до вирішення поставленої задачі; створити метод виявлення кібербулінгу в дописах соціальних інтернет-мереж; описати інформаційну структуру системи виявлення кібербулінгу в дописах; створити відповідну програмну реалізацію на основі створеного методу, виконати тестування створеного ПЗ; виконати дослідження ефективності створеного методу з використанням розробленого ПЗ. Результатами роботи системи є висновок стосовно наявності кібербулінга в дописі для аналізу та відсоткове значення ймовірності приналежності даних до класу кібербулінгу.

7. Календарний план виконання кваліфікаційної роботи бакалавра:

№	Назва етапів (розділів) кваліфікаційної роботи бакалавра	Термін виконання	Примітка
1	*Вибір напряму дослідження та узгодження тематики кваліфікаційної роботи бакалавра з керівником, складання календарного графіка виконання роботи*	*січень 2024*	Виконано
2	*Ознайомлення з предметною областю, формулювання мети та задач дослідження, визначення об'єкта та предмета дослідження*	*лютий 2024*	Виконано
3	*Проектування та розробка загальної архітектури програмного забезпечення, інтерфейсу користувача, вибір засобів реалізації програмного забезпечення*	*березень 2024*	Виконано
4	*Створення та тестування програмного забезпечення*	*квітень 2024*	Виконано
5	*Написання пояснювальної записки, урахування зауважень керівника, оформлення згідно вимог*	*травень 2024*	Виконано
6	*Розробка презентаційних матеріалів та попередній захист кваліфікаційної роботи*	*травень 2024*	Виконано
7	*Отримання відгуку керівника, рецензії, перевірка на плагіат, нормоконтроль*	*червень 2024*	Виконано
8	*Підготовка до захисту та захист кваліфікаційної роботи бакалавра*	*червень 2024*	Виконано

Виконавець: *студент групи КНс-21-1* _____ *Владислав АНДРОЩУК*
Курс, група виконавця — Підпис — Ім'я, ПРІЗВИЩЕ

Керівник: *викладач каф. КН* _____ *Марина МОЛЧАНОВА*
Науковий ступінь, посада — Підпис — Ім'я, ПРІЗВИЩЕ

Анотація

Тема кваліфікаційної роботи бакалавра: *«Метод виявлення кібербулінгу в дописах соціальних інтернет-мереж нейромережевими засобами»*

Виконавець кваліфікаційної роботи бакалавра: *студент групи КНс-21-1 Владислав АНДРОЩУК*

Керівник кваліфікаційної роботи бакалавра: *викладач каф. КН Марина МОЛЧАНОВА*

Кваліфікаційна робота бакалавра містить:

Пояснювальна записка				Кількість додатків
Сторінок	Рисунків	Таблиць	Джерел інформації	
66	28	4	31	4

Метою кваліфікаційної роботи бакалавра є спрощення експертизи виявлення кібербулінгу за рахунок автоматизованого виявлення кібербулінгу в дописах соціальних інтернет-мереж. Для розробки програмного забезпечення у вигляді інформаційної системи було використано мову програмування Python, а також хмарний сервіс Google Colab для навчання та збереження екземплярів нейромереж.

Розроблена система призначена для модераторів соціальних інтернет-мереж, а також може використовуватись науковцями для проведення наукових досліджень та розмітки нерозмічених даних для проведення наукових досліджень щодо протидії кібербулінгу

Напрямками практичного використання розробленої інформаційної системи визначено використання реалізації методу у вигляді надбудови до соціально-орієнтованих вебсервісів для протидії кібербулінгу.

Ключові слова: кібербулінг, BiLSTM, інформаційна система, система виявлення кібербулінгу, нейромережа.

Виконавець: *студент групи КНс-21-1* / *Владислав АНДРОЩУК*

Курс, група виконавця — Підпис — Ім'я, ПРІЗВИЩЕ

Зміст

Перелік скорочень

Скорочення, термін, позначення	Пояснення
DL	Глибоке навчання
ML	Машинне навчання
NLP	Обробка природної мови
RNN	Рекурентна нейронна мережа
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit
CNN	Конволюційні нейронні мережі
SVM	Метод опорних векторів
C-LSTM	Convolutional Long Short-Term Memory
RCNN	Recurrent Convolutional Neural Network
BERT	Bidirectional Encoder Representations from Transformers
BiLSTM	Bidirectional Long Short-Term Memory
SVC	Support Vector Classifier
COVID	Coronavirus Disease
TPU	Tensor Processing Unit
CPU	Central Processing Unit
NLTK	Natural Language Toolkit
ХНУ	Хмельницький національний університет.

Вступ

Метою кваліфікаційної роботи бакалавра було спрощення експертизи виявлення кібербулінгу за рахунок автоматизованого виявлення кібербулінгу в дописах соціальних інтернет-мереж, що досягалось шляхом розробки методу виявлення кібербулінгу в дописах соціальних інтернет-мереж нейромережевими засобами, а також відповідної програмної реалізації у вигляді інформаційної системи, яка буде використовувати розроблений метод.

Актуальність. В контексті сучасності, де взаємодія через соціальні мережі стає невід'ємною частиною життя, проблема кібербулінгу набуває особливої актуальності. Зростаюча кількість користувачів у цифровому просторі створює ідеальні умови для поширення ворожих або образливих повідомлень, які можуть серйозно впливати на психічне та емоційне становище людей.

Нейромережеві моделі можуть бути навчені розпізнавати ключові ознаки кібербулінгу, такі як агресивна лексика, образливі коментарі та загрози, забезпечуючи автоматизовану систему виявлення без необхідності ручного моніторингу. Здатність автоматично аналізувати величезні обсяги інформації за короткий проміжок часу дозволяє вчасно виявляти та реагувати на можливі випадки кібербулінгу. Застосування нейромережевих технологій дозволяє підвищити ефективність та точність цього процесу, що важливо для забезпечення безпеки та благополуччя користувачів соціальних мереж.

Таким чином, використання нейромереж для виявлення кібербулінгу в соціальних мережах визначається необхідністю забезпечення безпеки та психологічного комфорту користувачів у віртуальному просторі, що стає дедалі важливішим в аспекті формування здорового та етичного інтернет-середовища.

Об'єкт дослідження – процес виявлення кібербулінгу в дописах соціальних інтернет-мереж нейромережевими засобами.

Предмет дослідження – нейромережеві методи для роботи з текстовою інформацією.

Мета кваліфікаційної роботи бакалавра полягає в спрощенні експертизи виявлення кібербулінгу за рахунок автоматизованого виявлення кібербулінгу в дописах соціальних інтернет-мереж..

Завдання кваліфікаційної роботи бакалавра – виконати аналіз інформаційних моделей області виявлення кібербулінгу; виконати огляд теоретичних підходів та обрати підхід для нейромережевого виявлення кібербулінгу; провести аналіз існуючих публікацій за напрямком дослідження; провести аналіз існуючого програмного забезпечення області виявлення кібербулінгу в дописах соціальних інтернет-мереж; створити метод виявлення кібербулінгу в дописах соціальних інтернет-мереж; описати інформаційну структуру системи виявлення кібербулінгу в дописах; обрати набір даних для навчання нейромережевої компоненти методу; створити відповідну програмну реалізацію на основі створеного методу; виконати тестування створеного ПЗ; виконати дослідження ефективності створеного методу з використанням розробленого ПЗ.

Розділ 1 Характеристика предметної області виявлення кібербулінгу в дописах соціальних інтернет-мереж

1.1 Аналіз інформаційних моделей області виявлення кібербулінгу

Кібербулінг, або інтернет-мобінг, представляє собою нову форму агресії, яка відбувається в онлайн-середовищі. Атакуючи особу, нападник користується різними засобами спілкування, такими як соціальні мережі, електронна пошта та месенджери, з метою тиснути, завдавати шкоди та принижувати [1]. Це систематична та повторювана маніпуляційна поведінка, спрямована на індивіда чи групу, з метою залякування, викликання гніву або приниження. Основна мета кібербулінгу полягає в настанні психологічних чи емоційних негараздів жертв [2].

Спочатку однією з перших жертв кібербулінгу став американський підліток на ім'я Гіслан Раза. У 2002 році він створив відео, де відтворив сцену з фільму "Зоряні війни", замінюючи відомий меч на бейсбольну биту. Відео потрапило до рук його однокласників, які без його відома розмістили його в Інтернеті. Це призвело до того, що його почали дражнити та називати "дитиною зоряних війн", що вплинуло на його стосунки та психічний стан.

Небезпека інтернет-мобінгу полягає в тому, що віртуальна агресія може мати реальні наслідки, перетворюючись у фізичні знущання. Деякі можливості Інтернету також дозволяють агресорам діяти анонімно, що робить цю форму цькування особливо небезпечною.

Приклади кібербулінгу включають у себе такі ситуації:

– розповсюдження неправдивої інформації або публікація компрометуючих фотографій про когось у соціальних мережах;

– відправлення образливих повідомлень або загроз, спрямованих на приниження чи завдання шкоди через платформи обміну повідомленнями;

– піддавання себе під чуже ім'я та надсилання повідомлень іншим людям, при цьому представляючи себе за іншу особу.

Важливо відзначити, що особистий булінг та кібербулінг часто взаємопов'язані, проте кібербулінг залишає цифровий слід, у вигляді записів, які можуть служити доказами і допомагати зупинити цю форму тискання.

Закони, що забороняють булінг, а зокрема і кібербулінг, є відносно новими і на даний момент прийняті ще не в усіх країнах. Ось чому багато країн для боротьби з кібербулінгом використовують інші відповідні закони, наприклад, законодавство проти домагань.

Відповідно до законодавства України, згідно зі статтями 2 та 6 Закону "Про захист суспільної моралі", заборонено виробництво та розповсюдження будь-якої форми порнографічної продукції в Україні. В той же час, виробництво та розповсюдження продукції еротичного характеру та творів, що містять елементи насильства та жорстокості, дозволяється лише за умови дотримання законодавчих обмежень [3].

Згідно з нормами закону, забороняється виробництво та розповсюдження продукції, яка:

– пропагує війну, національну та релігійну ворожнечу, або веде до зміни конституційного ладу або територіальної цілісності України;

– поширює фашизм та неофашизм;

– ображає націю чи окремі особистості за національною ознакою;

– пропагує бузувірство, блюзнірство, або виявляє неповагу до національних і релігійних святинь;

– принижує особистість, включаючи знущання з фізичних вад чи каліцтва, а також з душевнохворих та літніх людей;

– пропагує невігластво та виявляє неповагу до батьків;

– рекламує наркоманію, токсикоманію, алкоголізм, тютюнопаління та інші шкідливі звички.

Характеристики кібербулінгу включають [4]:

– систематичні та повторювані дії, які можна визначити як кібербулінг;

– наявність різних сторін: кривдника (булі), потерпілого (жертву кібербулінгу) та спостерігачів, якщо такі існують;

– дії чи бездіяльність кривдника, що призводять до заподіяння психічної та/або фізичної шкоди, приниження, страху, тривоги, підпорядкування інтересам кривдника та/або призводять до соціальної ізоляції потерпілого.

До найпоширеніших видів кібербулінгу належать такі види, як: використання особистої інформації, анонімні погрози, хепіслепінг, обмовляння або зведення наклепів, переслідування, тролінг, флеймінг, онлайн-грумінг, секстинг [5]. Нижче наведено значення даних видів кібербулінгу.

Використання особистої інформації означає незаконне отримання конфіденційних даних через злам поштових скриньок, серверів, або соціальних мереж для подальшого переслідування особи.

Анонімні погрози включають надсилання загрозливих повідомлень електронною поштою, часто із вульгарною лексикою та образливим змістом.

Кіберпереслідування може відбуватися через мобільний зв'язок або електронну пошту, і може включати тривалі періоди приниження та шантажу жертви.

Тролінг передбачає розміщення провокаційних повідомлень для виклику конфліктів в інтернеті, що може призводити до взаємних образ.

Флеймінг – обмін гнівними повідомленнями між учасниками в інтернеті, часто в "публічних" місцях.

Обмови та зведення наклепів – поширення принизливої, неправдивої інформації з використанням комп'ютерних технологій.

Хепіслепінг – розповсюдження записів реальних сцен насильства через відео.

Секстинг – обмін інтимними матеріалами через мобільні телефони та соціальні мережі, що може призводити до кібербулінгу.

Секстинг може також стати перешкодою для майбутнього, оскільки інформація в інтернеті може впливати на процес навчання та пошук роботи.

Створення груп з образливими вмістом та розміщення інтимних фото колишніх коханих є прикладом кібербулінгу в соціальних мережах.

Ці дії можуть мати серйозні наслідки для жертви, включаючи агресивне переслідування та можливе розміщення матеріалів на сайтах з дитячою порнографією.

Учасники кібербулінгу зазвичай є особи в таких ролях: кривдник (булер), потерпілий (жертва), та спостерігачі.

Кривдник (булер) – це особа, яка бере участь у процесі, і вчиняє акти булінгу або цькування щодо іншого учасника колективу.

Потерпілий (жертва) – це особа, яка стала об'єктом булінгу або цькування.

Спостерігачі – це свідки чи безпосередні очевидці події булінгу.

Отже, виявлення і вчасна реакція на кібербулінг є важливим елементом захисту безпеки користувачів соціальних інтернет-мереж, тому автоматизація є актуальною задачею інформаційних технологій. Тому з проведеного аналізу, буде виконуватись автоматизоване виявлення кібербулінгу в дописах соціальних інтернет-мереж.

1.2 Огляд теоретичних підходів до виявлення кібербулінгу

У сучасну цифрову епоху широке використання соціальних мереж та онлайн-комунікацій породило нові проблеми, включаючи зростання кібербулінгу [6].

Завдяки анонімності та доступності Інтернету люди можуть деструктивно поводити себе в Інтернеті, переслідуючи або залякуючи інших, що призводить до руйнівних наслідків для жертв.

Тому в галузі інформаційних технологій наразі є декілька ключових підходів для вирішення проблеми виявлення кібербулінгу. Підхід з використанням машинного навчання. Машинне навчання є потужним інструментом у сфері штучного інтелекту. Особливо алгоритми машинного навчання можуть бути навчені виявляти закономірності в онлайн-спілкуванні, які вказують на поведінку, пов'язану з кібербулінгом.

Ці алгоритми спроможні виявляти випадки кібербулінгом в режимі реального часу, проаналізувавши величезні обсяги даних, що збираються з платформ соціальних мереж, месенджерів та інших онлайн-платформ. Це відкриває можливості для оперативного втручання та профілактичних заходів.

Одним із застосувань машинного навчання, яке може допомогти виявити кіберзалякування, є обробка природної мови. Алгоритми НЛП можуть аналізувати мову, використану в онлайн-спілкуванні, для визначення тону і настрою повідомлення, а також ідентифікації конкретних термінів чи фраз, пов'язаних з поведінкою, що залякує.

Наприклад, якщо людина часто використовує нецензурну лексику або робить загрозливі заяви, алгоритм може визначити це як потенційно образливу поведінку та повідомити відповідні органи.

За словами доктора Манджівана, використання машинного навчання для виявлення кібербулінгу має численні переваги, особливо з точки зору масштабованості. Звичайні методи запобігання кібербулінгу, такі як ручний моніторинг онлайн-платформ, можуть бути неефективними і вимагати багато часу, особливо для великих соціальних мереж із мільйонами користувачів.

Навпаки, алгоритми машинного навчання дозволяють вчасно та ефективно виявляти випадки кібербулінг та реагувати на них.

Однак цей підхід також породжує певні проблеми, зокрема, необхідність великих обсягів високоякісних даних для навчання алгоритмів, що є однією з найскладніших аспектів.

Навіть при тому, що кібербулінг вже досить поширене поняття, воно залишається маловивченою проблемою, особливо в контексті регіональних мов, до яких належить і українська. Недостатність загальнодоступних наборів даних, що містять прояви булінгу, становить проблему для вчених.

До засобів машинного навчання також належать і нейромережі, які також спроможні виконувати завдання виявлення кібербулінга.

Один з підходів до виявлення кібербулінгу за допомогою RNN включає використання моделей, таких як Long Short-Term Memory (LSTM) або Gated Recurrent Unit (GRU). Ці моделі здатні аналізувати текстові повідомлення та виявляти ознаки кібербулінгу, такі як образливі коментарі, загрози, негативні висловлення тощо [7].

Основна ідея полягає в тому, що рекурентні нейронні мережі можуть аналізувати послідовні дані, такі як текстові повідомлення в соціальних мережах, та виявляти шаблони, що вказують на наявність кібербулінгу. Наприклад, модель може виявити, що деякі типи коментарів часто супроводжуються негативними емоціями або агресивною лексикою, що свідчить про наявність кібербулінгу.

Для навчання таких моделей необхідна велика кількість данних, які включають в себе як позитивні, так і негативні приклади поведінки. Ці дані можуть бути розмічені експертами, які визначають, що являє собою кібербулінг, і на основі цього використовуються для навчання моделей RNN.

Після навчання модель може бути застосована для аналізу нових повідомлень і виявлення потенційних випадків кібербулінгу. Це дозволяє оперативно втручатися та надавати допомогу тим, хто стикається з цією проблемою в онлайн-середовищі.

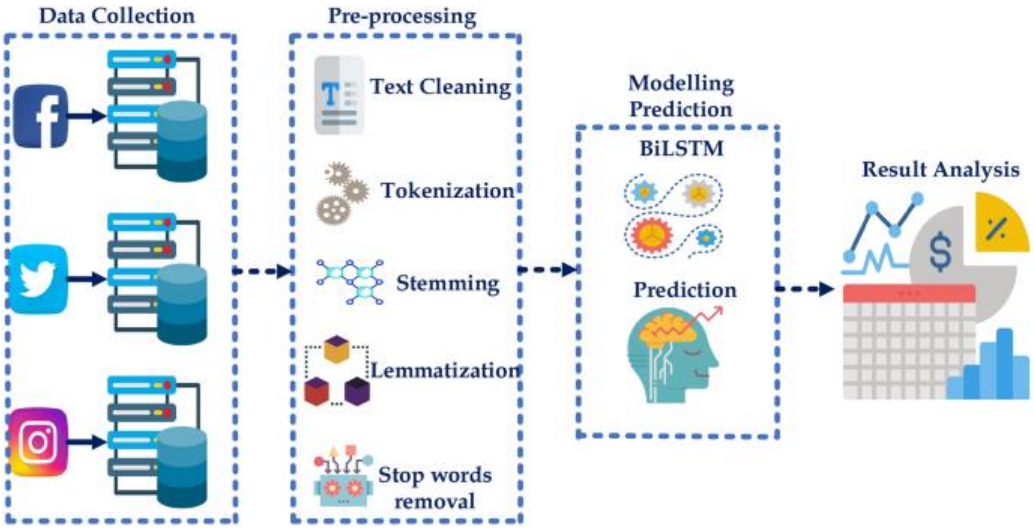


Рисунок 1.1 – Виявлення кібербулінгу нейромережею BiLSTM [9]

Двонаправлений LSTM, або BiLSTM, що є моделлю обробки послідовності, яка складається з двох LSTM: один приймає вхідні дані в прямому напрямку, а інший – у зворотному напрямку. BiLSTM ефективно збільшують обсяг інформації, доступної для мережі, покращуючи контекст, доступний для алгоритму (наприклад, знання того, які слова безпосередньо

слідують і передують слову в реченні) [8]. Процес виявлення кібербулінгу за допомогою нейромережі BiLSTM наведено на рисунку 1.1.

Отже, технологічні досягнення, такі як машинне навчання, дають надію на підвищення ефективності виявлення та запобігання кібербулінгу. В рамках роботи з огляду теоретичних підходів буде використано нейромережевий підхід, а саме з архітектурою BiLSTM.

1.3 Аналіз існуючих публікацій щодо автоматизації процесу виявлення кібербулінгу

Кібербулінг в Інтернеті є небезпечним і призводить до нещасних випадків, таких як самогубства, депресія тощо, тому необхідно контролювати його поширення. Тому виявлення кібербулінгу є життєво важливим на платформах соціальних мереж [10]. Основна мета цього дослідження – ідентифікація коментарів із кібербулінгом. Потрібні дані беруться з веб-сайту машинного навчання «Kaggle» та «Github». У цьому проекті використовуються три різних алгоритми машинного навчання: логістична регресія, метод опорних векторів, багатокласовий наївний Баєс для визначення найкращого алгоритму машинного навчання за метриками оцінки для передбачення коментарів. Основна мета дослідження є передбачення коментарів із цькуванням та їх сили, такої як легкий, сильний, помірний.

Інформаційно-комунікаційні технології збагатили соціальні мережі та полегшили комунікацію. Однак кібербулінг на цих платформах мав серйозні наслідки. Механізми, залежні від користувача, такі як звітність, блокування та видалення повідомлень з кібербулінгом онлайн, є ручними і неефективними. Представлення тексту "bag-of-words" без метаданих обмежує класифікацію тексту з кібербулінгом в кіберпросторі. У дослідженні [11] розроблено автоматизовану систему виявлення кібербулінгу двома підходами: традиційним машинним навчанням та передавальним навчанням. В даному дослідженні використані дані AMiCA, які включають значну кількість контексту кібербулінгу

та структурований процес анотації. Текстові, сентиментальні та емоційні, статичні та контекстуальні вбудовування слів, психолінгвістичні, термінологічні та токсичні функції були використані в традиційному підході машинного навчання. В дослідженні використовуються функції токсичності для виявлення кібербулінгу. Також дослідження використовує функції психолінгвістики з інструменту Linguistic Inquiry and Word (LIWC) 2022, а також лексикон Empath для виявлення кібербулінгу. Вбудовування контексту ggeluBert, tnBert та DistilBert мають подібну ефективність, проте вибрані вбудовування DistilBert для вищого значення F-виміру. Текстові функції, вбудовування DistilBert та функції токсичності, які встановили новий стандарт, були топ-три унікальними функціями, коли подавалися окремо. Використання комбінації текстових, сентиментальних, вбудовувань DistilBert, психолінгвістичних та токсичних функцій у модель логістичної регресії підняло її ефективність до значення F-виміру 64,8%, перевершивши Linear SVC за швидкістю тренування та ефективністю обробки високоінформаційних функцій. Підхід передавального навчання полягав у доналаштуванні оптимізованих версій попередньо навчених мовних моделей, а саме DistilBert, DistilRoBerta та Electra-small, які виявилися більш швидкими у вирахуванні тренувань, ніж їх базові форми. Доналаштований DistilBert мав найвище значення F-виміру на рівні 72,42%, перевищивши традиційний метод машинного навчання. Дослідження встановило, що передавальне навчання було найкращим для підвищеної ефективності та зменшення зусиль, оскільки воно вимагало відсутності інженерії ознак та попередньої обробки даних.

Відповідно, методи DL мають перевагу в області виявлення порівняно з традиційними методами ML. Ще однією критичною моделлю для послідовних даних є рекурентна нейронна мережа, яку використовують для вирішення численних завдань обробки природної мови. RNN є моделлю на основі нейронних мереж, що використовує механізми пам'яті для збереження попереднього стану для наступного часового кроку. Проте RNN може мати проблему зниклого градієнта, якщо дані містять ознаки довгих послідовностей.

Тому Hochreiter і Schmidhuber запропонували нову модель RNN на основі короткочасної пам'яті LSTM. Нейромережа може керувати та передбачати важливі приховані ознаки з великими інтервалами в часовому ряді та подолати недоліки оригінального RNN. Крім того, Chung et al. [12] запропонували модель з воротами рекурентної одиниці (GRU), щоб спростити архітектуру LSTM. У їхній моделі ворота забуття та введення замінені на «ворота оновлення» для прискорення виконання та зменшення використання пам'яті. Вчені розробили систему виявлення кібербулінгу для педагогів, яка спрямована на соціальні мережеві взаємини студентів в Instagram, та використовували LSTM для класифікації повідомлень що містять булінг. Оскільки булінг нараховується, важливо визначити, чи мають місце повторні атаки проти одного й того ж учня. Загалом RNN використовується в численних дослідженнях для ефективного вилучення прихованих ознак з текстових даних. Таким чином, моделі на основі RNN, такі як проста RNN, GRU та LSTM, застосовуються до численних задач класифікації та регресії з послідовними даними.

Конволюційні нейронні мережі використовуються для збереження семантики, усунення витрат часу на вилучення ознак та ефективного підвищення точності класифікації. Загалом моделі CNN використовуються для вилучення прихованих ознак зображення.

Останніми роками CNN застосовується до вилучення прихованих ознак з текстових даних на основі векторів слів, і це може покращити їхню ефективність. Цей метод також може бути застосований до тексту для вилучення прихованих локальних ознак текстових послідовностей. Вчений [13] запропонував, що модель CNN може використовуватися для вилучення прихованих ознак з тексту для завдань класифікації речень, і що попередньо навчені вектори слів word2vec можуть бути використані як вхідні ознаки для моделей CNN. CNN використовує різну кількість ядер згортання для вилучення багаторівневих ознак з тексту згідно із зваженим підходом до обміну ваг. Крім того, для вилучення важливих ознак використовується підхід макс-пулінгу, а для визначення ймовірності, що кожна ознака належить певній категорії,

використовується функція SoftMax. Таким чином, метод CNN широко застосовується в дослідженнях класифікації тексту.

Вчені [14] виявили, що поточна технологія виявлення кібербулінгу може класифікувати лише бульозне мовлення для одного тематичного розділу на конкретній платформі; тому ці дослідники використали CNN, LSTM, двонаправлену LSTM та двонаправлену LSTM з механізмом уваги, щоб подолати цей недолік. Результати вищезазначених авторів вказують, що моделі DL можуть досягати високої ефективності з різними темами на різних соціальних платформах, таких як Formspring, Twitter та Wikipedia.

Автори [15] порівняли SVM та логістичну регресію із трьома методами глибокого навчання (тобто CNN, C-LSTM та CNN-LSTM) для вирішення питань класифікації тексту. їхні результати вказують на те, що C-LSTM перевершив інші моделі на наборі даних Formspring, що стосується кібербулінгу. Однак комбінація CNN та RNN також проявляє переваги в завданнях класифікації тексту.

Автори [16] запропонували рекурентну CNN, щоб отримати семантичні вектори слів та використовували макс-пулінг для вилучення критичних ознак для отримання всього вектора тексту. RCNN проявив високу ефективність в класифікації. Крім того, подвійні представлення кодера від трансформера – це потужна модель навчання представлення, з тренуванням на двох завданнях, таких як моделювання мови з маскою та передбачення наступного речення. BERT належить до передньо навченої моделі для отримання текстових представлень та покращення ефективності в завданні підзадачі. Тому багато дослідників використовують BERT як передньо навчену модель та налаштовують моделі в багатьох завданнях.

З огляду сучасних публікацій в області виявлення кібербулінгу, нейромережеві засоби є ключовими для вирішення даної проблеми. Проте, даний напрямок потребує доопрацьовування та подальших досліджень.

1.4 Аналіз існуючих програмних рішень

З огляду важливості теми кібербулінгу, вже ведуться розробки відповідного програмного забезпечення. Наприклад, на офіційному сайті StopBullying.gov [17] є опис застосунків та програмного забезпечення для цифрового моніторингу для батьків.

Батьки, які хочуть захистити своїх дітей від кібербулінгу, шкідливої цифрової поведінки та впливу контенту для дорослих, можуть використовувати програмне забезпечення для батьківського контролю та моніторингу, яке допоможе їм налаштувати системи, менш агресивні для їхніх дітей. Вигляд сайту наведено на рисунку 1.2

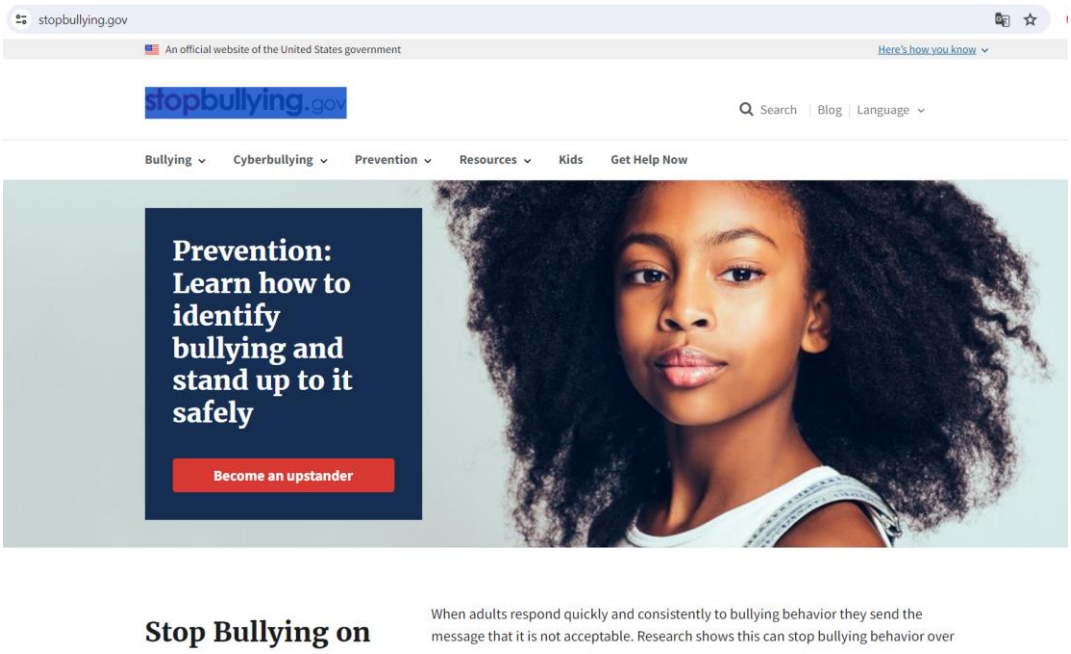


Рисунок 1.2 – Вигляд сайту StopBullying.gov

Доступні безкоштовні варіанти програмного забезпечення та програми, що допоможуть батькам обмежувати контент, блокувати домени або переглядати дії своїх дітей в Інтернеті, включаючи соціальні мережі, не заглядаючи щодня в налаштування своєї дитини. Більшість варіантів безкоштовного програмного забезпечення надають деякі функції безкоштовно, але за надійнішу інформацію стягується плата.

При виборі програмного забезпечення батьку слід враховувати вік дитини, використання пристрою та цифрову поведінку: те, що підходить для обмеження десятирічній дитині, може бути непотрібним для підлітка.

Також ще одним ресурсом є фонд «Cybersmile», що є некомерційною організацією, яка отримала багато нагород і присвячена підтримці цифрового благополуччя та боротьбі з будь-якими формами онлайн булінгом та жорстокості [18]. Фонд працює над сприянням доброзичливості, різноманітності та інклюзивності, створюючи безпечнішу та позитивнішу цифрову спільноту. Зображення офіційного сайту наведено на рисунку 1.3.

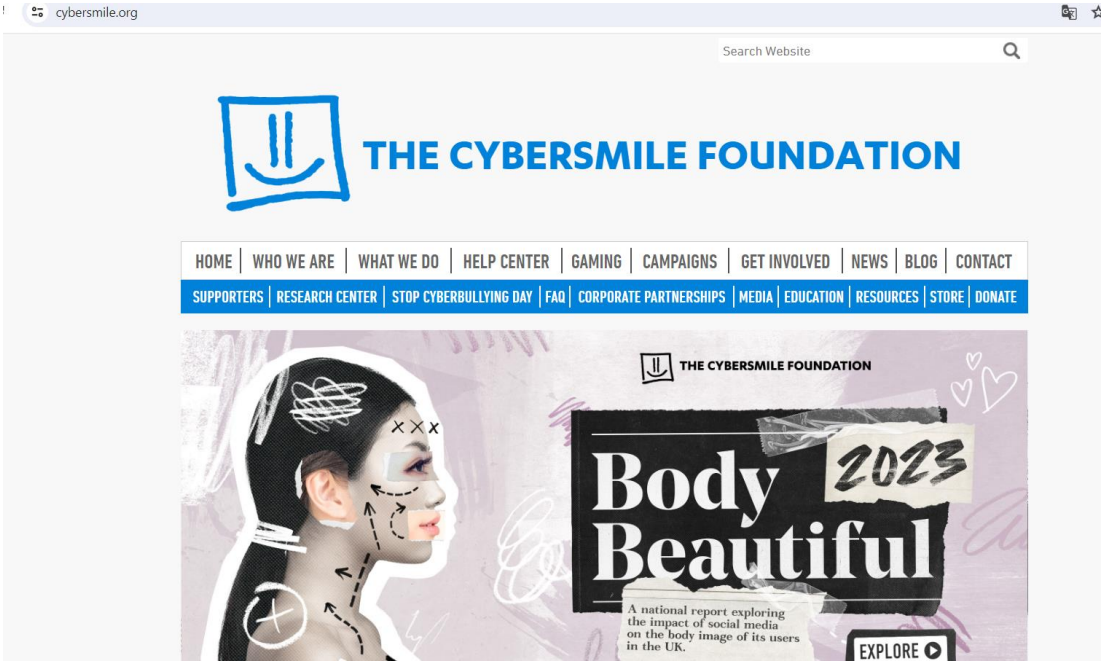


Рисунок 1.3 – The cybersmile foundation

За допомогою освіти, досліджень, просвітницьких кампаній та підтримки позитивного цифрового громадянства зменшуються випадки кібербулінгу. Через професійні служби допомоги та підтримки фахівцями надається можливість постраждалим та їхнім сім’ям відновити контроль над своїм життям.

Заснована в 2010 році, Cybersmile стала однією з провідних некомерційних організацій, що бореться з кібербулінгом і підтримує цифрове благополуччя. Зареєстрована як некомерційна організація 501(c)3 у США та як благодійна організація у Великобританії – Cybersmile надає експертну

підтримку, ресурси та консультації окремим особам, урядам, корпораціям та навчальним закладам у всьому світі.

Отже, зважаючи на недостатній рівень розвитку існуючого програмного забезпечення, подальша розробка є актуальною.

1.5 Мета, задачі та вимоги до реалізації інформаційної системи

Метою кваліфікаційної роботи бакалавра є спрощення експертизи виявлення кібербулінгу за рахунок автоматизованого виявлення кібербулінгу в дописах соціальних інтернет-мереж. Мета досягається шляхом розробки методу виявлення кібербулінгу в дописах соціальних інтернет-мереж нейромережевими засобами, та відповідного програмного забезпечення.

Для досягнення поставленої мети слід вирішити такі завдання:

– виконати аналіз інформаційних моделей області виявлення кібербулінгу;

– виконати огляд теоретичних підходів та обрати підхід для нейромережевого виявлення кібербулінгу;

– провести аналіз існуючих публікацій за напрямком дослідження;

– провести аналіз існуючого програмного забезпечення області виявлення кібербулінгу в дописах соціальних інтернет-мереж;

– створити метод виявлення кібербулінгу в дописах соціальних інтернет-мереж;

– описати інформаційну структуру системи виявлення кібербулінгу в дописах;

– обрати набір даних для навчання нейромережевої компоненти методу;

– створити відповідну програмну реалізацію на основі створеного методу;

– виконати тестування створеного ПЗ;

– виконати дослідження ефективності створеного методу з використанням розробленого ПЗ.

Розділ 2 Розробка методу виявлення кібербулінгу в дописах соціальних інтернет-мереж

2.1 Схема та кроки методу виявлення кібербулінгу в дописах соціальних інтернет-мереж

Метод виявлення кібербулінгу в дописах соціальних інтернет-мереж призначений для нейромережевого виявлення кібербулінга в текстових дописах, що публікуються в соціальних мережах. Схема та кроки методу наведені на рисунку 2.1.

Рисунок 2.1 – Схема методу виявлення кібербулінгу в дописах соціальних інтернет-мереж

Вхідними даними методу виявлення кібербулінгу в дописах соціальних інтернет-мереж є векторизатор, яким оброблялись дані під час навчання нейромережі, збережена нейромережева навчена модель для виявлення кібербулінгу та користувацький допис для аналізу вмісту кібербулінгу.

Першим кроком є попередня обробка допису для аналізу, що включає в себе видалення стоп-символів та стоп-слів, таких як розділові знаки, спеціальні символи тощо, а також перетворення у числову послідовність за допомогою збереженого векторизатора. Така попередня обробка допомагає зменшити розмірність даних та підготувати їх для подальшого аналізу за допомогою збереженої нейромережевої моделі BiLSTM.

Другим кроком є нейромережева оцінка наявності кібербулінгу в дописі. Нейромережа приймає на вхід текстові дані (перетворені на першому кроці у числову послідовність) та повертає прогноз щодо наявності кібербулінгу в тестовому дописі соціальних інтернет-мереж.

Третім кроком є опрацювання отриманих даних прогнозу та формування висновків користувачам. Висновок включає в себе відповідь стосовно наявності або відсутності кібербулінгу, а також відсоткову оцінку контенту як такого що містить кібербулінг, та як такого, що не містить його.

Вихідними даними роботи методу є висновок стосовно наявності кібербулінга в дописі для аналізу та відсоткове значення ймовірності приналежності даних до класу кібербулінга.

Отже, запропонований метод виявлення кібербулінгу в дописах соціальних інтернет-мереж призначений для перетворення вхідних даних у вигляді векторизатора, яким оброблялись дані під час навчання нейромережі, збереженої нейромережевої навченої моделі для виявлення кібербулінгу та користувацького допису для аналізу вмісту кібербулінгу у вихідні дані у вигляді висновку стосовно наявності кібербулінга в дописі для аналізу та відсоткового значення ймовірності приналежності даних до класу кібербулінга.

2.2 Аналіз та автоматизація обробки потоків даних

Внутрішній механізм, що забезпечує функціонування інформаційної системи виявлення кібербулінгу, що є прямою імплементацією створеного методу наведено на рисунку 2.2.

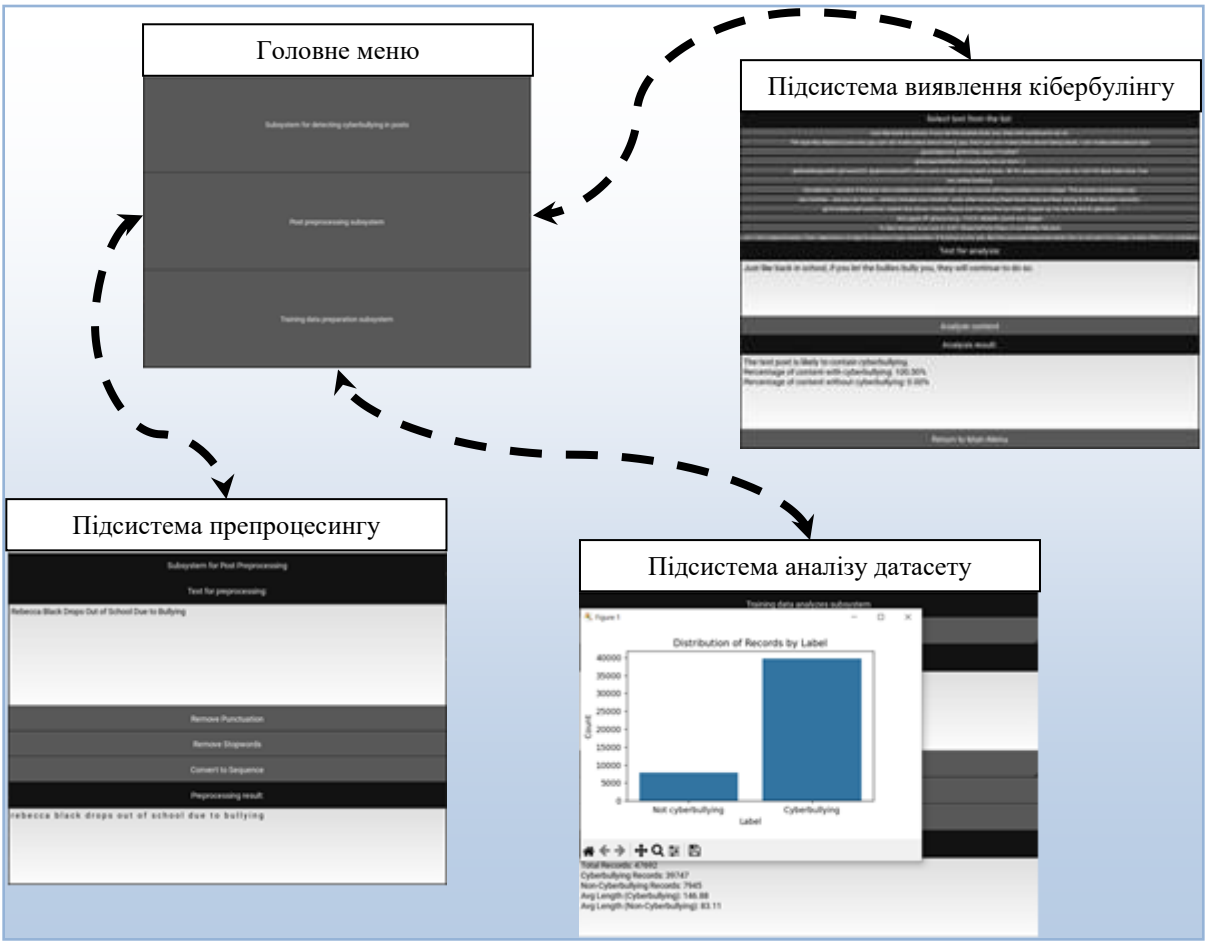


Рисунок 2.2 – Схема навігації між підсистемами інформаційної системи виявлення кібербулінгу

Необхідно надати можливість зручного користування підсистемами, тому важливо забезпечити переходи між підсистемами. Інформаційна система з точки зору користувача складається із головного меню та трьох підсистем: «Підсистеми виявлення кібербулінгу», «Підсистеми препроцесингу» та «Підсистеми аналізу датасету».

Підсистема виявлення кібербулінгу призначена для виявлення кібербулінгу в дописах соціальних інтернет-мереж нейромережевими засобами та виконує таку групу функцій:

– вибір допису для аналізу з датасета;

– уведення тексту для аналізу власноруч;

– аналіз текстового представлення з метою виявлення наявності кібербулінгу;

– виведення статистики виконаного аналізу щодо наявності кібербулінгу.

Підсистема препроцесингу призначена для виконання попередньої обробки заданого тексту, та виконує таку групу функцій:

– уведення тексту для подальшої обробки;

– видалення стоп-символів;

– видалення стоп-слів;

– перетворення тексту у числове представлення;

– виведення результатів користувачу.

Підсистеми аналізу датасету призначена для виконання аналізу вмісту набору даних, включно з графічною інтерпретаціює, та виконує таку групу функцій:

– перегляд вмісту датасету;

– додавання до датасету нового маркованого запису;

– виведення загальної кількість записів;

– виведення кількості записів, що містять кібербулінг;

– виведення кількості записів, що не містять кібербулінг;

– виведення середньої довжини записів, що містять кібербулінг;

– виведення середньої довжини записів, що не містять кібербулінг;

– виведення графічного представлення статистики.

Отже, було проведено аналіз та автоматизація обробки потоків даних інформаційної системи виявлення кібербулінгу.

2.3 Розробка архітектури нейронної мережі для виявлення кібербулінгу

Архітектура нейромережі BiLSTM, що є складовою методу виявлення кібербулінгу в дописах соціальних інтернет-мереж наведено на рисунку 2.3.

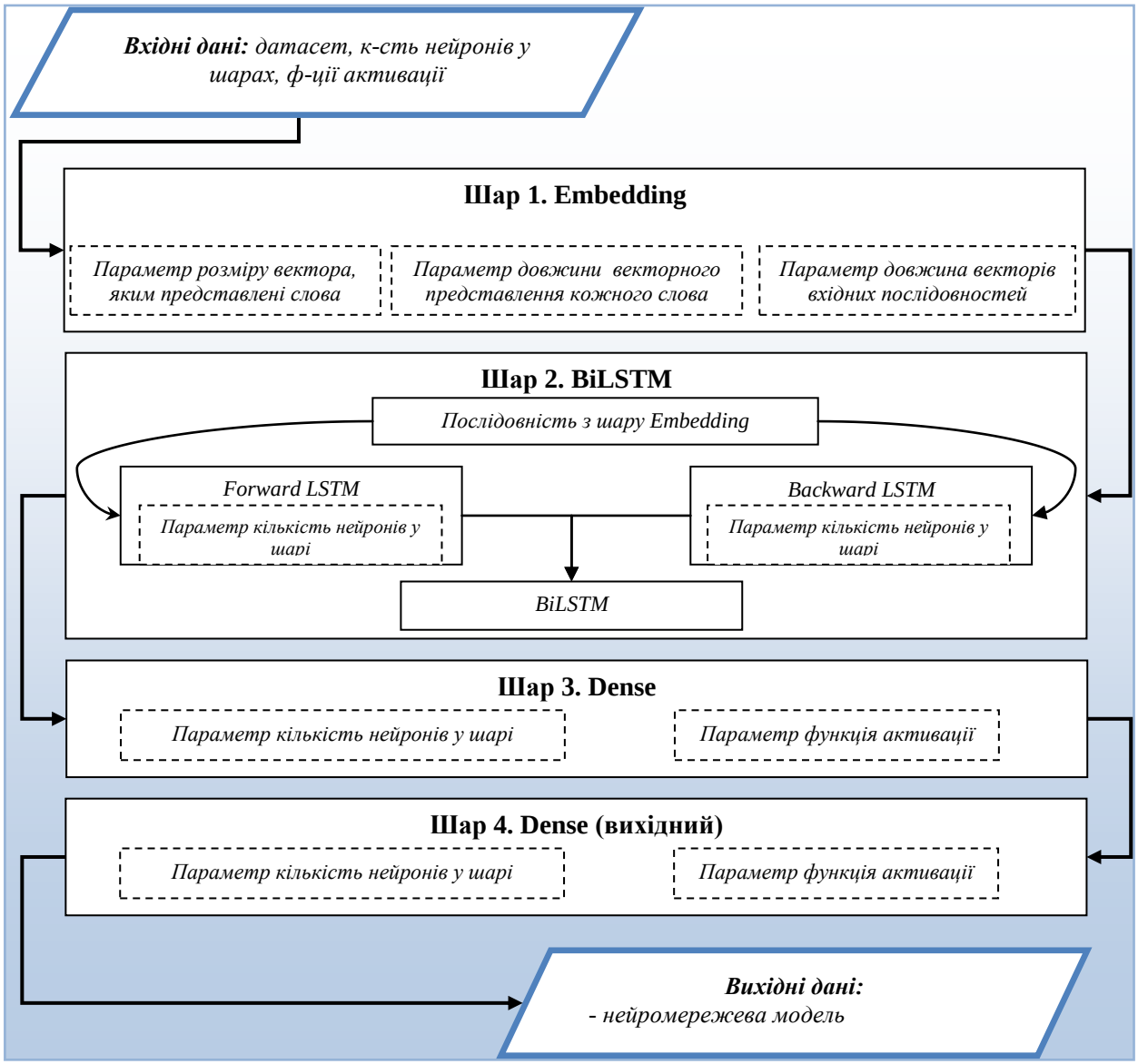


Рисунок 2.3 – Архітектура нейромережевої моделі BiLSTM

Вхідними даними є датасет, кількість нейронів у шарах, функції активації. Решта параметрів архітектури розраховується в залежності від вказаних даних.

Першим шаром є шар Embedding. Цей шар перетворює слова або символи у вектори фіксованої довжини. Він допомагає моделі розуміти семантичне значення слів та їх взаємозв'язки в контексті. Параметр розміру вектора, яким представлені слова є розміром словника, тобто кількість унікальних слів у вхідних даних. Відповідно, залежить від множини датасета. В рамках дослідження установлено в 5000. Параметр довжини векторного представлення кожного слова вказує на розмір векторів для кожного слова, в даному випадку буде встановлений у значення 50. Параметр довжина векторів вхідних послідовностей визначає максимальну довжину вхідної послідовності, тобто максимальну кількість слів у кожному вхідному дописі соціальних інтернет-мереж. В рамках дослідження даний параметр встановлено в значення 100.

Після шару Embedding дані переходять до шару BiLSTM. Цей шар є розширенням звичайного LSTM, яке обробляє вхідну послідовність в обох напрямках (вперед і назад) за допомогою двох LSTM шарів. Кожен LSTM шар має 64 нейрони. Forward LSTM – перший LSTM шар, який обробляє вхідну послідовність вперед (зліва направо). Backward LSTM – другий LSTM шар, який обробляє вхідну послідовність назад (справа наліво).

Bidirectional LSTM об'єднує виходи обох LSTM шарів (вперед та назад) у єдину послідовність. Ця послідовність містить інформацію з обох напрямків та використовується для подальшої обробки моделлю. Вихід Bidirectional LSTM є конкатенацією виходів з обох напрямків, що дозволяє моделі використовувати інформацію з контексту з обох сторін вхідної послідовності.

У передостанньому шарі Dense відбувається повне з'єднання кожного входу з кожним виходом, що включає в себе всі вхідні дані з попереднього шару. Кількість нейронів в рамках дослідження – 64. Це вказує, скільки внутрішніх нейронів буде у шарі, які будуть мати ваги, які будуть навчатися під час процесу навчання. Функція активації ReLU дозволяє шару нейронів вибирати, які сигнали проходять через нього (які приймають значення від нуля до нескінченності) і які – ні.

Останній шар нейромережі однойменний попередньому, Dense. Оскільки шар є кінцевим, функція активації для цього шару буде softmax. Кількість нейронів визначається кількість унікальних міток класів у датасеті. В рамках дослідження це 2 (кібербулінг або не кібербулінг). Softmax видає вектор ймовірностей для кожного класу, де кожне значення відображає ймовірність, що вхідний зразок належить до цього класу. Сума всіх ймовірностей вихідного вектора дорівнює 1, що дозволяє використовувати його для отримання ймовірностей класів.

Отже, було сформовано нейромережеву архітектуру для методу виявлення кібербулінгу в дописах соціальних інтернет-мереж, що складається із чотирьох основних шарів: Embedding, що перетворює слова або символи у вектори фіксованої довжини; BiLSTM, який є конкатенацією виходів з обох напрямків, що дозволяє моделі використовувати інформацію з контексту з обох сторін вхідної послідовності; та 2-х повнозв'язних шарів Dense. Створена модель потребує навчання, для реалізації можливості визначення кібербулінгу в дописах соціальних інтернет-мереж.

2.4 Основні етапи навчання нейромережі для виявлення кібербулінгу

Сформовану вище архітектуру нейромережі необхідно навчити, так як навчена нейромережева модель є складовою методу виявлення кібербулінгу в дописах соціальних інтернет-мереж призначений для нейромережевого виявлення кібербулінга в текстових дописах, що публікуються в соціальних мережах. Етапи навчання нейромережі наведені на рисунку 2.4.

Вхідними даними є датасет з текстами що містять кібербулінг, які взяті з соціальних інтернет-мереж, сформована модель BiLSTM та параметри навчання моделі.

На першому етапі здійснюється попередня обробка даних датасету, що включає в себе видалення стоп-символів та стоп-слів, таких як розділові знаки та спеціальні символи, а також перетворення тексту у числову послідовність. Ця

обробка сприяє зменшенню розмірності даних і готує їх для подальшого навчання нейромережі.

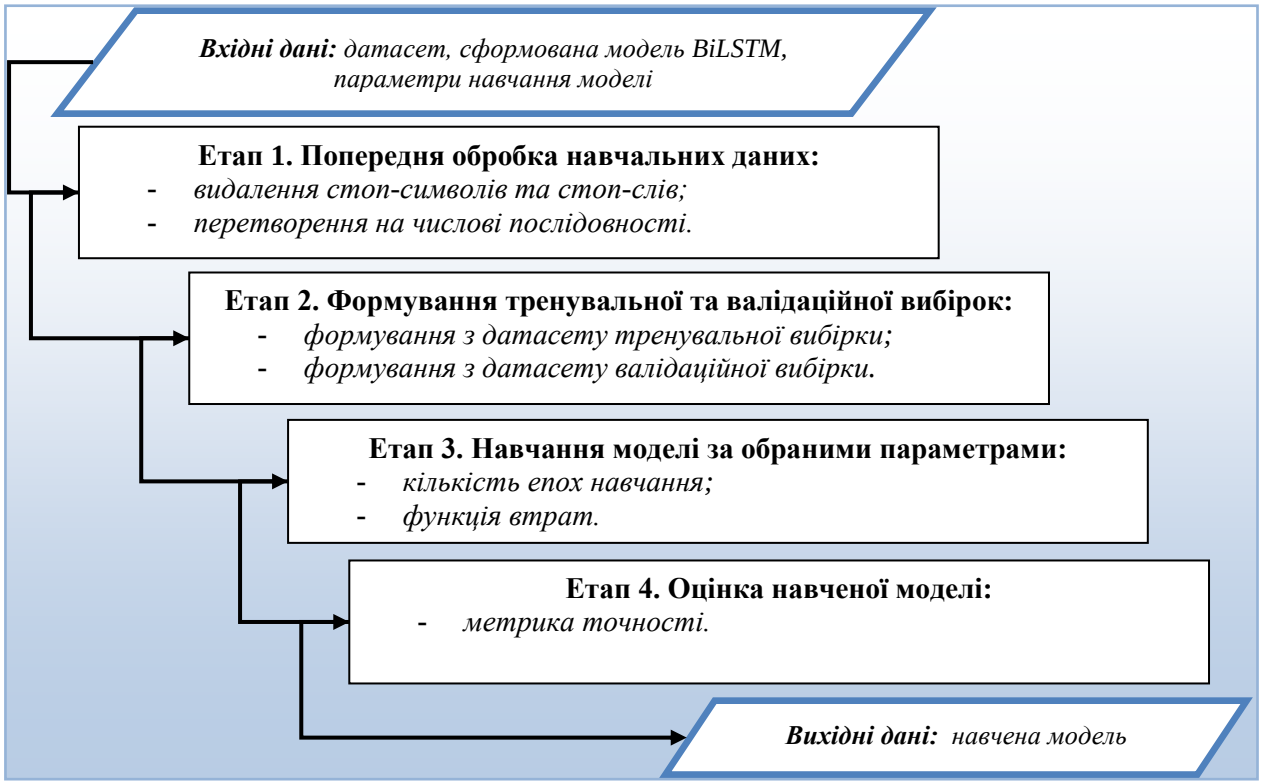


Рисунок 2.4 – Етапи навчання нейромережі BiLSTM

Етап формування тренувальної та валідаційної вибірок є критичним для успішного навчання нейромережевої моделі. Ці дві вибірки використовуються для навчання та оцінки моделі. Для формування цих вибірок, спершу виконується аналіз датасету, якщо дані є незбалансованими, то береться однакова кількість зразків з кібербулінгом та без кібербулінгу, та в подальшому ділиться у пропорції 80 на 20, де 80% це тренувальні дані, а 20 % – валідаційні. Після формування цих двох вибірок модель готова до навчання.

Наступним етапом є етап навчання моделі за обраними параметрами. Одним із параметрів є оптимізатор, що використовується для оновлення ваг моделі на кожному кроці навчання. У рамках дослідження буде використано Adam, який є адаптивним оптимізатором, який ефективно пристосовує швидкість навчання для кожного параметра. Кількість епох визначається

експериментально, проте для базової моделі кількість епох буде рівна 10. У якості функції втрат буде використано кросентропічну функцію втрат.

Етап навчання дуже тісно взаємодіє з етапом оцінки за метриками, адже ще під час навчання модель на кожній епосі робить оцінку продуктивності. У якості метрики навчання використовується метрика accuracy.

Вихідними даними є навчена збережена нейромережева модель, що спроможна виявляти кібербулінг у дописах соціальних інтернет-мереж, та є складовою запропонованого в кваліфікаційній роботі бакалавра методу.

Отже, наведено етапи навчання нейромережі на основі розробленої архітектури, яка є складовою методу виявлення кібербулінгу в дописах соціальних інтернет-мереж призначений для нейромережевого виявлення кібербулінга в текстових дописах, що публікуються в соціальних мережах. Результатом виконання вказаних етапів є навчена збережена нейромережа BiLSTM.

2.5 Проектна архітектура системи та взаємозв’язок компонентів

Наступним кроком проєктування інформаційної системи є побудова проектної архітектури. Проектна архітектура інформаційної системи виявлення кібербулінгу представлена на рисунку 2.5.

Підсистема навчання нейромережі призначена для навчання та збереження нейромережевої моделі BiLSTM за створеною архітектурою. Підсистема є допоміжною складовою, яка відповідає за побудова моделі за вказаною архітектурою, підготовка навчальних даних до навчання нейромережі, навчання нейромережевої моделі, збереження навченої нейромережевої моделі у файловій системі, формування оцінки якості моделі за метриками. Навчання нейромережі здійснюється на підібраному сформованому наборі даних.

Навчена нейромережева модель BiLSTM є прямим продуктом роботи підсистеми навчання нейромережі. Модель в подальшому використовується як компонента підсистеми виявлення кібербулінгу. Підсистема виявлення

кібербулінгу відповідає за вибір допису для аналізу з датасета, можливість уведення тексту для аналізу власноруч, виконання аналізу текстового представлення з метою виявлення наявності кібербулінгу, виведення статистики виконаного аналізу щодо наявності кібербулінгу.

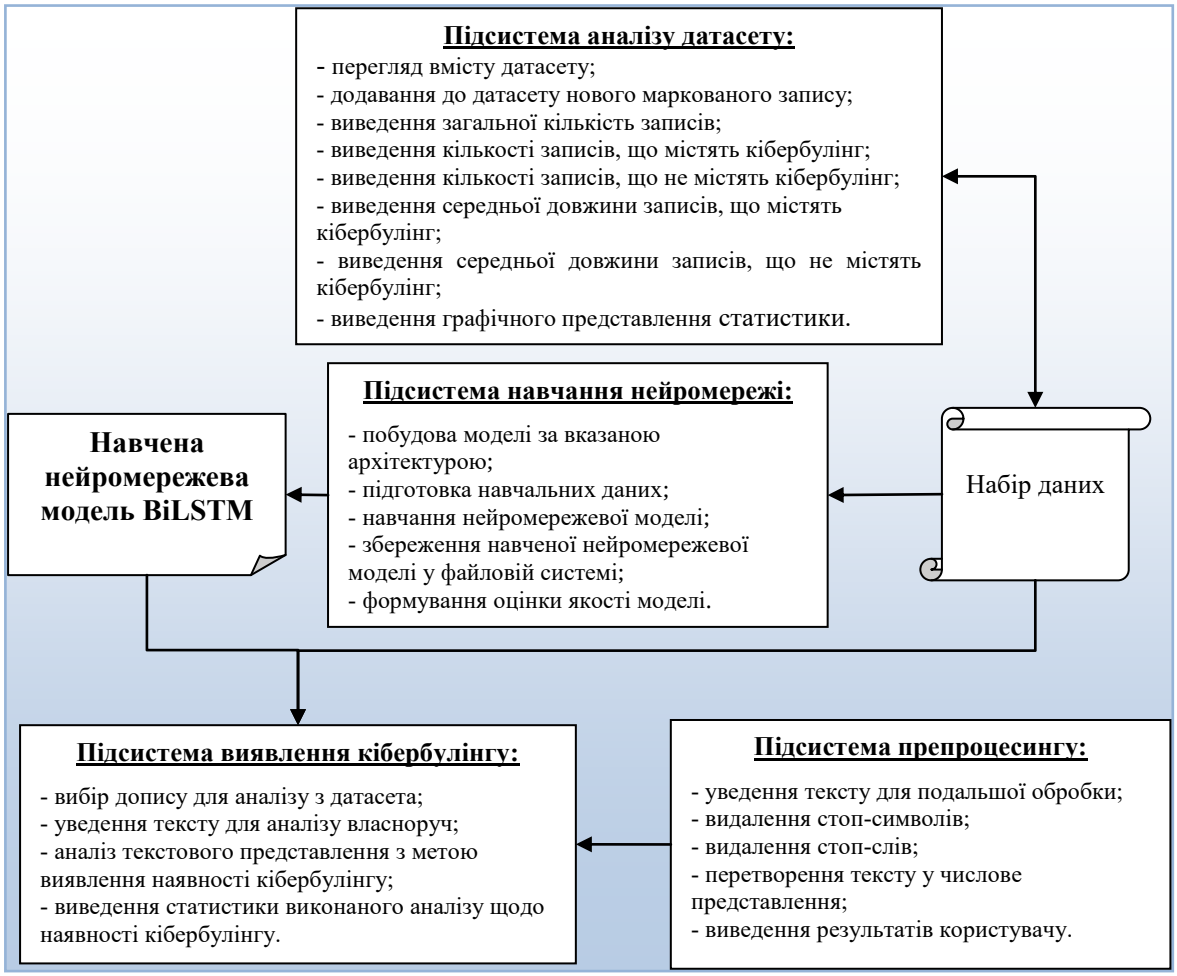


Рисунок 2.5 – Проектна архітектура інформаційної системи виявлення кібербулінгу

Підсистема препроцесингу використовується підсистемою виявлення кібербулінгу, яка користувацький допис перетворює у числову послідовність. Також вона відповідає за можливість перегляду усіх етапів попередньої обробки, таких як видалення стоп-символів, видалення стоп-слів, перетворення тексту у числове представлення, виведення результатів користувачу.

Підсистема аналізу датасету призначена для статистичної оцінки наявного його змісту, оскільки дата сети часто бувають не збалансовані, містити

порожні дані, тощо. Підсистема аналізу включає в себе такі можливості, як перегляд вмісту датасету, додавання до датасету нового маркованого запису, виведення загальної кількість записів, виведення кількості записів, що містять кібербулінг, виведення кількості записів, що не містять кібербулінг, виведення середньої довжини записів, що містять кібербулінг, виведення середньої довжини записів, що не містять кібербулінг, виведення графічного представлення статистики.

Отже, таким чином наведено проектну архітектуру інформаційної системи виявлення кібербулінгу в текстових дописах, що публікуються в соціальних інтернет-мережах. Подана архітектура включає в себе усі етапи та компоненти, які забезпечують реалізацію і апробацію запропонованого методу виявлення кібербулінгу в текстових дописах.

2.6 Підготовка робочих вхідних даних для інформаційної системи виявлення кібербулінгу

Оскільки використання соціальних медіа стає все більш поширеним для різних вікових груп, переважна більшість громадян покладаються на це важливе середовище для повсякденного спілкування. Загальне поширення соціальних медіа означає, що кібербулінг може ефективно впливати на будь-кого в будь-який час і будь-де, а також відносну анонімність, Інтернет робить такі особисті атаки такими, які важче зупинити, ніж традиційний булінг. 15 квітня 2020 року ЮНІСЕФ випустив попередження у відповідь на підвищений ризик кібербулінгу під час пандемії COVID-19 через масове закриття шкіл, збільшення часу перед екраном і зменшення соціальної взаємодії віч-на-віч [19].

Статистика кібербулінгу відверто тривожна: 36,5% учнів середньої та старшої школи відчували кібербулінг, а 87% спостерігали кібербулінг, наслідки якого варіюються від зниження успішності до депресії та суїцидальних думок.

У світлі вищевикладеного матеріалу, було сформовано набір даних, що містить понад 47 000 твітів, позначених відповідно до класу кібербулінгу [20]:

– вік;

– етнічна приналежність;

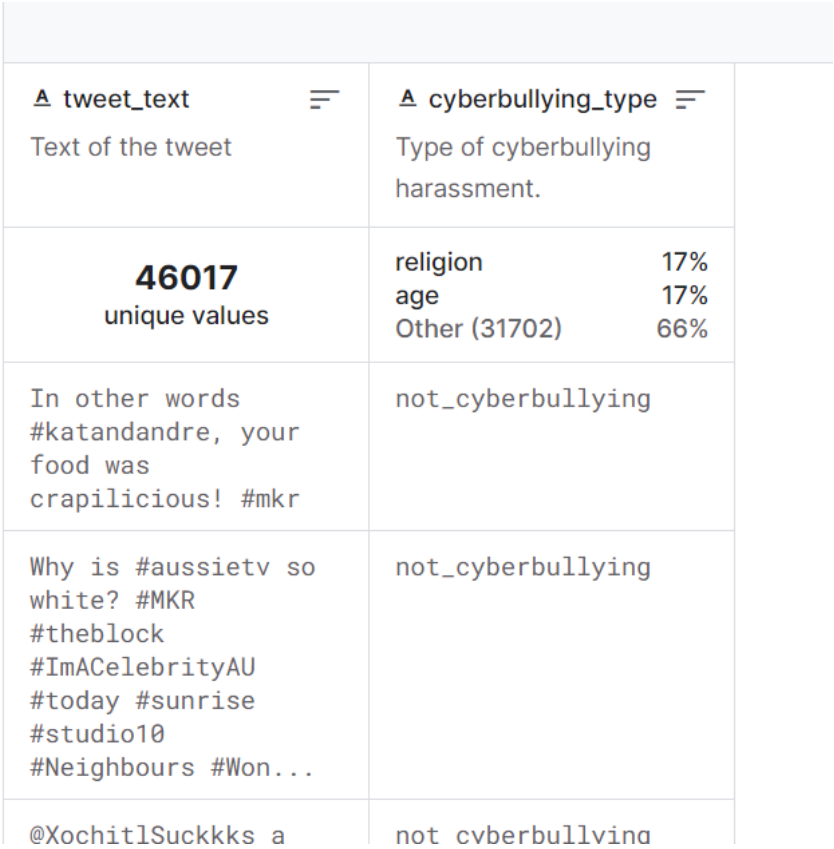
– стать;

– релігія;

– інші види кібербулінгу;

– не кібербулінг

Дані англомовні та збалансовані, містять приблизно по 8000 кожного класу. Проте у рамках дослідження буде використано тільки 2 класи – кібербулінг , який об’єднає всі види кібербулінгу та не кібербулінг. Приклад даних з датасету наведено на рисунку 2.6.



tweet_text Text of the tweet	cyberbullying_type Type of cyberbullying harassment.
46017 unique values	religion 17% age 17% Other (31702) 66%
In other words #katandandre, your food was crapilicious! #mkr	not_cyberbullying
Why is #aussietv so white? #MKR #theblock #ImACelebrityAU #today #sunrise #studio10 #Neighbours #Won...	not_cyberbullying
@XochitlSuckkks a	not_cyberbullying

Рисунок 2.6 – Приклад даних датасету

Після злиття даних, клас Cyberbullying налічує 39747, а клас Not cyberbullying налічує 7945 записів. Для навчання нейромережі дані будуть використані наступним чином: з класу Cyberbullying буде випадковим чином

обрано 8000 зразків, відповідно, навчальна множина даних буде складатись з 15 945 зразків, рівномірно розподілених між двома цільовими класами.

Отже, для навчання нейромережі BiLSTM буде використано набір даних Cyberbullying Classification, що налічує 39747 зразків, з яких буде обрано 8000 зразків, що містять кібербулінг, та 7945 записів без кібербулінгу.

2.7 Особливості використання спеціалізованих програмних компонентів інформаційної системи виявлення кібербулінгу

Окрім цього для реалізації методу виявлення кібербулінгу в дописах соціальних інтернет-мереж необхідно використати спеціалізовані програмні розширення.

Torch – це фреймворк для наукових обчислень та розробки машинного навчання, який забезпечує швидкість та гнучкість. Одна з основних бібліотек у ньому – PyTorch. PyTorch використовується для розробки та навчання нейронних мереж, а також для наукових обчислень [21].

Keras – високорівневий API для машинного навчання, який спрощує процес створення та навчання моделей. Зазвичай використовується як інтерфейс для інших бібліотек, зокрема TensorFlow, щоб забезпечити простоту та ефективність у розробці [22].

Keras є однією з найпопулярніших бібліотек для реалізації моделей машинного навчання та глибокого навчання. Вона забезпечує простий та інтуїтивно зрозумілий інтерфейс для створення, навчання та валідації нейронних мереж. Основною перевагою Keras є його високорівневий API, який дозволяє розробникам швидко та легко створювати складні моделі з найменшими зусиллями.

Бібліотека має вбудовані функції для створення різних типів моделей, таких як звичайні нейронні мережі, згорткові нейронні мережі, рекурентні нейронні мережі та їх комбінації. Крім того, Keras надає можливість легко

налаштовувати гіперпараметри моделі, вибирати різні функції активації, функції втрат та оптимізатори, щоб досягти оптимальної продуктивності.

TensorFlow – це відкрита бібліотека для числових обчислень, яка стала основою для багатьох проектів з машинного навчання. Використовується для розробки та навчання глибоких нейронних мереж, а також для широкого спектру завдань в галузі машинного навчання [23].

TensorFlow є однією з найпопулярніших та найбільш впливових бібліотек для реалізації моделей машинного навчання та глибокого навчання. Однією з ключових переваг TensorFlow є його висока продуктивність та масштабованість, що дозволяє обробляти великі обсяги даних та виконувати складні обчислення на різних пристроях, включаючи CPU, GPU та TPU. Бібліотека також надає інструменти для оптимізації роботи з пам'яттю та обчисленнями.

Крім того, TensorFlow активно розвивається та підтримується великою спільнотою розробників, яка постійно вносить внески у розширення можливостей бібліотеки та удосконалення її функціональності. Це робить TensorFlow однією з найбільш важливих та перспективних бібліотек для реалізації проектів з машинного навчання та глибокого навчання.

Os модуль стандартної бібліотеки Python, який забезпечує функції для взаємодії з операційною системою. Використовується для виконання операцій з файловою системою, директоріями та іншими операціями, пов'язаними з операційною системою [24].

Оскільки в рамках теми є потреба працювати з текстом, то також використовується бібліотека Natural Language Toolkit, популярна бібліотека для обробки природної мови у середовищі Python. Вона надає широкий спектр інструментів для роботи з текстовими даними, включаючи токенізацію, лематизацію, векторизацію, роботу з частотними словниками, виявлення частин мови, роботу з синтаксичними деревами та багато іншого [25].

NLTK має простий та зрозумілий інтерфейс, що робить її доступною для використання як початківцям, так і досвідченим дослідникам у галузі обробки природної мови. Бібліотека містить велику кількість корпусів текстових даних,

граматик та інших ресурсів, які можна використовувати для навчання моделей та проведення досліджень.

NLTK підтримує різноманітні завдання в галузі обробки природної мови, включаючи аналіз сентименту, класифікацію текстів, розпізнавання іменованих сутностей, машинний переклад та інші. Вона стала важливим інструментом для багатьох досліджень та проектів у галузі обробки текстів, машинного навчання та штучного інтелекту.

Matplotlib – це бібліотека для створення візуалізацій у середовищі Python. Вона дозволяє створювати різноманітні графіки, діаграми, гістограми, карти, зображення та інші типи візуалізацій для аналізу даних та представлення результатів [26]. Буде використана для візуалізації даних датасету та побудови графіків статистики навчання нейромережі.

Matplotlib надає широкий спектр можливостей для створення візуальних представлень даних. Вона підтримує різні типи графіків, включаючи лінійні графіки, точкові графіки, гістограми, кругові діаграми, контурні графіки, теплові карти та багато інших.

Однією з особливостей Matplotlib є те, що вона інтегрується з іншими бібліотеками для наукових обчислень у Python, такими як NumPy та Pandas, що дозволяє зручно працювати з числовими даними та створювати візуалізації на їх основі.

Matplotlib має простий та зрозумілий інтерфейс, що робить її популярним інструментом для створення візуалізацій у середовищі Python. Вона використовується як для створення простих графіків, так і для складних візуалізацій у наукових дослідженнях, аналізі даних, веб-розробці та інших галузях.

Kivy – це відкрита бібліотека для розробки крос-платформенних додатків та інтерфейсів користувача на мові програмування Python [27]. Вона надає можливості для створення інтерактивних програм з графічним інтерфейсом, які можуть працювати на різних операційних системах, включаючи Windows, macOS, Linux, Android і iOS. Kivy дозволяє розробникам створювати додатки з

різноманітними функціями, від мультимедійних програм до ігор та освітніх додатків.

Однією з ключових особливостей Kivy є його мультидотикове введення, що дозволяє створювати інтерфейси, що реагують на декілька дотиків одночасно, що особливо важливо для сенсорних пристроїв. Крім того, Kivy має потужні засоби для роботи з графікою, звуком і відео, що дозволяє розробникам створювати додатки з вражаючими візуальними та аудіо ефектами. Він також підтримує анімацію, що дозволяє створювати динамічні та привабливі інтерфейси.

Отже, для створення інтерфейсів користувача буде використано бібліотеку Kivy, яка надає можливості для створення інтерактивних програм з графічним інтерфейсом, які можуть працювати на різних операційних системах. Для побудови графіків результатів навчання нейромережі та статистики вмісту датасету буде використано Matplotlib. Для попередньої обробки текстових даних датасету для виявлення кібербулінгу буде використано бібліотеку NLTK, а для навчання нейромережі буде використано бібліотеку TensorFlow.

2.8 Висновки до розділу 2

У рамках виконання другого розділу запропоновано метод виявлення кібербулінгу в дописах соціальних інтернет-мереж, що призначений для перетворення вхідних даних у вигляді векторизатора, яким оброблялись дані під час навчання нейромережі, збереженої нейромережевої навченої моделі для виявлення кібербулінгу та користувацького допису для аналізу вмісту кібербулінгу у вихідні дані у вигляді висновку стосовно наявності кібербулінга в дописі для аналізу та відсоткового значення ймовірності приналежності даних до класу кібербулінга.

Проведено аналіз та автоматизація обробки потоків даних інформаційної системи виявлення кібербулінгу, що основана для реалізації методу.

Створено нейромережеву архітектуру для методу виявлення кібербулінгу в дописах соціальних інтернет-мереж, що складається із чотирьох основних шарів: Embedding, що перетворює слова або символи у вектори фіксованої довжини; BiLSTM, який є конкатенацією виходів з обох напрямків, що дозволяє моделі використовувати інформацію з контексту з обох сторін вхідної послідовності; та 2-х повнозв'язних шарів Dense.

Наведено етапи навчання нейромережі на основі розробленої архітектури, результатом виконання вказаних етапів є навчена збережена нейромережа BiLSTM.

Наведено проектну архітектуру інформаційної системи виявлення кібербулінгу в текстових дописах, що публікуються в соціальних інтернет-мережах. Архітектура складається із 4-х підсистем: «Підсистеми навчання нейромережі», «Підсистеми препроцесингу», «Підсистеми аналізу датасету» та головної «Підсистеми виявлення кібербулінгу», а також набору даних і навченої нейромережевої моделі BiLSTM. Подана архітектура включає в себе усі етапи та компоненти, які забезпечують реалізацію і апробацію запропонованого методу виявлення кібербулінгу в текстових дописах.

Здійснено підготовку навчальних даних. Для навчання нейромережі BiLSTM буде використано набір даних Cyberbullying Classification, що налічує 39747 зразків, з яких буде обрано 8000 зразків, що містять кібербулінг, та 7945 записів без кібербулінгу.

Наведено особливості використання спеціалізованих програмних компонентів, а саме для створення інтерфейсів користувача буде використано бібліотеку Kivy, яка надає можливості для створення інтерактивних програм з графічним інтерфейсом, які можуть працювати на різних операційних системах. Для побудови графіків результатів навчання нейромережі та статистики вмісту датасету буде використано Matplotlib. Для попередньої обробки текстових даних датасету для виявлення кібербулінгу буде використано бібліотеку NLTK, а для навчання нейромережі буде використано бібліотеку TensorFlow.

В подальшому за методом виявлення кібербулінгу в дописах соціальних інтернет-мереж необхідно розробити застосунок, за допомогою якого провести дослідження ефективності розробленого методу. Для доведення коректності результатів метод потрібно окремо функціонально дослідити й протестувати.

Розділ 3 Експериментальне дослідження методу виявлення кібербулінгу в дописах соціальних інтернет-мереж

3.1 Визначення шляхів дослідження та засобів створення інформаційної системи виявлення кібербулінгу

Експериментальне дослідження методу виявлення кібербулінгу в дописах соціальних інтернет-мереж є важливим кроком для оцінки ефективності та точності розробленого методу. Для цього дослідження планується створення програмної системи з функціоналом, спрямованим на аналіз текстових даних з соціальних мереж з метою виявлення ознак кібербулінгу. Інформаційна система буде включати в себе модулі для попередньої обробки тексту, класифікації за допомогою нейромережевої моделі, а також модулі для аналізу результатів.

Дослідження буде включати тестування розробленого методу на різних наборах користувацьких даних. Планується виконати порівняльний аналіз результатів роботи розробленого методу з результатами роботи існуючих методів виявлення кібербулінгу. Параметри для порівняння будуть обрані на основі важливості для ефективного виявлення кібербулінгу, таких як точність та F1-міра.

Основними метриками, які будуть використовуватися для оцінки ефективності розробленого методу, будуть метрики класифікації, такі як точність та F1-міра. Крім того, планується виконання аналізу навчальних і валідаційних даних.

3.2 Вибір засобів розробки інформаційної системи виявлення кібербулінгу

Для реалізації інформаційної системи виявлення кібербулінгу буде використано мову програмування Python, так як вона має велику кількість спеціалізованих програмних компонентів для реалізації за стосунків на базі ШІ. Для навчання нейромережі буде використано Google Colab, що є безкоштовним

хмарним сервісом, а для побудови інтерфейсів користувача буде використано середовище програмування PyCharm.

Python вважається однією з найпопулярніших та ефективних мов програмування для реалізації нейромереж і штучного інтелекту взагалі, завдяки декільком ключовим перевагам. Спільнота Python має доступ до широкого вибору бібліотек та фреймворків для глибокого навчання, включаючи такі популярні інструменти, як TensorFlow, PyTorch, Keras і інші. Ця різноманітність дозволяє вибирати ті інструменти, які найкраще відповідають конкретним завданням і сприяє швидкому розвитку нових рішень у галузі штучного інтелекту [28].

Однією з ключових переваг Python є його лаконічний та легкий для читання синтаксис, що полегшує розробку та відлагодження коду. Це робить мову привабливою для широкого кола розробників, включаючи тих, хто не є експертами у галузі машинного навчання. Більше того, Python є мультиплатформеною мовою, що дозволяє розробникам реалізовувати та запускати нейромережеві застосунки на різних операційних системах, що сприяє універсальності та доступності розроблених рішень.

Для навчання нейромережі буде використано Google Colab, що є безкоштовний сервіс від Google, який надає можливість запускати та навчати нейронні мережі в хмарному середовищі, використовуючи ресурси від Google, включаючи графічний процесор та графічний процесор Tensor Processing Unit. Основні переваги Google Colab полягають у тому, що він безкоштовний, має широкий вибір установлених бібліотек для машинного навчання та глибокого навчання, включаючи TensorFlow, PyTorch, Keras, інші, а також в доступності інтерфейсу через веб-браузер [29].

Завдяки використанню обчислювальних ресурсів Google, включаючи GPU та TPU, Colab дозволяє навчати складні моделі нейронних мереж на великих даних без необхідності інвестування в власне обладнання. Крім того, Colab надає можливість спільного користування ноутбуками, що дозволяє спільно працювати над проектами з колегами або ділитися результатами з

іншими користувачами. Даний ресурс буде використано для навчання та збереження нейромережі BiLSTM.

Для створення інтерфейсу користувача буде використано середовище розробки PyCharm, що є інтегрованим середовищем розробки для мови програмування Python, розроблене компанією JetBrains. PyCharm надає розширені можливості для розробки програм на Python, забезпечуючи розумне автодоповнення коду, підказки та аналіз помилок, що допомагає підвищити продуктивність розробника [30].

Основні функції та переваги PyCharm включають розширені можливості рефакторингу коду, інтеграцію з системами контролю версій (наприклад, Git), підтримку віртуальних середовищ (наприклад, virtualenv), а також інтеграцію з популярними фреймворками та бібліотеками Python, такими як Django, Flask, NumPy, Pandas тощо.

Отже, для програмної реалізації інформаційної системи виявлення кібербулінгу буде використано мову програмування Python, для навчання нейромережі буде використано Google Colab, що є безкоштовним хмарним сервісом, а для побудови інтерфейсів користувача буде використано середовище програмування PyCharm.

3.3 Структура та функціональне призначення програмних складових інформаційної системи виявлення кібербулінгу

Структура програмних складових інформаційної системи виявлення кібербулінга у текстових дописах соціальних інтернет-мереж наведена на рисунку 3.1.

Клас «MainMenuLayout» є підкласом «BoxLayout». Він представляє головне меню інформаційної системи з виявлення кібербулінгу. Головне меню містить три кнопки: «Cyberbullying Detection Subsystem», «Preprocessing Subsystem» і «Dataset Analysis Subsystem».

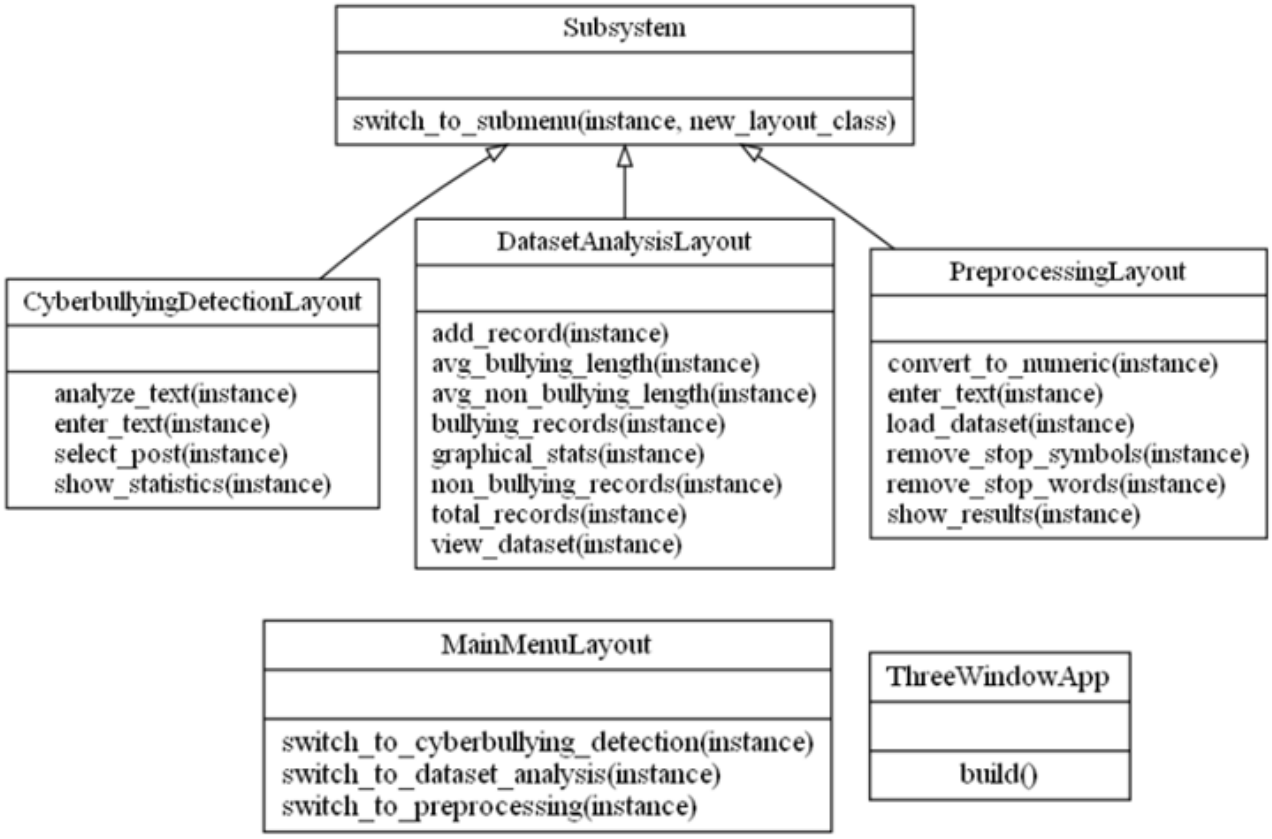


Рисунок 3.1 – Діаграма класів інформаційної системи виявлення кібербулінгу

Базовий клас «Subsystem» містить метод switch_to_submenu, що дозволяє змінювати вміст головного вікна на вміст підсистеми. Кожен з класів підсистем успадковує цей клас і має свої власні методи та вміст.

Клас «CyberbullyingDetectionLayout» є підкласом Subsystem. Цей клас відповідає за підсистему виявлення кібербулінгу. Клас має на меті реалізацію функціоналу з вибіру допису для аналізу з датасета (метод «select_post»), аналіз текстового представлення з метою виявлення наявності кібербулінгу (метод «analyze_text»), виведення статистики виконаного аналізу щодо наявності кібербулінгу (метод «show_statistics»).

Клас «PreprocessingLayout» є підкласом «Subsystem». Цей клас відповідає за підсистему попередньої обробки тексту. Метод «load_dataset» відповідає за завантаження датасета з CSV файлу. Метод «remove_stop_symbols» відповідає за видалення стоп-символів з тексту. Виконує операції видалення певних символів з тексту, які не несуть семантичної інформації. Метод «remove_stop_words» відповідає за видалення стоп-слів з тексту. Виконує операцію видалення слів, які

є дуже поширеними та не несуть значення для аналізу. Метод «convert_to_numeric» відповідає за перетворення тексту в числове представлення. Метод «show_results» відповідає за відображення результатів попередньої обробки тексту.

Клас «DatasetAnalysisLayout» є підкласом «Subsystem». Цей клас відповідає за підсистему аналізу набору даних. Метод «view_dataset» відповідає за перегляд набору даних. Метод «add_record» відповідає за додавання нового запису до набору даних. Метод «total_records» відображає загальну кількість записів у наборі даних. Метод «bullying_records» відображає кількість записів, що містять кібербулінг, а метод «non_bullying_records» відображає кількість записів, що не містять кібербулінг. Метод «avg_bullying_length» відображає середню довжину записів, що містять кібербулінг, а метод «avg_non_bullying_length» відображає середню довжину записів, що не містять кібербулінг. Метод «graphical_stats» відображає графічне представлення статистики.

Отже, було створено структуру та описано функціональне призначення складових інформаційної системи виявлення кібербулінгу. Створена структура складається із 3-х основних підсистем: виявлення кібербулінгу, препроцесингу та аналізу вмісту датасету, що спроектована відповідними класами та у подальшому буде реалізована.

3.4 Особливості реалізації програмних складових інформаційної системи виявлення кібербулінгу

Оскільки однією зі складових частин інформаційної системи виявлення кібербулінгу є навчена нейромережа, спершу було виконано її навчання та збереження. Для навчання нейромережі використано хмарний сервіс Google Colab, що дозволяє навчати нейромережу у хмарному середовищі, використовуючи ресурси від Google не встановлюючи програмні компоненти на локальний комп’ютер. Нейромережа навчалась за створеною вище архітектурою

та описаною вище послідовністю. Етап завантаження датасету наведено на рисунку 3.2.

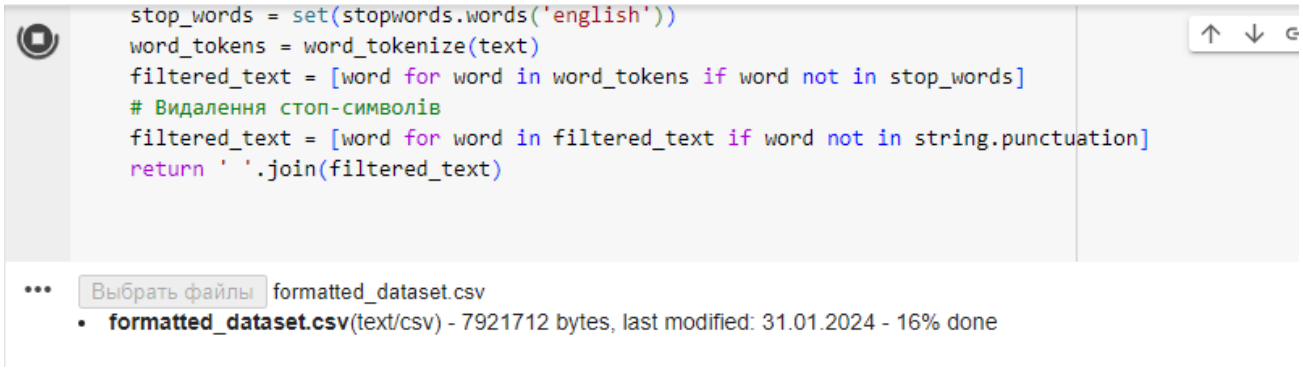


Рисунок 3.2 – Етап завантаження датасета

Етап старту процесу навчання наведено на рисунку 3.3.

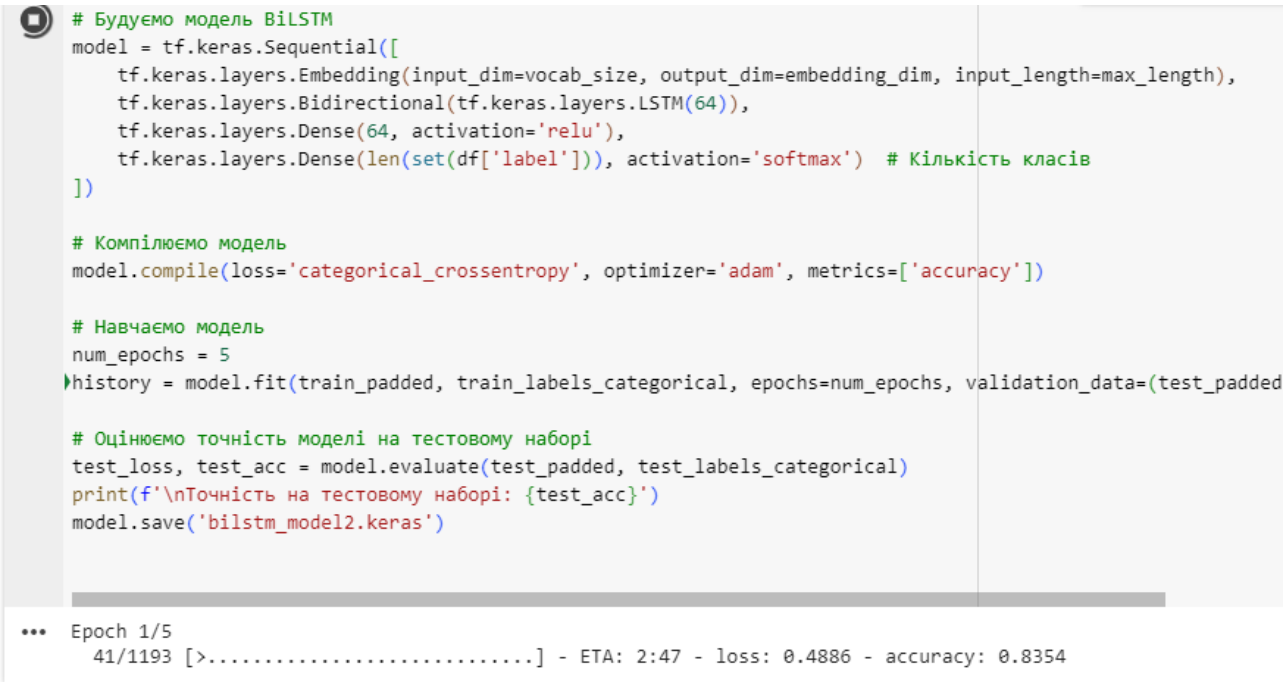


Рисунок 3.3 – Процес навчання нейромережі

По завершенні навчання нейромережа була завантажена на жорсткий диск, та була виведена статистика змін точності та функції втрат за епохами (рисунок 3.4). Надалі нейромережа використовувалась як складова частина інформаційної системи виявлення кібербулінгу, що має графічний інтерфейс.

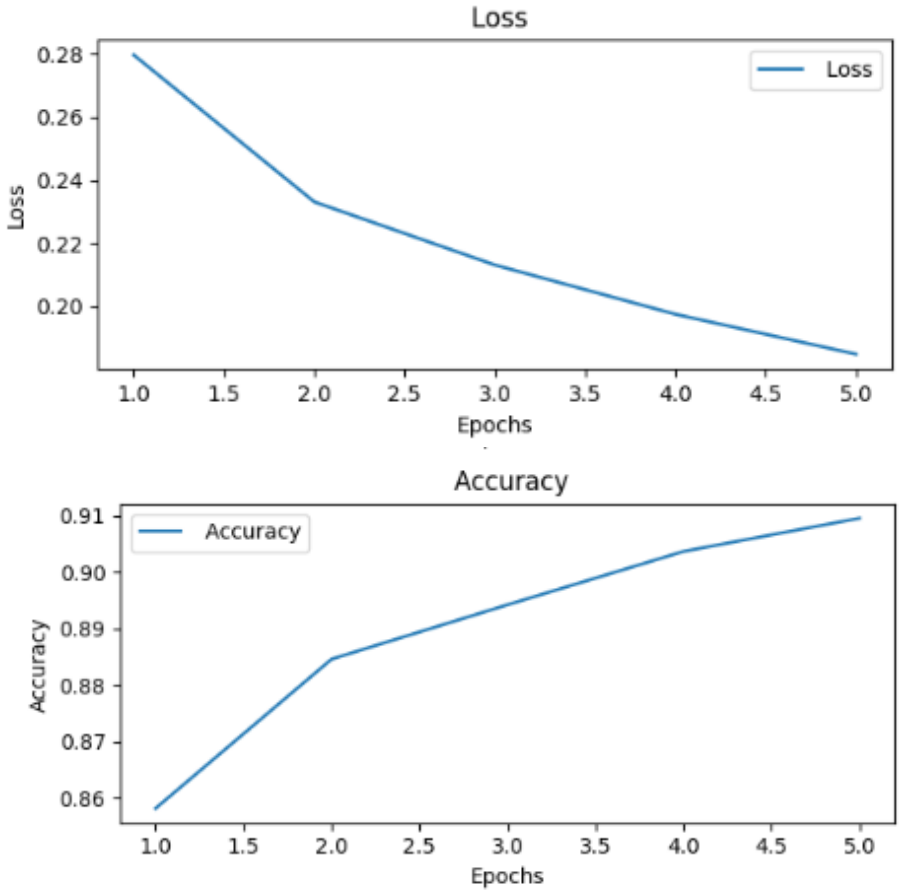


Рисунок 3.4 – Статистика показників Loss та Accuracy за епохами

Для використання у застосунку з віконним інтерфейсом навченої моделі використовуються можливості бібліотеки Keras для завантаження моделі нейронної мережі з файлу. Також, окрім самої нейромережі завантажується токенізатор. Підсистема виявлення кібербулінгу може як визначати кібербулінг з обраного користувачем повідомлення з набору даних, так і писати текст для перевірки на наявність кібербулінгу власноруч. Інтерфейс підсистеми виявлення кібербулінгу наведено на рисунку 3.5.

Для визначення кібербулінгу користувацький текст спершу проходить попередню обробку, що включає в себе видалення символів пунктуації та стоп слів. Стоп-слова завантажуються з набору стоп-слів для англійської мови за допомогою бібліотеки NLTK, текст розбивається на окремі слова за допомогою word_tokenize і кожне слово перевіряється, чи не є воно стоп-словом. Якщо слово не належить до списку стоп-слів, то воно додається до нового списку. Далі

текст подається векторизатору для подальшої обробки нейромережею. Приклад виведення результату обробки наведено на рисунку 3.6.

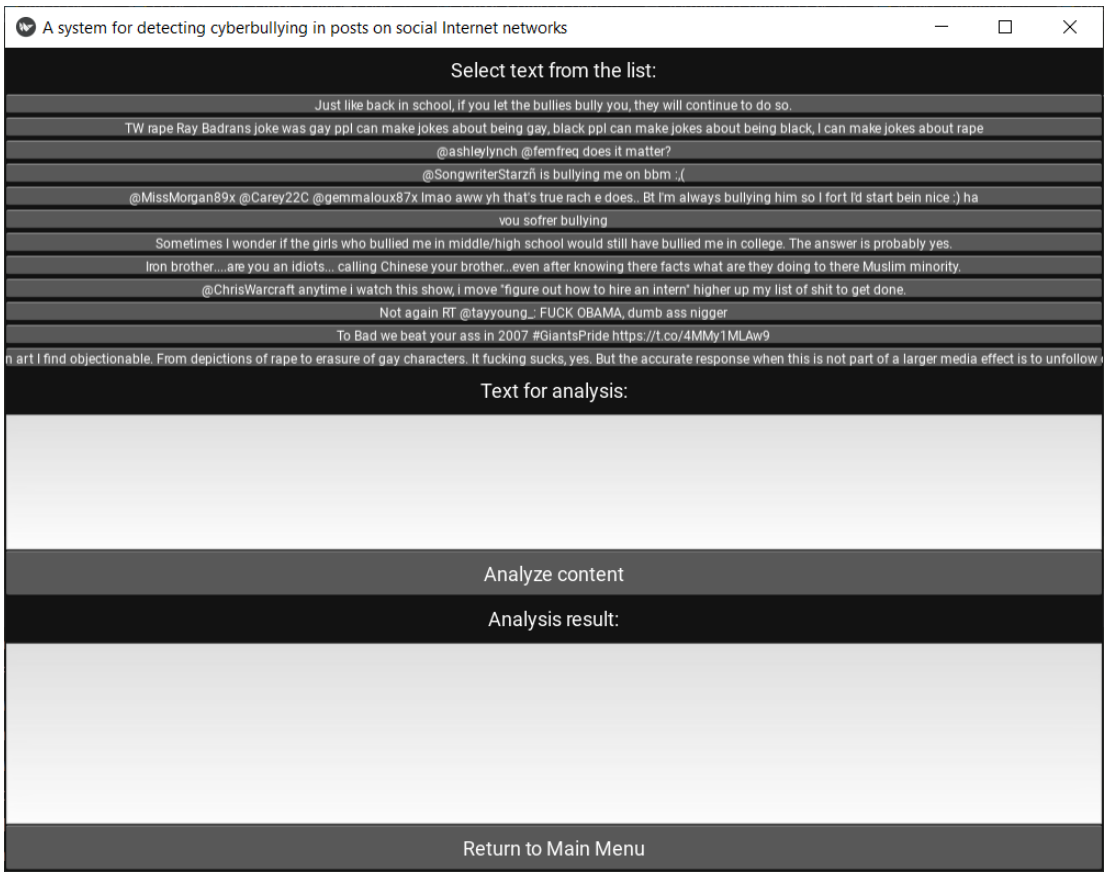


Рисунок 3.5 – Інтерфейс підсистеми по визначенняю кібербулінгу

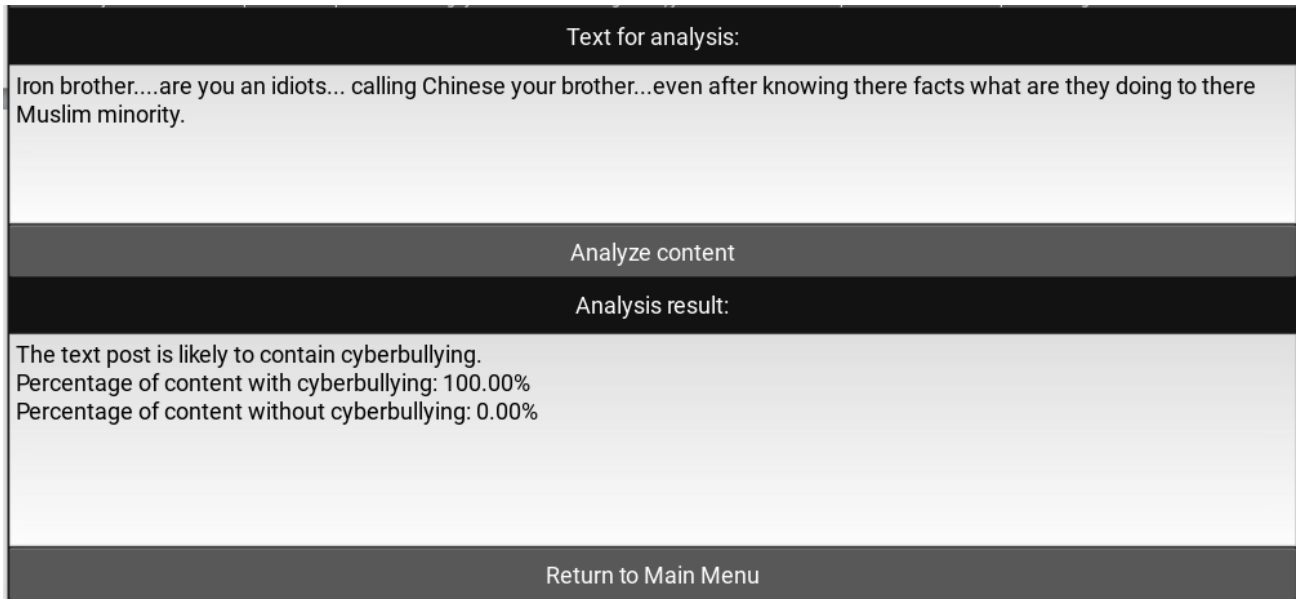


Рисунок 3.6 – Вивід результату дослідження тексту з кібербулінгом

Отже, були наведені особливості реалізації інформаційної системи виявлення кібербулінгу, що використовує нейромережевий підхід.

3.5 Тестування інформаційної системи та вимоги до розгортання

Створена інформаційна система виявлення кібербулінгу за текстовим дописом соціальних інтернет-мереж протестована засобами тест-кейсів з метою виявлення у подальшому можливих помилок в роботі.

Є потреба дослідити ефективність підсистеми з навчання нейромережі, а саме її основного результату, можливості збереження навчених нейромережевих моделей у файловій системі. Для цього був створений відповідний тест-кейс, кроки якого описані у таблиці 3.1

Таблиця 3.1 – Тест-кейс 00001

<table>
<tr><td>Тест-кейс ID: 00001</td><td>Приорітет: 1</td><td>Створено: 4.03.2024, Владислав АНДРОЩУК</td></tr>
<tr><td colspan="3">Назва: Перевірка збереження навчених моделей у файловій системі</td></tr>
<tr><td colspan="2">Кроки</td><td>Очікуваний результат</td></tr>
<tr><td colspan="2">1. Відкрити програмний модуль підсистеми для навчання нейромережі
2. Встановити параметри кількості епох в 15.
3. Виконати запуск блокнота</td><td>Відкрито хмарний сервіс Google Colab

Проходження процесу навчання у вигляді проходження епох в Google Colab, збереження навченої моделі по завершенню навчання.</td></tr>
<tr><td colspan="3">Результат виконання тест-кейсу: перевірку пройдено успішно.</td></tr>
</table>

Після запуску блокнота та виконання кроків, що описані у таблиці 3.1, можна переконатись, що після завершення навчання модель дійсно зберігається у файловій системі. Результат зображено на рисунку 3.7.

Рисунок 3.7 – Збереження навченої моделі в файловій системі

Таблиця 3.2 – Тест-кейс 00002

Тест-кейс ID: 00002	**Приорітет:** 1	**Створено:** 5.03.2024, Владислав АНДРОЩУК
Назва: Перевірка виведення графіків статистики навчання за епохами		
Кроки		**Очікуваний результат**
1. Відкрити програмний модуль підсистеми для навчання нейромережі 2. Встановити параметри кількості епох в 15. 3. Виконати запуск блокнота 4. Перевірити наявність 2-х графіків, що відображають значення функції втрат та точності за епохами.		Відкрито хмарний сервіс Google Colab Проходження процесу навчання у вигляді проходження 15 епох в Google Colab. Виведення графіків точності та втрат
Результат виконання тест-кейсу: перевірку пройдено успішно.		

Наступним тестовим випадком буде дослідження коректності функції виведення статистики з навчання у вигляді діаграм підсистеми з навчання

нейромережі. Для цього був створений відповідний тест-кейс, кроки якого описані у таблиці 3.2.

Після запуску блокнота та виконання кроків, що описані у таблиці 3.2, можна переконатись, що після завершення навчання моделі будуть виведені графіки зі статистики навчання. Результат зображено на рисунку 3.8.

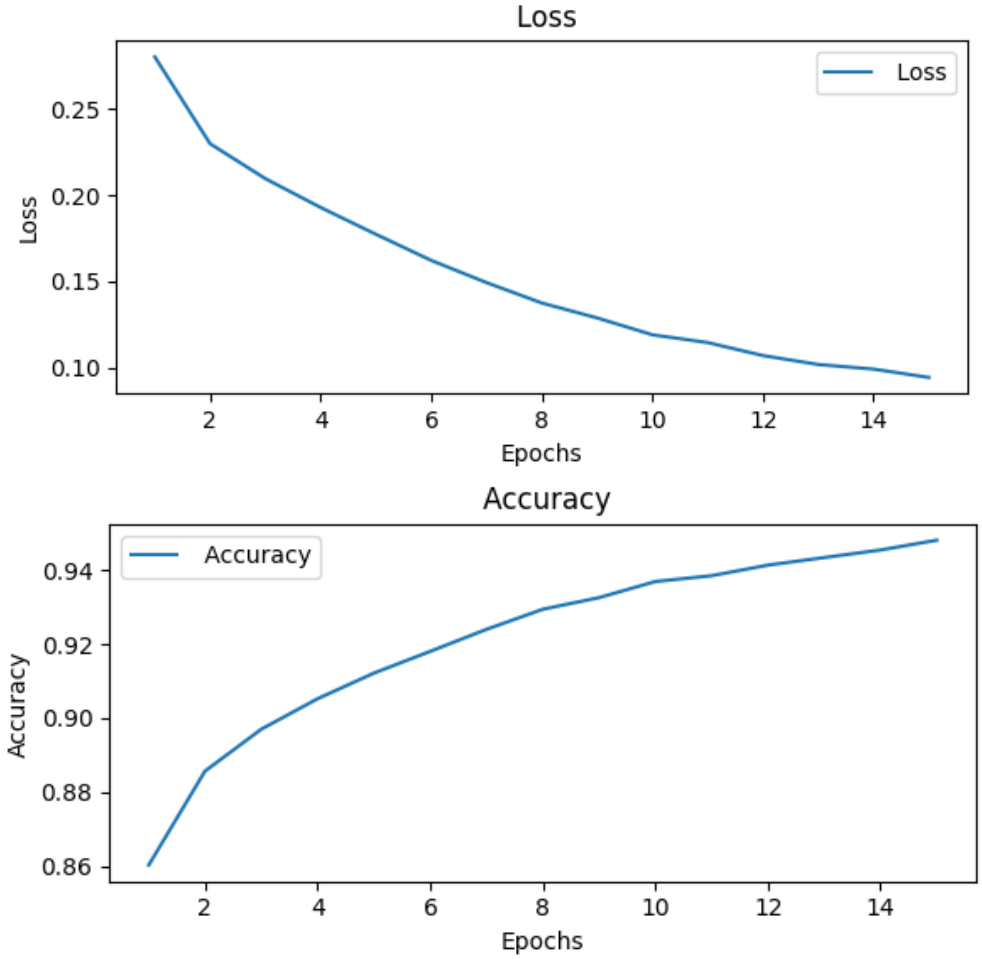


Рисунок 3.8 – Графіки точності та втрат

Наступним тестовим випадком буде дослідження коректності функції аналізу датасету підсистеми аналізу навчальних даних. Для цього був створений відповідний тест-кейс, кроки якого описані у таблиці 3.3.

Таблиця 3.3 – Тест-кейс 00003

Тест-кейс ID: 00003	**Приорітет:** 1	**Створено:** 5.03.2024, Владислав АНДРОЩУК
Назва: Перевірка виведення статистики аналізу з навчального набору даних		
Кроки		**Очікуваний результат**
1. Відкрити затсосунок 2. Перейти на підсистему аналізу навчальних даних, натиснувши на кнопку «Training data analyzes subsystem». 3. Натиснути кнопку «Show Dataset Statistics» 4. Перевірити наявність графіка розподілу даних в датасеті по категоріях та числове виведення статистики по довжині постів по категоріях.		Відкрився застосунок Виконано перехід на підсистему «Training data analyzes subsystem» Виведено графік та статистику
Результат виконання тест-кейсу: перевірку пройдено успішно.		

Після запуску віконного застосунку та виконання кроків, що описані у таблиці 3.3, можна переконатись, що буде виведено статистику розподілу даних в датасеті та виведено відповідну графічну гістограму. Результат зображено на рисунку 3.9.

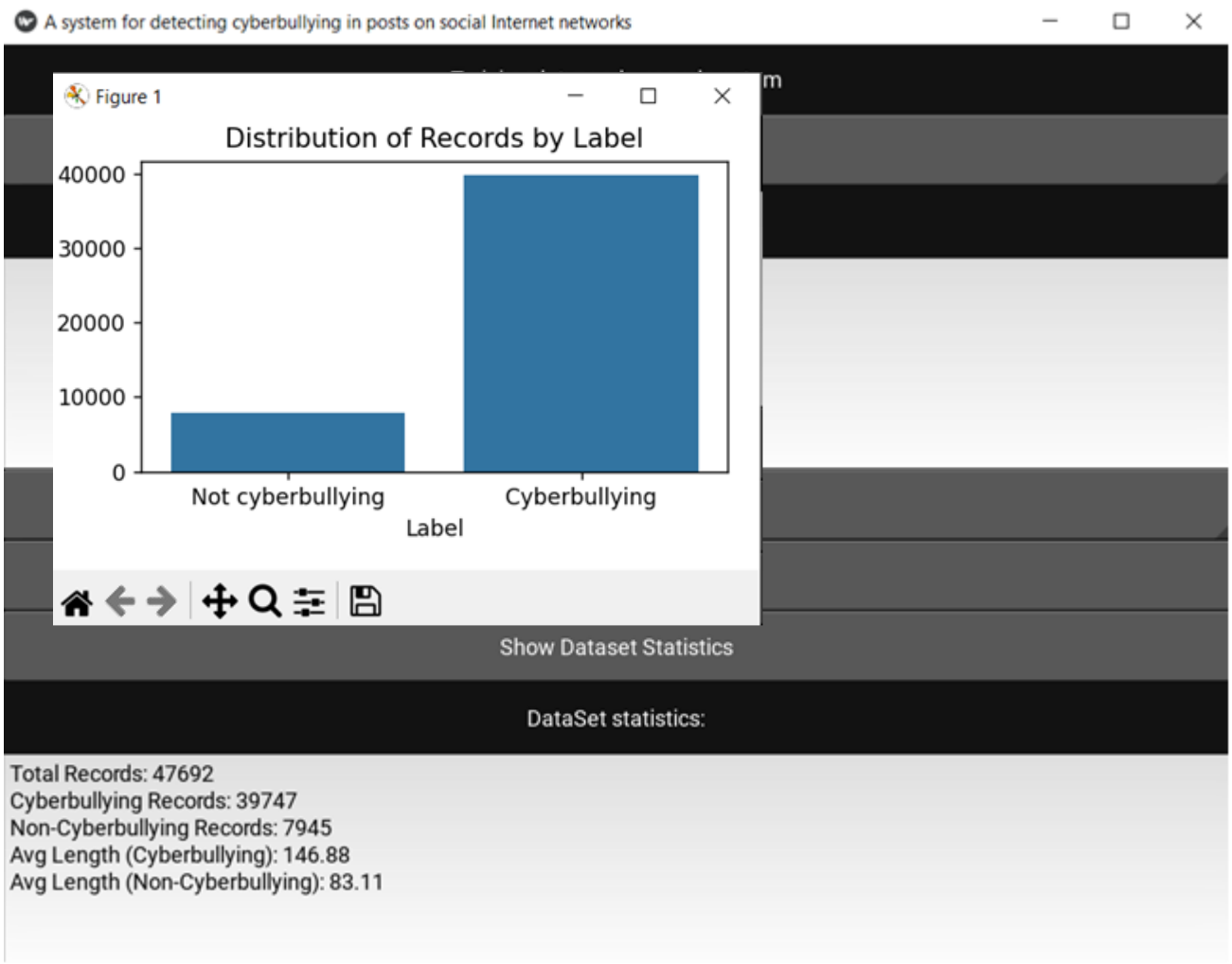


Рисунок 3.8 – Виведення статистики вмісту навчальних даних

Отже, з проведеного тестування програмної реалізації інформаційної системи, заявлений функціонал працює коректно, непрацюючих функцій не виявлено.

3.6 Аналіз функціональності інформаційної системи виявлення кібербулінгу

Після тестування інформаційної системи виявлення кібербулінгу необхідно провести аналіз її функціональності. Застосунок для виявлення кібербулінгу має віконну реалізацію, а також підсистему для навчання нейромережі, яка не має графічного інтерфейсу і є допоміжною підсистемою. Спершу будуть розглянуті особливості функціональності застосунку з графічним

інтерфейсом користувача і перша форма це є форма головного меню, що наведена на рисунку 3.9.

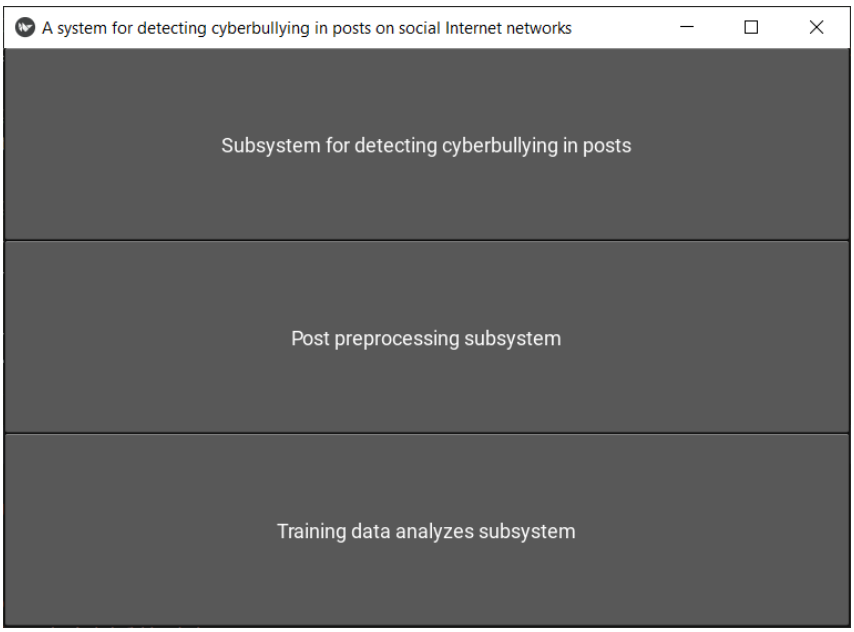


Рисунок 3.9 – Головне меню інформаційної системи виявлення кібербулінгу

З головного меню можна здійснити перехід на одну із 3-х підсистем, що також мають графічний інтерфейс. Це «Підсистема виявлення кібербулінгу», «Підсистема препроцесингу» та «Підсистема аналізу датасету». Для переходу на «Підсистему препроцесингу» необхідно натиснути кнопку «Post preprocessing subsystem», відбудеться відповідний перехід на форму (рисунок 3.10).

Дана підсистема спрямована на розуміння користувача, яким чином здійснюється підготовка даних для навчання нейромережі та для ідентифікації кібербулінгу вже навченими екземплярами нейромережі BiLSTM. На рисунку 3.10 наведено приклад виконання функції видалення пунктуації. Для виконання решти заявленого функціоналу необхідно натиснути на однойменні кнопки.

Також з головного меню інформаційної системи виявлення кібербулінгу можна перейти на підсистему виявлення кібербулінгу. Для цього необхідно натиснути на кнопку «Subsystem for detecting cyberbullying in posts». Дана форма є головною підсистемою, інтерфейс наведено на рисунку 3.11.

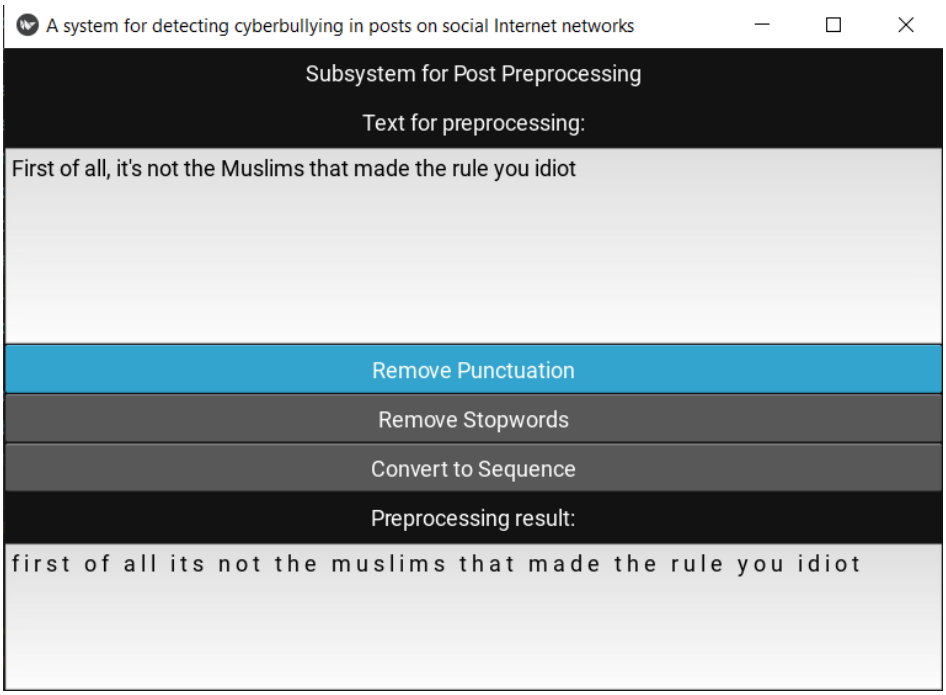


Рисунок 3.10 – Підсистема препроцесингу

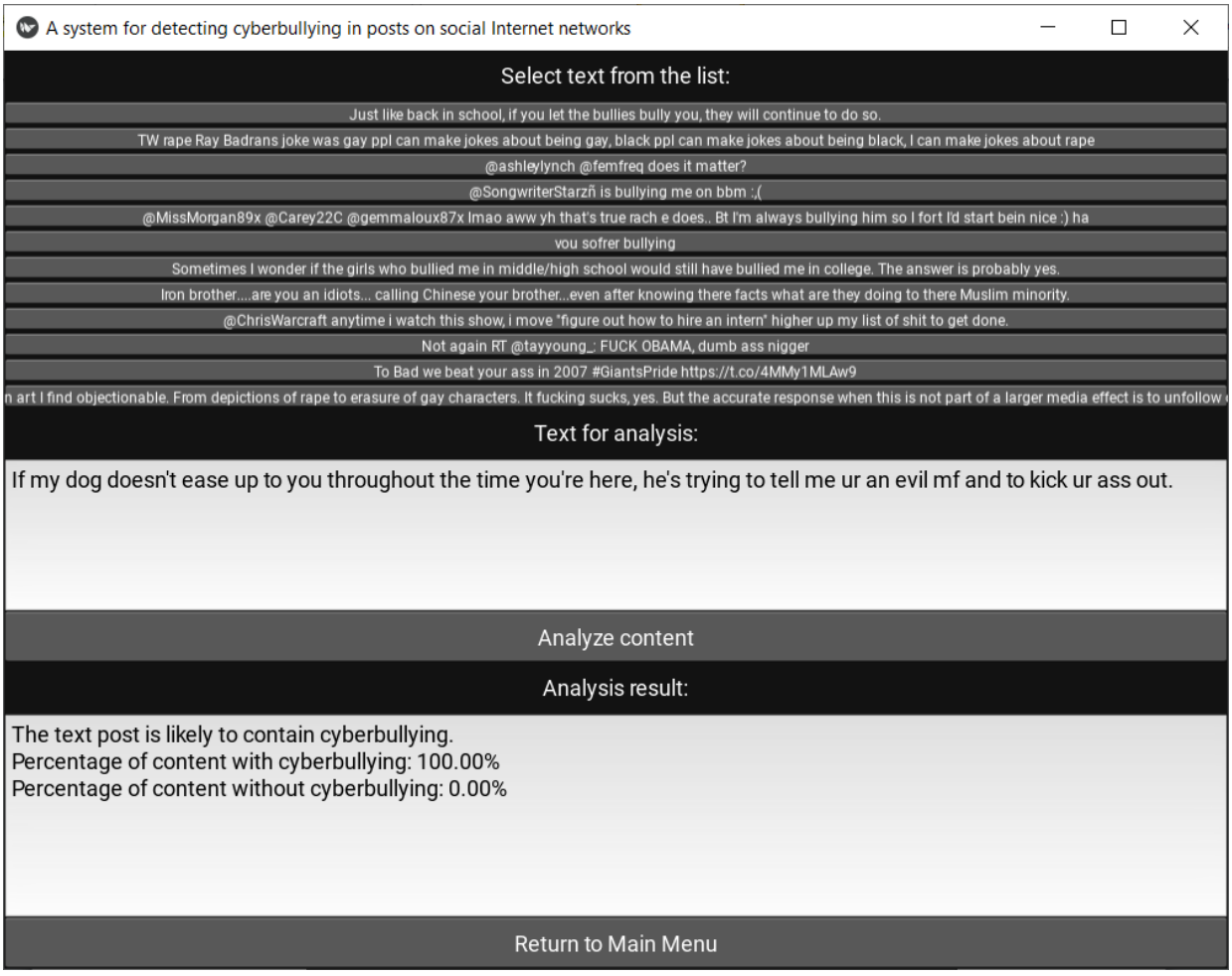


Рисунок 3.11 – Підсистема виявлення кібербулінгу

Для використання даної підсистеми можна обрати пост для аналізу з датасету, скориставшись частиною екрану з перерахованими твітами в полі «Select text from list:»б, або ж написати текст власноруч, увівши його у поле «Text for analisis». Для отримання прогнозу стосовно вмісту кібербулінгу, необхідно натиснути кнопку «Analyze content». Результат виконання аналізу контенту наведено на рисунку 3.11 у полі «Analysis result:».

Для використання функціоналу роботи з датасетом необхідно натиснути кнопку головного меню «Training data analyzes subsystem». Екран форми наведено на рисунку 3.12.

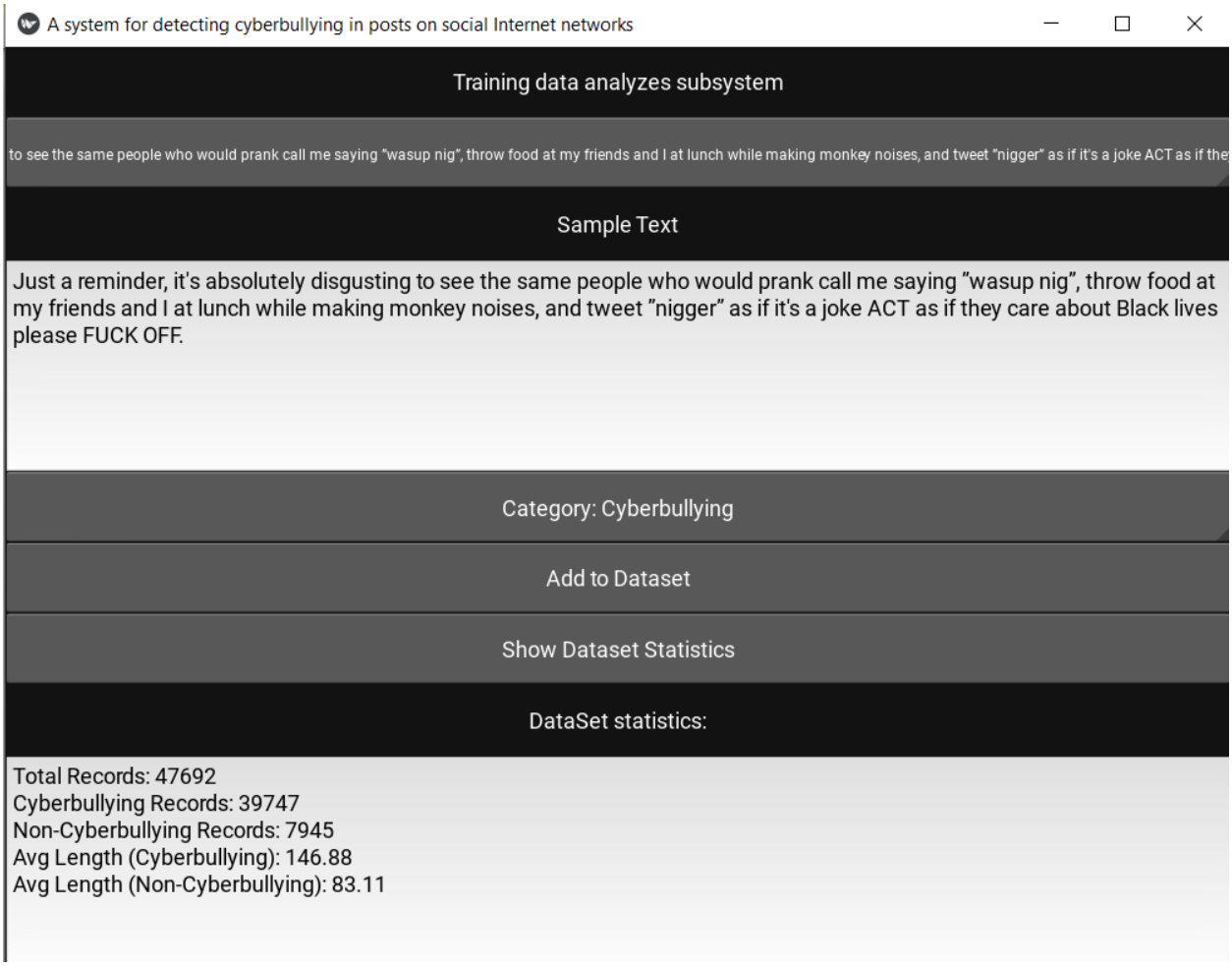


Рисунок 3.12 – Форма підсистеми аналізу датасету

Тут є можливість як переглянути вже існуючі записи, натиснувши на випадаючий список вгорі екрану, після чого відкриється перелік для відображення тексту (рисунок 3.13), так і є можливість додавати дані в датасет,

увівши у текстове поле «Sample Text» текст, та обравши у випадаючому списку Category відповідну категорію, після чого натиснувши кнопку «Add to Dataset».

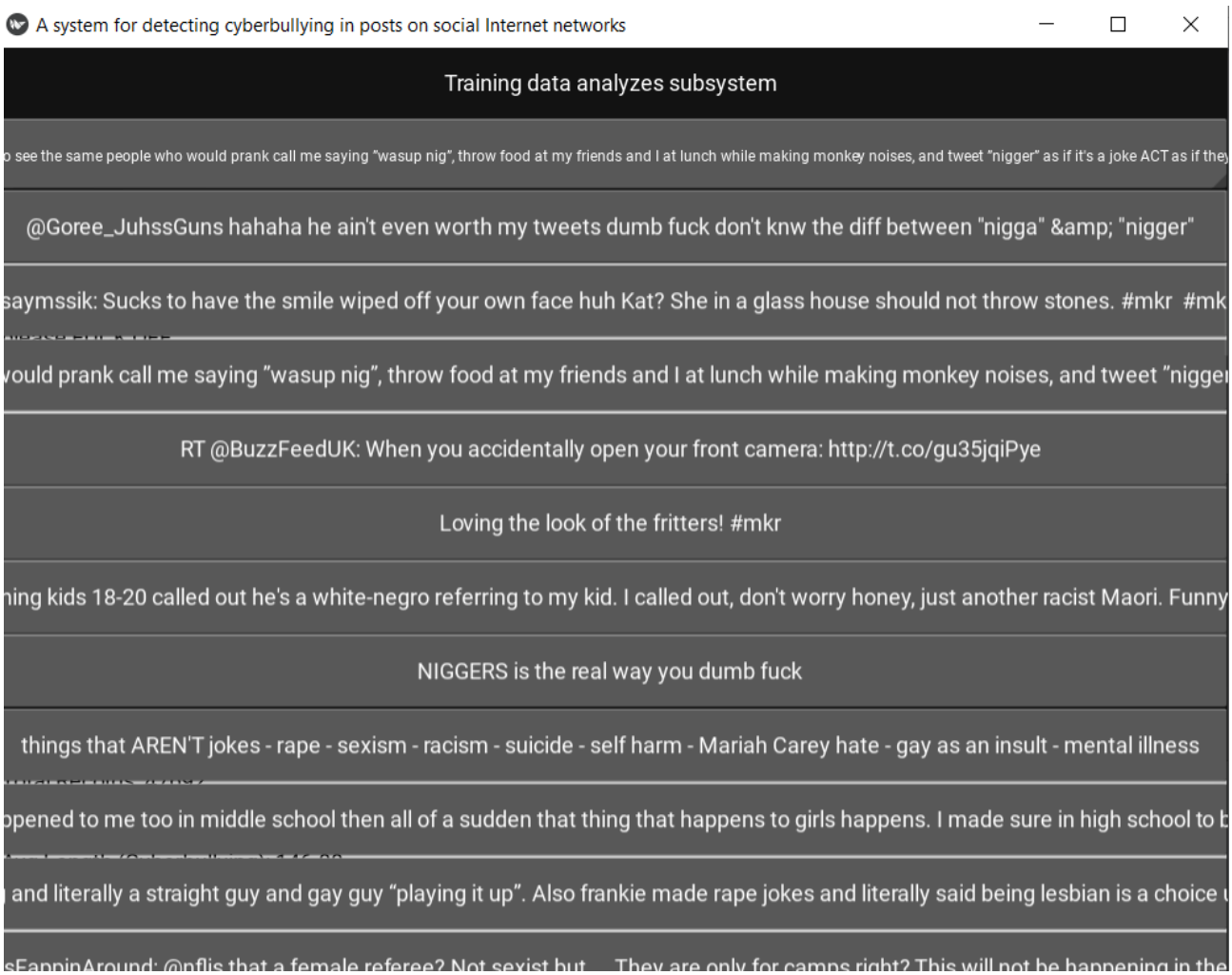


Рисунок 3.13 – Фрагмент випадаючого переліку з твітами для перегляду

Що стосується модулю для навчання нейромереж, то для розгортання даної підсистеми необхідно використати веббраузер, та перейти на сайт «https://colab.google/». Далі відкрити створений блокнот та по черзі виконати всі комірки з програмним кодом, змінюючи параметри кількості епох, кількості нейронів у шарах нейромережі та розмір батча для експериментів (рисунок 3.14). Для запуску комірок блокнота використовується знак запуску у верхньому лівому кутку.

```
model = tf.keras.Sequential([
    tf.keras.layers.Embedding(input_dim=vocab_size, output_dim=embedding_dim, input_length=max_length),
    tf.keras.layers.Bidirectional(tf.keras.layers.LSTM(64)),
    tf.keras.layers.Dense(64, activation='relu'),
    tf.keras.layers.Dense(len(set(df['label'])), activation='softmax')  # Кількість класів
])
model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])

num_epochs = 15
history = model.fit(train_padded, train_labels_categorical, epochs=num_epochs, validation_data=(test_padded, test_labels_categorical))

test_loss, test_acc = model.evaluate(test_padded, test_labels_categorical)
print(f'\nТочність на тестовому наборі: {test_acc}')
model.save('bilstm_model2.keras')
```

Рисунок 3.14 – Параметри нейромережі

Отже, висвітлено аналіз функціональності системи з виявлення кібербулінгу, що функціонує на базі методу виявлення кібербулінгу в дописах соціальних інтернет-мереж та складається із десктопного застосунку із 3-х підсистем з графічним інтерфейсом, блокноту для побудови нейромережевої моделі та збереження її на жорсткий диск та датасету.

3.7 Результати досліджень

Розроблена програмна реалізація дозволила успішно виявляти кібербулінг у англомовних дописах соціальних інтернет-мереж. Відсоток точності за метрикою Accuracy на тренувальному наборі становив близько 95 %. Були проведені експерименти з навчання нейромережі, які мали на меті сприяти покращенню базового відсотку ідентифікації, зміни стосувались кількості нейронів в шарах, кількості епох навчання та довжини словника і вхідних послідовностей. Фрагмент з графічного представлення функції втрат та точності для 25 епох навчання наведено на рисунку 3.15.

Як видно з рисунку 3.15, функція втрат все ще продовжує спадати, а графік точності все ще має деяку тенденцію до зростання. Тому є потреба дослідити процес навчання для більшої кількості епох. Збільшивши кількість епох до 50 результат дав приріст на навчальній вибірці, склавши показник точності 0.97, проте на валідаційній вибірці приросту в точності не спостерігалось. Результати експерименту наведені в таблиці 3.4.

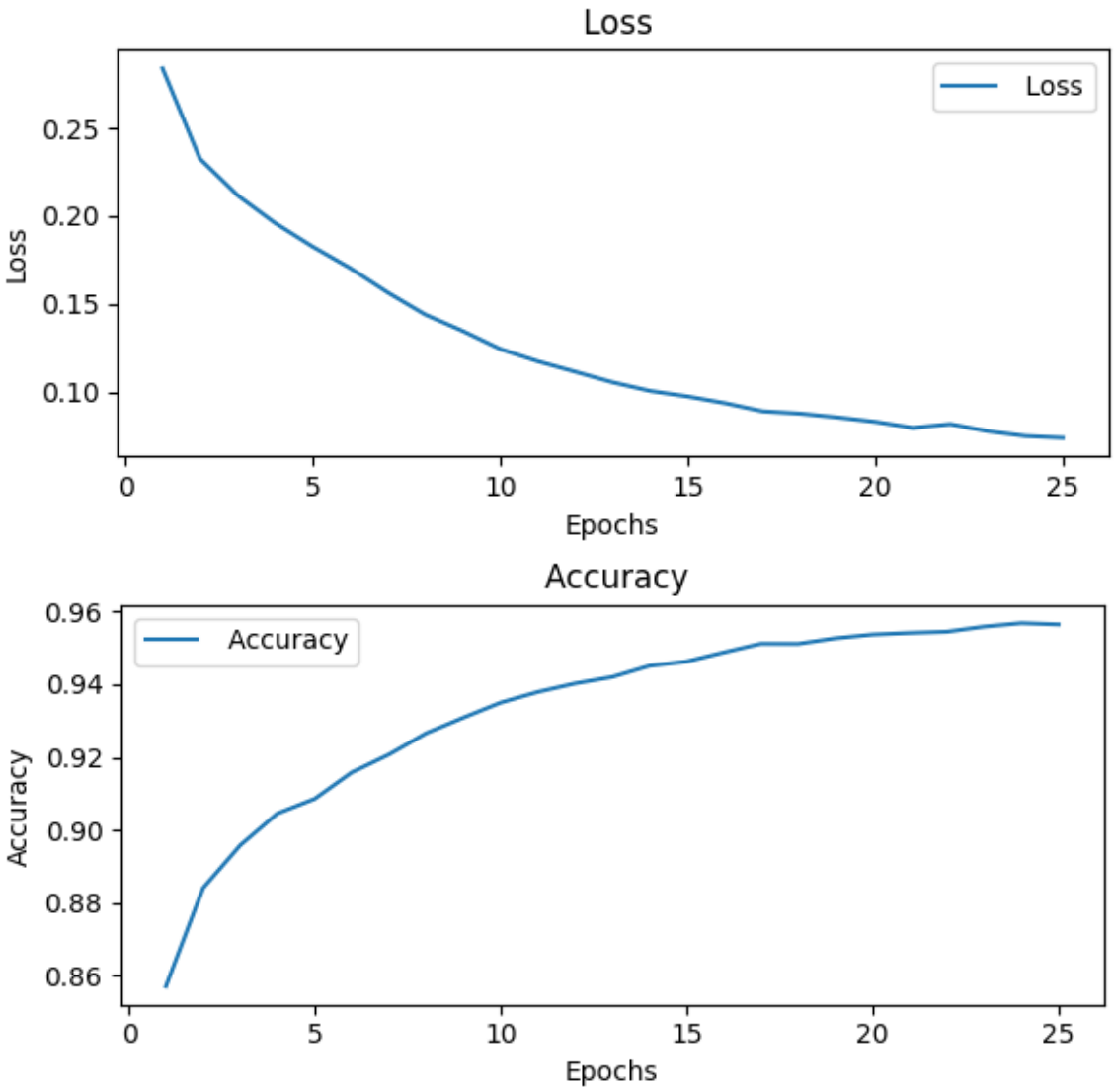


Рисунок 3.15 – Статистика навчання за 25-тьма епохами

По результатам експерименту побудовано графік (рисунок 3.16), що ілюструє значення метрик для вказаних в таблиці 3.4 варіантів дослідження.

З таблиці 3.4 та графіку 3.16 видно, що всі параметри знаходяться приблизно на одному рівні, однак версія BiLSTM з кількістю епох 25 та кількістю нейронів в шарах LSTM та Dense 64 і розміром словника в 5000 показала кращий результат, досягнувши значень 0.96 метрикою Accuracy та 0.92 метрикою F1-міри. При цьому функція втрат склала 0.11.

Таблиця 3.4 – Параметри нейромережі та результати за метриками

Параметри нейромережі:	**Accuracy**	**Loss**	**F1-міра**
К-сть епох: 5 К-сть нейронів в шарах LSTM та Dense: 64 Розмір словника: 5000 (BiLSTM 1)	0.90	0.17	0.89
К-сть епох: 7 К-сть нейронів в шарах LSTM та Dense: 64 Розмір словника: 5000 (BiLSTM 2)	0.91	0.14	0.9
К-сть епох: 15 К-сть нейронів в шарах LSTM та Dense: 64 Розмір словника: 5000 (BiLSTM 3)	0.92	0.12	0.91
К-сть епох: 25 К-сть нейронів в шарах LSTM та Dense: 64 Розмір словника: 5000 (BiLSTM 4)	0.96	0.11	0.92
К-сть епох: 15 К-сть нейронів в шарах LSTM та Dense: 128 Розмір словника: 5000 (BiLSTM 5)	0.93	0.12	0.92
К-сть епох: 15 К-сть нейронів в шарах LSTM та Dense: 32 Розмір словника: 5000 (BiLSTM 6)	0.94	0.15	0.91
К-сть епох: 15 К-сть нейронів в шарах LSTM та Dense: 32 Розмір словника: 3000 (BiLSTM 7)	0.91	0.18	0.87
К-сть епох: 15 К-сть нейронів в шарах LSTM та Dense: 128 Розмір словника: 3000 (BiLSTM 8)	0.92	0.16	0.91

Повторення експерименту на валідаційній вибірці дало дещо гірші результати, які становили 0.86 за метрикою Accuracy та 0.83 за метрикою F1-міри.

В подальшому було проаналізовано датасет на предмет довжини твітів. Проведений аналіз показав, що середня довжина твітів з кібербулінгом становила 146.88 символів, у той час як середня довжина твітів без кібербулінгу становила 83.11 символів. Розподіл твітів за довжиною для категорії «Cyberbullying» наведено на рисунку 3.17.

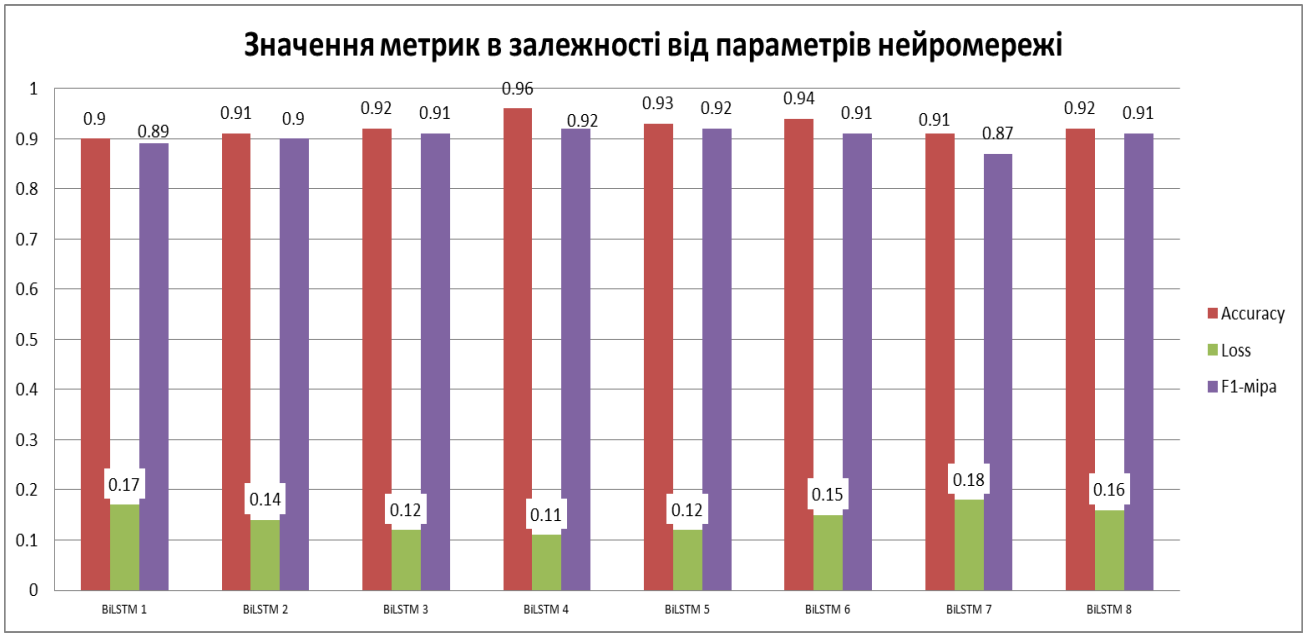


Рисунок 3.16 – Графік значень метрик в залежності від параметрів нейромережі

Розподіл твітів за довжиною для категорії «Not cyberbullying» наведено на рисунку 3.18.

Рисунок 3.17 – Розподіл твітів за довжиною для категорії «Cyberbullying»

Як видно з розподілів твітів за довжиною, твіти що містять кібербулінг є як правило, довшими. У той час як твіти без кібербулінгу здебільшого не перевищують значення в 250 символів.

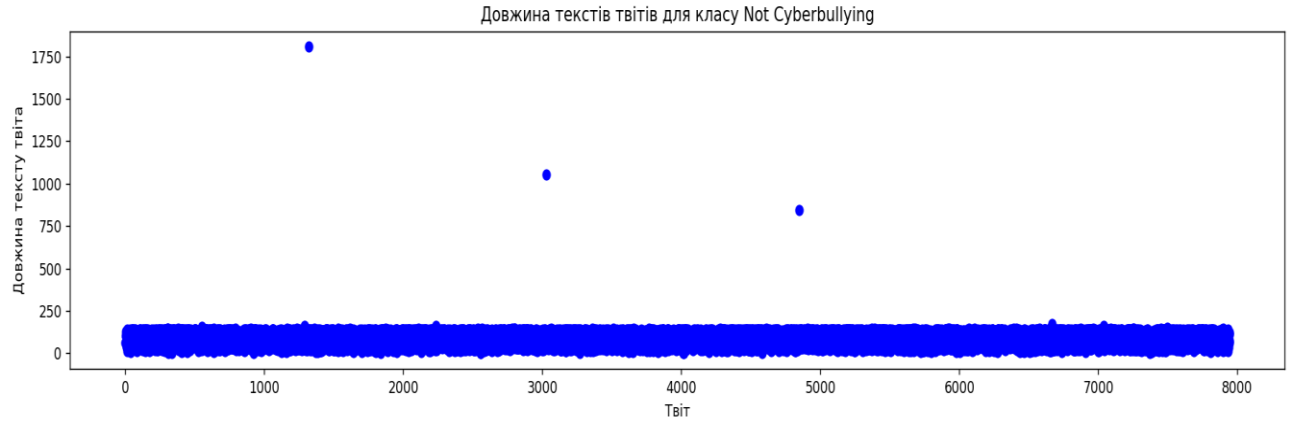


Рисунок 3.18 – Розподіл твітів за довжиною для категорії «Not cyberbullying»

Наступним кроком експерименту було відділення усіх твітів, що складаються менше ніж 6 слів, а також відділення твітів, що складаються із довжини більше 1000 символів. У свою чергу, це дещо зменшило розмірність вибірки, і вона стала становити 37995 твітів категорії «Cyberbullying» та 6884 твітів категорії «Not cyberbullying». Після використання даних маніпуляцій над віибіркою, нейромережу було заново перенавчено.

Як видно з рисунку 3.19 (а) та 3.19 (б), результати навчання покращились, і на 15-ти епохах точність склала 0.97, а в експерименті без фільтрації даних точність склала 0.92, у той час як функція втрат склала 0.05, на ряду з значенням 0.12 до фільтрації. Перевіривши дану модель на валідаційних даних, метрики також покращили значення, склавши 0.91 за метрикою Accuracy та 0.89 за метрикою F1-міри.

Виконавши перенавчання на 25 епохах, показники дещо зросли і склали 0.989 за метрикою Accuracy та 0.95 за метрикою F1-міри на навчальному наборі та 0.928 за метрикою Accuracy та 0.912 за метрикою F1-міри.

Також у подальшому варіант моделі з 25 епохами було збережено та досліджено із застосунку з графічним інтерфейсом. В цілому, такий розрив між навчальними та тестовими результатами пояснюється тим, що окрім простої фільтрації вибірки, не всі дані датасету мають коректну розмітку. Наприклад, зустрічаються твіти не лише англійською, а і португальською мовою: «*bullying é o oq passa o número 24 da caderneta de chamada !! #semgraça!*, Not cyberbullying».

Хоча і твіт має розмітку як не кібербулінг, проте, його переклад «*Знущання - це те, про що говорить номер 24 у книзі викликів!!*» не на стільки однозначний.

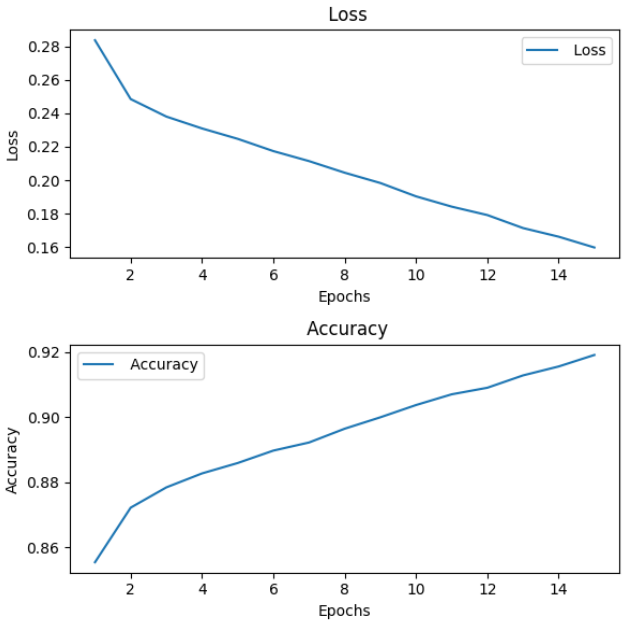


Рисунок 3.19 (а) – Значення метрик втрат та точності до фільтрації

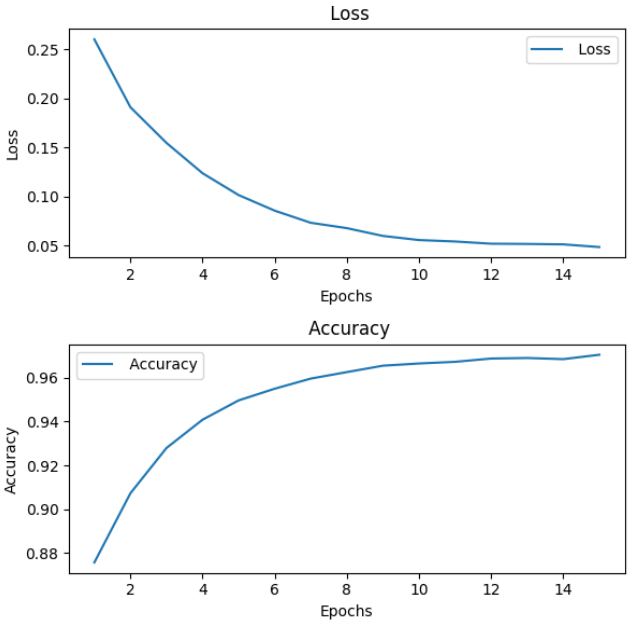


Рисунок 3.19 (б) – Значення метрик втрат та точності після фільтрації

Програмно даний твіт класифікується як кібербулінг на 55%.

Тому подальші дослідження будуть спрямовані на подальший аналіз датасету та його доповнення іншими даними, а також на покращення архітектури нейромережі з метою отримання більш точних результатів класифікації.

Отже, було досліджено ефективність запропонованого методу виявлення кібербулінгу у дописах соціальних інтернет-мереж. Запропонований метод показав високу ефективність, досягнувши показників 0.928 за метрикою Accuracy та 0.912 за метрикою F1-міри. Практичне тестування роботи моделі в межах інформаційної системи виявлення кібербулінгу показало її спроможність щодо ефективного виявлення кібербулінгу, та може використовуватись у вигляді надбудови до соціально-орієнтованих вебсервісів для протидії кібербулінгу.

3.8 Висновки до розділу 3

Було виконано експериментальне дослідження методу виявлення кібербулінгу в дописах соціальних інтернет-мереж, що є важливим кроком для оцінки ефективності та точності розробленого методу. У якості метрик, які будуть використовуватися для оцінки ефективності розробленого методу, були обрані метрики класифікації точність та F1-міра.

Для реалізації інформаційної системи виявлення кібербулінгу обрано мову програмування Python, для навчання нейромережі використано Google Colab, що є безкоштовним хмарним сервісом, а для побудови інтерфейсів користувача використано середовище програмування PyCharm.

Було створено структуру та описано функціональне призначення складових системи виявлення кібербулінгу, яка складається із 3-х основних підсистем: виявлення кібербулінгу, препроцесингу та аналізу вмісту датасету.

Наведені особливості реалізації інформаційної системи виявлення кібербулінгу, що використовує нейромережевий підхід. Для створеної програмної реалізації проведено тестування, яке підтвердило, що заявлений функціонал працює коректно, непрацюючих функцій не виявлено.

Висвітлено аналіз функціональності інформаційної системи виявлення кібербулінгу, що функціонує на базі методу виявлення кібербулінгу в дописах соціальних інтернет-мереж та складається із десктопного застосунку із 3-х підсистем з графічним інтерфейсом, блокноту для побудови нейромережевої моделі та збереження її на жорсткий диск та датасету.

Досліджено ефективність запропонованого методу, що показав високу ефективність, досягнувши показників 0.928 за метрикою Accuracy та 0.912 за метрикою F1-міри. Практичне тестування роботи моделі показало її спроможність щодо ефективного виявлення кібербулінгу, та може використовуватись у вигляді надбудови до соціально-орієнтованих вебсервісів для протидії кібербулінгу.

Загальні висновки

Метою кваліфікаційної роботи бакалавра було спрощення експертизи виявлення кібербулінгу за рахунок автоматизованого виявлення кібербулінгу в дописах соціальних інтернет-мереж.

Для досягнення поставленої мети були поставлені та вирішені такі завдання:

– Виконано аналіз предметної області виявлення кібербулінгу, в рамках якого з’ясовано, що виявлення і вчасна реакція на кібербулінг є важливим елементом захисту безпеки користувачів соціальних інтернет-мереж, тому автоматизація є актуальною задачею інформаційних технологій. Помічено, що сучасні технологічні досягнення, на кшталт машинного навчання, дають надію на підвищення ефективності виявлення та запобігання кібербулінгу. Тому в рамках роботи з розглянутих теоретичних підходів було обрано нейромережевий підхід, а саме нейромережу з архітектурою BiLSTM.

– Створено метод виявлення кібербулінгу в дописах соціальних інтернет-мереж, який призначений для перетворення вхідних даних у вигляді векторизатора, яким оброблялись дані під час навчання нейромережі, збереженої нейромережевої навченої моделі для виявлення кібербулінгу та користувацького допису для аналізу вмісту кібербулінгу у вихідні дані у вигляді висновку стосовно наявності кібербулінга в дописі для аналізу та відсоткового значення ймовірності приналежності даних до класу кібербулінга.

– Створено нейромережеву архітектуру для методу виявлення кібербулінгу в дописах. Наведено етапи навчання нейромережі на основі розробленої архітектури та підготовлено навчальні дані.

– Наведено проектну архітектуру інформаційної системи виявлення кібербулінгу в текстових дописах, що публікуються в соціальних інтернет-мережах. Архітектура складається із 4-х підсистем: «Підсистеми навчання нейромережі», «Підсистеми препроцесингу», «Підсистеми аналізу датасету» та головної «Підсистеми виявлення кібербулінгу», а також набору даних і навченої

нейромережевої моделі BiLSTM. Подана архітектура включає в себе усі етапи та компоненти, які забезпечують реалізацію і апробацію запропонованого методу виявлення кібербулінгу в текстових дописах.

– Створено відповідну програмну реалізацію на основі створеного методу та виконано тестування створеного ПЗ.

– Виконано дослідження ефективності створеного методу з використанням розробленого ПЗ, що показав високу ефективність, досягнувши показників 0.928 за метрикою Accuracy та 0.912 за метрикою F1-міри. Практичне тестування роботи моделі показало її спроможність щодо ефективного виявлення кібербулінгу, та може використовуватись у вигляді надбудови до соціально-орієнтованих вебсервісів для протидії кібербулінгу.

Результат, отриманий в ході розробки КРБ, цілком відповідає поставленому завданню. За темою кваліфікаційної роботи бакалавра автором виконано наукову публікацію «Метод нейромережевого виявлення кібербулінгу з використанням хмарних сервісів та об'єктно-орієнтованої моделі» у фаховому журналі [31].

Перелік посилань

1. Кібербулінг: як протистояти URL: https://supportme.org.ua/needle-and-bullying/cyberbullying

2. Кібербулінг. URL: https://www.unicef.org/ukraine/cyberbulling

3. Кібербулінг та кібергрумінг: поняття, протидія, відповідальність. URL: https://legalaid.gov.ua/publikatsiyi/kiberbuling-ta-kibergruming-ponyattya-protydiya-vidpovidalnist/

4. Кібербулінг: що це, яким він буває та як від нього захистити свою дитину. URL: https://www.telegraf.in.ua/kremenchug/10082081-kberbulng-scho-ce-yakim-vn-buvaye-ta-yak-vd-nogo-zahistiti-svoyu-ditinu.html

5. Кібербулінг в освітньому середовищі. URL: https://wiki.legalaid.gov.ua/index.php/Кібербулінг_в_освітньому_середовищі

6. Machine learning to the rescue: Preventing cyberbullying in real time URL: https://lens.monash.edu/@politics-society/2023/03/28/1385576/machine-learning-to-the-rescue-preventing-cyberbullying-in-real-time

7. DEA-RNN: A Hybrid Deep Learning Approach for Cyberbullying Detection in Twitter Social Media Platform. URL: https://ieeexplore.ieee.org/document/9718597

8. Bidirectional LSTM. URL: https://paperswithcode.com/method/bilstm

9. Cyberbullying detection solutions based on deep learning architectures. URL: https://link.springer.com/article/10.1007/s00530-020-00701-5

10. Detection of cyber bullying on social media using machine learning. URL: https://jespublication.com/upload/2022-V13I7091.pdf

11. Cyberbullying Detection in Social Networks: A Comparison Between Machine Learning and Transfer Learning Approaches. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10122521

12. Classifying the Severity of Cyberbullying Incidents by Using a Hierarchical Squashing-Attention Network. URL: https://www.mdpi.com/2076-3417/12/7/3502

13. Kim Y. Convolutional neural networks for sentence classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25–29 October 2014; pp. 1746–1751

14. Agrawal, S.; Awekar, A. Deep learning for detecting cyberbullying across multiple social media platforms. In Advances in Information Retrieval; Springer: Berlin/Heidelberg, Germany, 2018; pp. 141–153

15. Rosa, H.; Matos, D.M.; Ribeiro, R.; Coheur, L.; Carvalho, J.P. A "Deeper" look at detecting cyberbullying in social networks. In Proceedings of the 2018 International Joint Conference on Neural Networks, Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–8

16. Devlin, J.; Chang, M.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 North American Chapter of the Association for Computational Linguistics-Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186

17. StopBullying.gov. URL: https://www.stopbullying.gov/

18. The cybersmile foundation. URL: https://www.cybersmile.org/downloadable-resources

19. Під час глобальної пандемії COVID-19 ризики онлайн-насильства над дітьми підвищуються. URL: https://www.unicef.org/ukraine/прес-релізи/під-час-глобальної-пандемії-covid-19-ризики-онлайн-насильства-над-дітьми-підвищуються

20. Cyberbullying Classification. URL: https://www.kaggle.com/datasets/andrewmvd/cyberbullying-classification

21. Torch 2.2.0. URL: https://pypi.org/project/torch/

22. Keras. URL: https://keras.io/

23. TensorFlow. URL: https://www.tensorflow.org/

24. Os. URL: https://docs.python.org/uk/3/library/os.html

25. Natural Language Toolkit. URL: https://www.nltk.org/

26. Pyplot tutorial. URL: https://matplotlib.org/stable/tutorials/pyplot.html

27. Kivy: The Open Source Python App Development Framework. URL: https://kivy.org/

28. Що таке мова програмування Python. URL: https://freehost.com.ua/ukr/faq/wiki/chto-takoe-jazik-programmirovanija-python/

29. Google Colab. URL: https://colab.research.google.com/

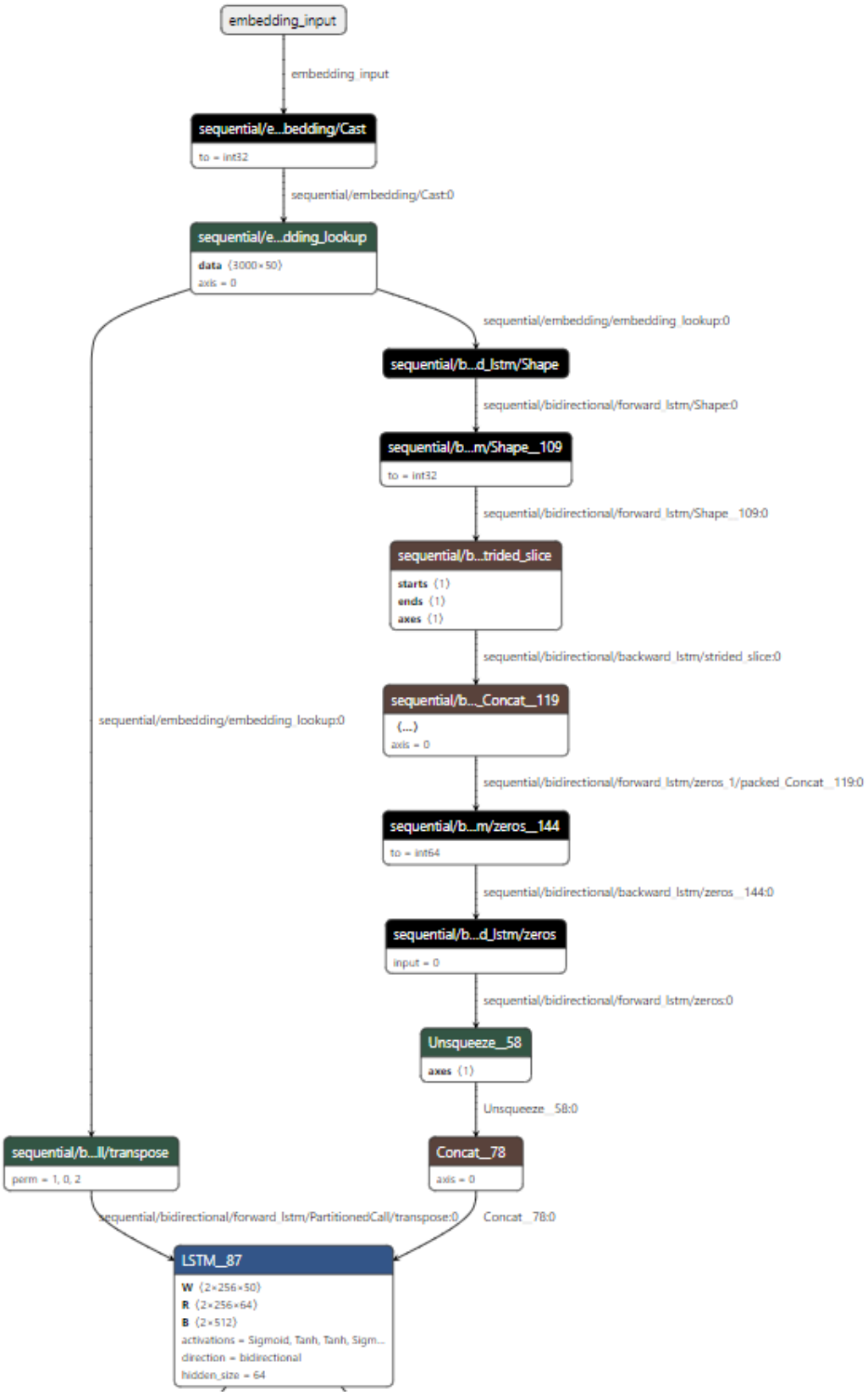
30. PyCharm. URL: https://uk.wikipedia.org/wiki/PyCharm

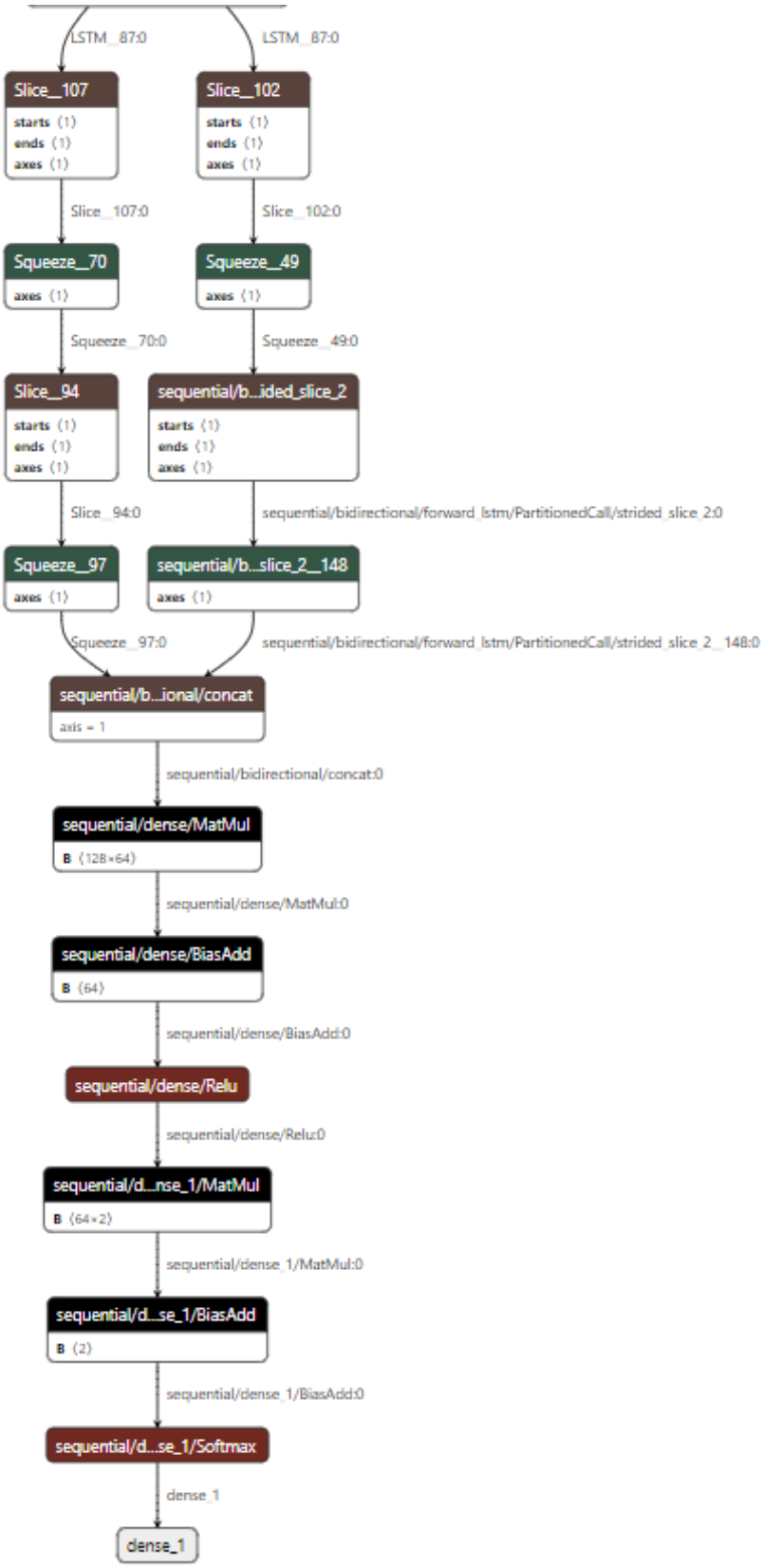
31. Молчанова М.О., Мазурець О.В., Собко О.В., Кліменко В.І., Андрощук В.І. Метод нейромережевого виявлення кібербулінгу з використанням хмарних сервісів та об'єктно-орієнтованої моделі. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2024. №2 (333). С. 200-206.

ДОДАТКИ

Додаток А

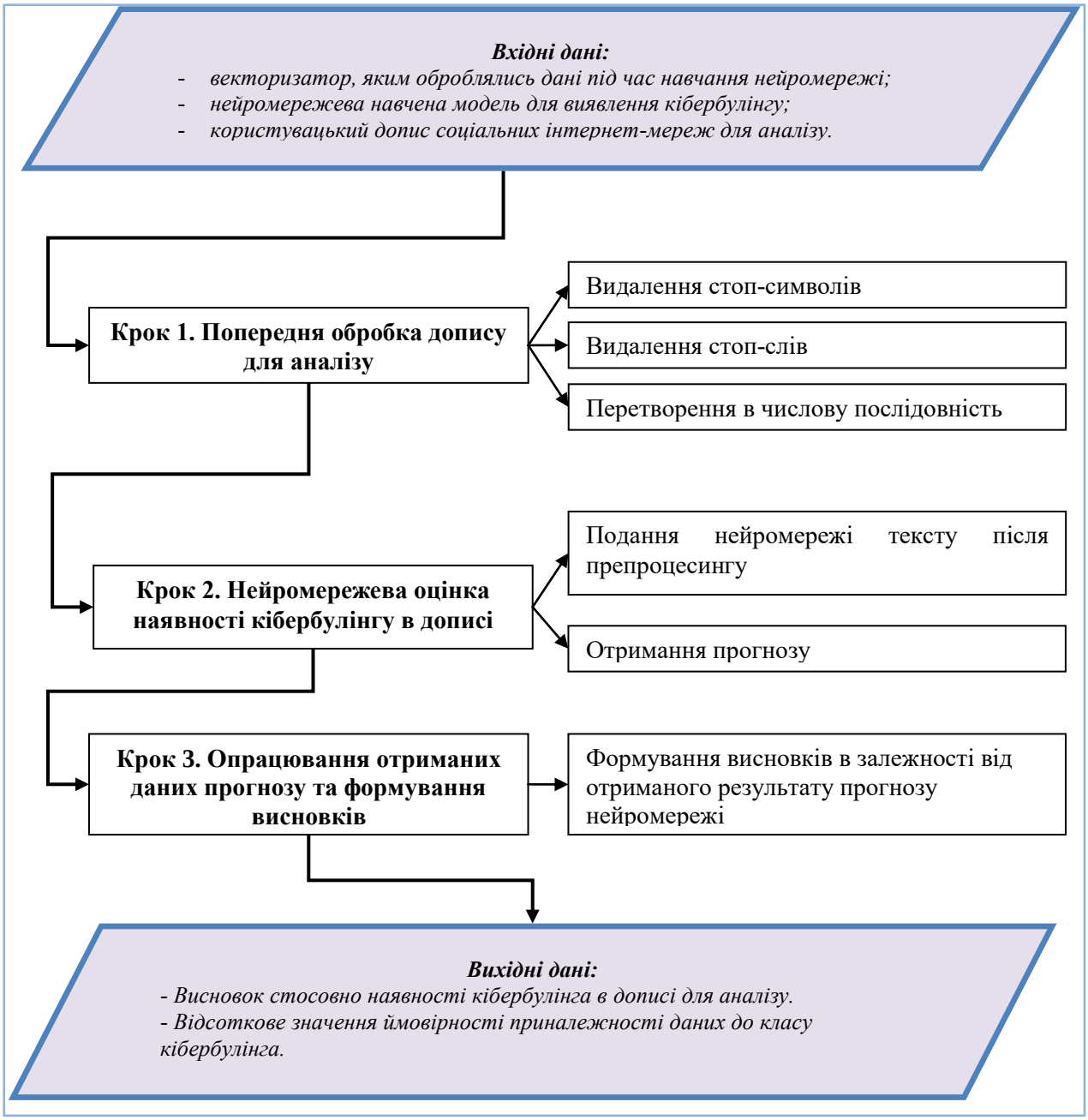
Архітектура використовуваної нейромережі BiLSTM




LSTM_87:0
LSTM_87:0
Slice_107
starts ⟨1⟩
ends ⟨1⟩
axes ⟨1⟩
Slice_102
starts ⟨1⟩
ends ⟨1⟩
axes ⟨1⟩
Slice_107:0
Slice_102:0
Squeeze_70
axes ⟨1⟩
Squeeze_49
axes ⟨1⟩
Squeeze_70:0
Squeeze_49:0
Slice_94
starts ⟨1⟩
ends ⟨1⟩
axes ⟨1⟩
sequential/b...ided_slice_2
starts ⟨1⟩
ends ⟨1⟩
axes ⟨1⟩
Slice_94:0
sequential/bidirectional/forward_lstm/PartitionedCall/strided_slice_2:0
Squeeze_97
axes ⟨1⟩
sequential/b...slice_2_148
axes ⟨1⟩
Squeeze_97:0
sequential/bidirectional/forward_lstm/PartitionedCall/strided_slice_2_148:0
sequential/b...ional/concat
axis = 1
sequential/bidirectional/concat:0
sequential/dense/MatMul
B ⟨128×64⟩
sequential/dense/MatMul:0
sequential/dense/BiasAdd
B ⟨64⟩
sequential/dense/BiasAdd:0
sequential/dense/Relu
sequential/dense/Relu:0
sequential/d...nse_1/MatMul
B ⟨64×2⟩
sequential/dense_1/MatMul:0
sequential/d...se_1/BiasAdd
B ⟨2⟩
sequential/dense_1/BiasAdd:0
sequential/d...se_1/Softmax
dense_1
dense_1

Додаток Б

Схема методу виявлення кібербулінгу в дописах соціальних інтернет-мереж



Додаток В

Проектна архітектура системи та взаємозв'язок компонентів

Додаток Г

Презентаційний матеріал

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

МЕТОД ВИЯВЛЕННЯ КІБЕРБУЛІНГУ В ДОПИСАХ СОЦІАЛЬНИХ ІНТЕРНЕТ-МЕРЕЖ НЕЙРОМЕРЕЖЕВИМИ ЗАСОБАМИ

Виконав:
студент групи КНс-21-1
Владислав АНДРОЩУК
Керівник:
викладач кафедри КН
Марина МОЛЧАНОВА

Актуальність

В контексті сучасності, де взаємодія через соціальні мережі стає невід'ємною частиною життя, проблема кібербулінгу набуває особливої актуальності. Зростаюча кількість користувачів у цифровому просторі створює ідеальні умови для поширення ворожих або образливих повідомлень, які можуть серйозно впливати на психічне та емоційне становище людей.

Нейромережеві моделі можуть бути навчені розпізнавати ключові ознаки кібербулінгу, такі як агресивна лексика, образливі коментарі та загрози, забезпечуючи автоматизовану систему виявлення без необхідності ручного моніторингу. Здатність автоматично аналізувати величезні обсяги інформації за короткий проміжок часу дозволяє вчасно виявляти та реагувати на можливі випадки кібербулінгу. Застосування нейромережевих технологій дозволяє підвищити ефективність та точність цього процесу, що важливо для забезпечення безпеки та благополуччя користувачів соціальних мереж.

Таким чином, використання нейромереж для виявлення кібербулінгу в соціальних мережах визначається необхідністю забезпечення безпеки та психологічного комфорту користувачів у віртуальному просторі, що стає дедалі важливішим в аспекті формування здорового та етичного інтернет-середовища.

Мета і задачі роботи

Метою кваліфікаційної роботи бакалавра є спрощення експертизи виявлення кібербулінгу за рахунок автоматизованого виявлення кібербулінгу в дописах соціальних інтернет-мереж.

Для досягнення поставленої мети слід вирішити такі **завдання**:

- виконати аналіз інформаційних моделей області виявлення кібербулінгу;
- виконати огляд теоретичних підходів та обрати підхід для нейромережевого виявлення кібербулінгу;
- провести аналіз існуючих публікацій за напрямком дослідження;
- провести аналіз існуючого програмного забезпечення області виявлення кібербулінгу в дописах соціальних інтернет-мереж;
- створити метод виявлення кібербулінгу в дописах соціальних інтернет-мереж;
- описати інформаційну структуру системи виявлення кібербулінгу в дописах;
- обрати набір даних для навчання нейромережевої компоненти методу;
- створити відповідну програмну реалізацію на основі створеного методу;
- виконати тестування створеного ПЗ;
- виконати дослідження ефективності створеного методу з використанням розробленого ПЗ.

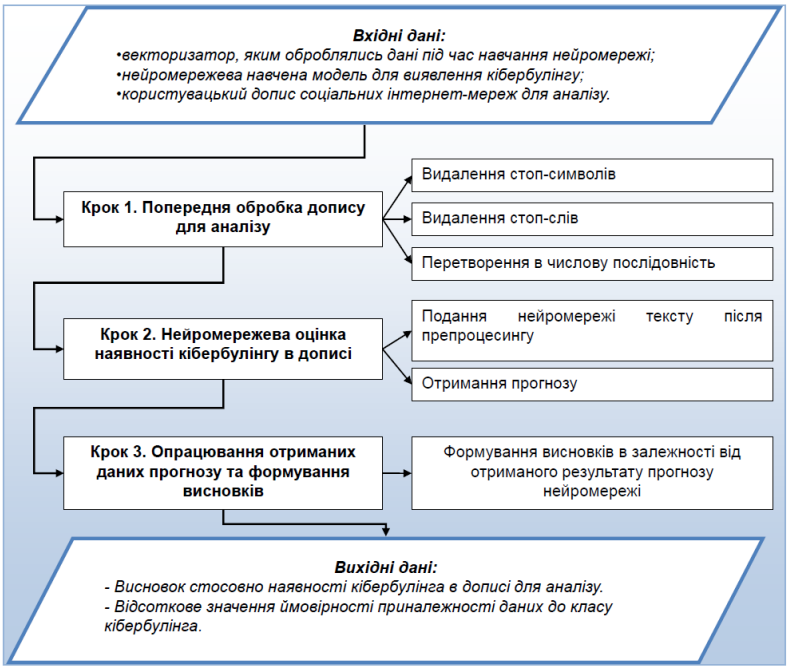
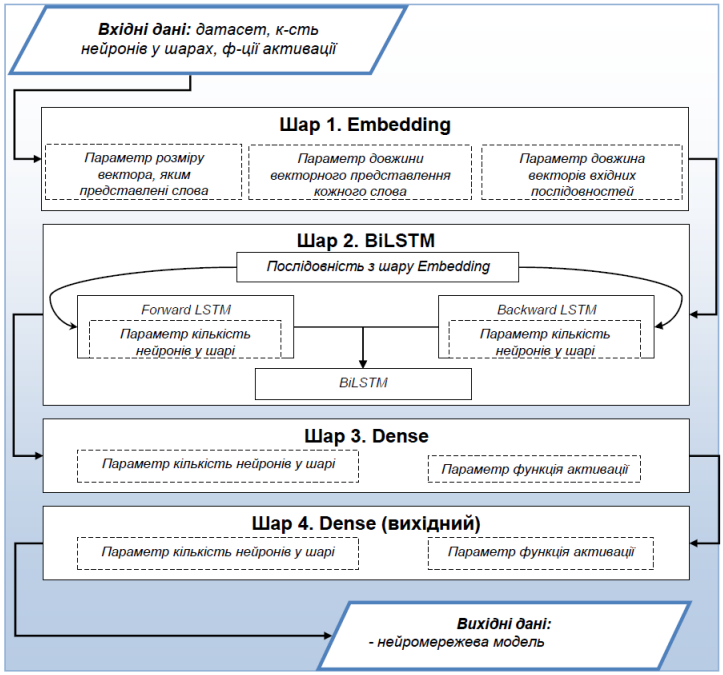
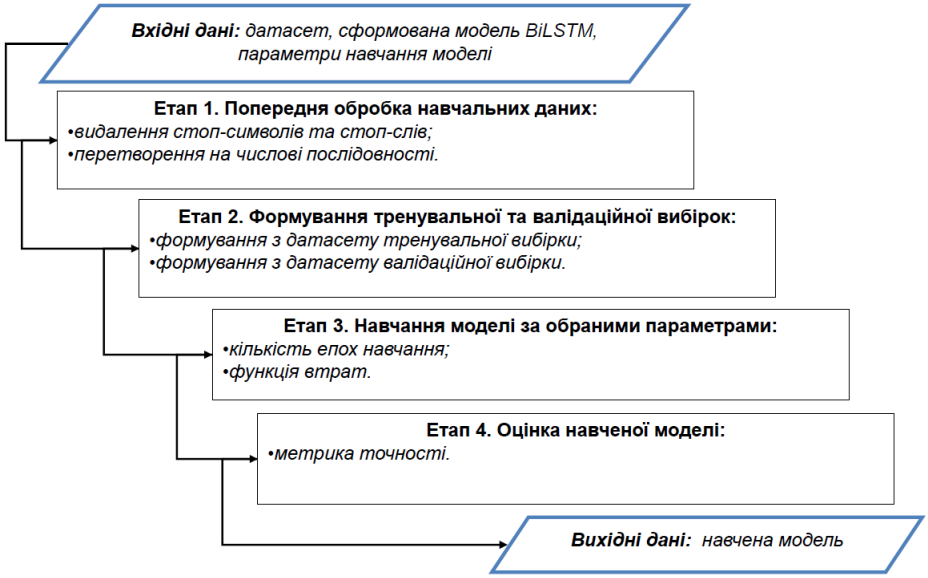


Схема методу виявлення кібербулінгу в дописах соціальних інтернет-мереж

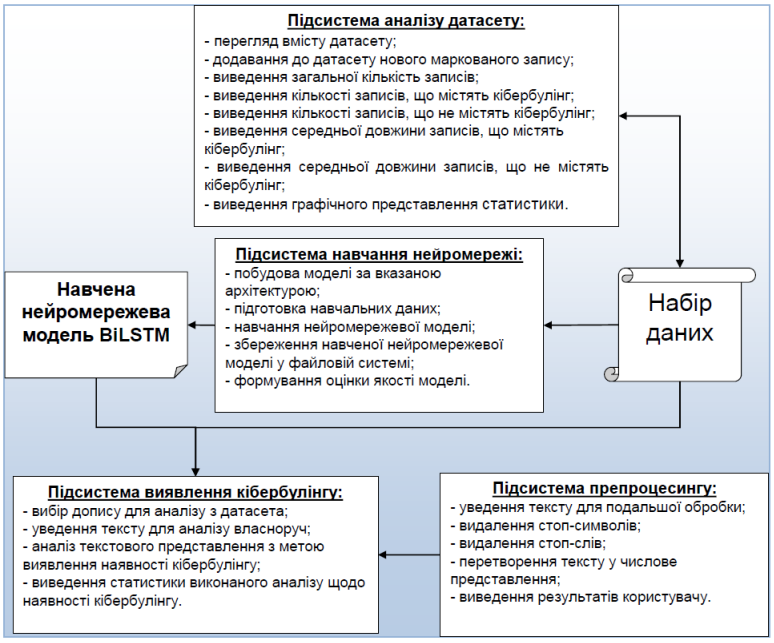
Схема навігації між підсистемами

Архітектура нейромережевої моделі BiLSTM





Етапи навчання нейромережі BiLSTM



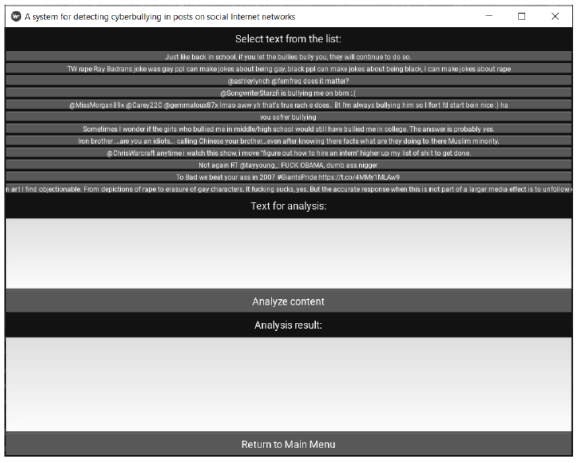
Проектна архітектура інформаційної системи виявлення кібербулінгу

tweet_text Text of the tweet	cyberbullying_type Type of cyberbullying harassment.
46017 unique values	religion 17% age 17% Other (31702) 66%
In other words #katandandre, your food was crapilicious! #mkr	not_cyberbullying
Why is #aussietv so white? #MKR #theblock #ImACelebrityAU #today #sunrise #studio10 #Neighbours #Won...	not_cyberbullying
@XochitlSuckkks a	not_cyberbullying

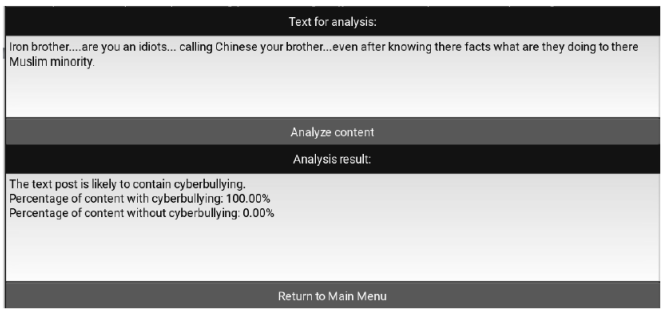
Набір даних дослідження

Для навчання нейромережі BiLSTM буде використано набір даних Cyberbullying Classification, що налічує 39747 зразків, з яких обрано 8000 зразків, що містять кібербулінг, та 7945 записів без кібербулінгу.

Інформаційна система виявлення кібербулінгу



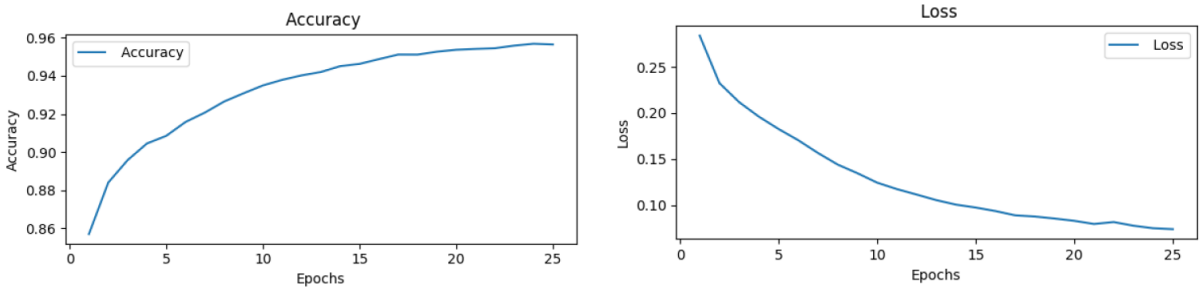
Інтерфейс підсистеми по визначенняю кібербулінгу



Вивід результату дослідження тексту з кібербулінгом

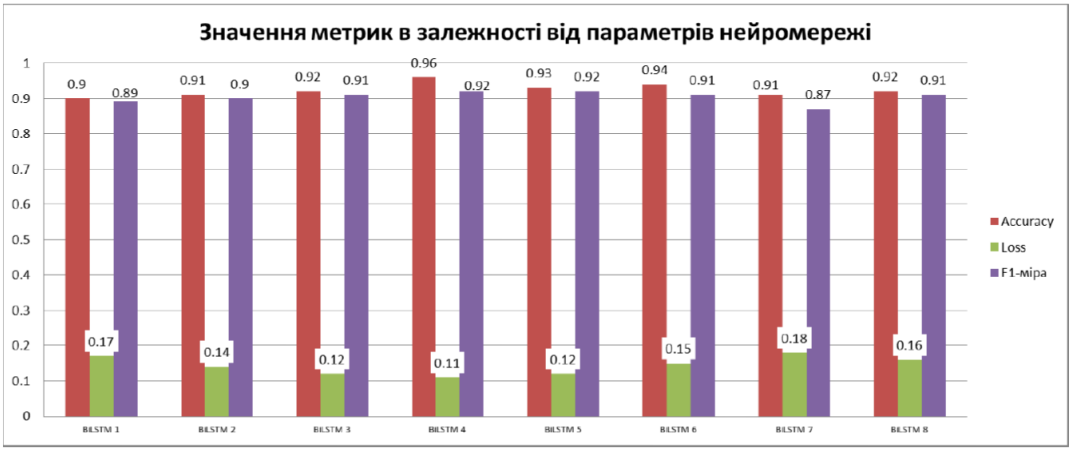
Результати досліджень

Розроблена програмна реалізація дозволила успішно виявляти кібербулінг у англомовних дописах соціальних інтернет-мереж. Відсоток точності за метрикою Accuracy на тренувальному наборі становив близько 95 %. Були проведені експерименти з навчання нейромережі, які мали на меті сприяти покращенню базового відсотку ідентифікації, зміни стосувались кількості нейронів в шарах, кількості епох навчання та довжини словника і вхідних послідовностей.



Статистика навчання за 25-тьма епохами

Результати досліджень



Версія BiLSTM з кількістю епох 25 та кількістю нейронів в шарах LSTM та Dense 64 і розміром словника в 5000 показала кращий результат, досягнувши значень 0.96 метрикою Accuracy та 0.92 метрикою F1-міри. При цьому функція втрат склала 0.11.

Висновки

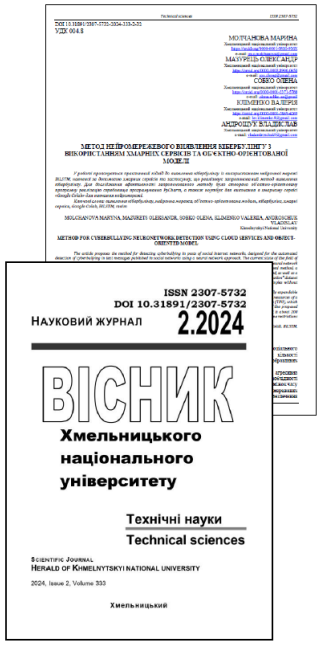
Було досягнуто мету кваліфікаційної роботи бакалавра – спрощення експертизи виявлення кібербулінгу за рахунок автоматизованого виявлення кібербулінгу в дописах соціальних інтернет-мереж.

Для досягнення поставленої мети були поставлені та вирішені такі завдання:

- Виконано аналіз предметної області виявлення кібербулінгу, в рамках якого з'ясовано, що виявлення і вчасна реакція на кібербулінг є важливим елементом захисту безпеки користувачів соціальних інтернет-мереж, тому автоматизація є актуальною задачею інформаційних технологій.
- Створено метод виявлення кібербулінгу в дописах соціальних інтернет-мереж
- Створено нейромережеву архітектуру для методу виявлення кібербулінгу в дописах. Наведено етапи навчання нейромережі на основі розробленої архітектури та підготовлено навчальні дані.
- Наведено проектну архітектуру інформаційної системи виявлення кібербулінгу в текстових дописах, що публікуються в соціальних інтернет-мережах.
- Створено відповідну програмну реалізацію на основі створеного методу та виконано тестування створеного ПЗ.
- Виконано дослідження ефективності створеного методу з використанням розробленого ПЗ, що показав високу ефективність, досягнувши показників 0.928 за метрикою Accuracy та 0.912 за метрикою F1-міри. Практичне тестування роботи моделі показало її спроможність щодо ефективного виявлення кібербулінгу, та може використовуватись у вигляді надбудови до соціально-орієнтованих вебсервісів для протидії кібербулінгу.

За темою кваліфікаційної роботи бакалавра автором виконано наукову публікацію «Метод нейромережевого виявлення кібербулінгу з використанням хмарних сервісів та об'єктно-орієнтованої моделі» у фаховому журналі:

Молчанова М.О., Мазурець О.В., Собко О.В., Кліменко В.І., Андрощук В.І. Метод нейромережевого виявлення кібербулінгу з використанням хмарних сервісів та об'єктно-орієнтованої моделі. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2024. №2 (333). С. 200-206.



ISSN 2307-5732
DOI 10.31891/2307-5732
НАУКОВИЙ ЖУРНАЛ 2.2024

ВІСНИК
Хмельницького національного університету

Технічні науки
Technical sciences

SCIENTIFIC JOURNAL
HERALD OF KHMELNYTSKYI NATIONAL UNIVERSITY

2024, Issue 2, Volume 333

Хмельницький

UNICHECK
by Turnitin

Ім'я користувача:
Кафедра КН

ID перевірки:
1016374210

Дата перевірки:
19.06.2024 07:58:26 EEST

Тип перевірки:
Doc vs Internet + Library

Дата звіту:
19.06.2024 08:48:53 EEST

ID користувача:
100005671

Назва документа: **КНс-21-1 Андрощук_ЗАПИСКА**

Кількість сторінок: **70** Кількість слів: **11857** Кількість символів: **96506** Розмір файлу: **2.02 MB** ID файлу: **1016181930**

Виявлено модифікації тексту (можуть впливати на відсоток схожості)

11.2%
Схожість

Найбільша схожість: **4.01%** з джерелом з Бібліотеки (ID файлу: **1016177779**)

7.46% Джерела з Інтернету	808	Сторінка 72
6.97% Джерела з Бібліотеки	112	Сторінка 77

0% Цитат

Вилучення цитат вимкнене

Вилучення списку бібліографічних посилань вимкнене

0%
Вилучень

Немає вилучених джерел

Модифікації

Виявлено модифікації тексту. Детальна інформація доступна в онлайн-звіті.

Замінені символи	1
Підозріле форматування	15 сторінок

Wed Jun 19 07:53:41 EEST 2024, Петровський Сергій Степанович, Хмельницький національний університет, ХНУ

Anti-Plagiarism v-15.257

Максимальне співпадіння з одним документом 4.0%

Словники перевірки: en_US, ru_RU, ua_UA. **Помилок в документах: 15%**

	Документ		Сумарний збіг по Базі Даних	
ID: 131448 Назва: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод виявлення кібербулінгу в дописах соціальних інтернет-мереж нейромережевими засобами Додано в БД: 2024-06-19 Автора: Владислав АНДРОЩУК Керівники: Марина МОЛЧАНОВА Консультанти: Опоненти:	Символи	Лексеми	Символи	Лексеми
	77306	1132	4471 (6%)	67 (6%)

Джерело плагіату

ID	Опис	Наявність плагіату в документі	
		Символи	Лексеми

РІШЕННЯ ЕКСПЕРНОЇ КОМІСІЇ КАФЕДРИ КОМП'ЮТЕРНИХ НАУК
ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ ДО ЗАХИСТУ

Підтверджуємо ознайомлення з результатом звіту подібності щодо роботи, генерованого системою виявлення текстових збігів/ідентичності/схожості:

Назва: *Метод виявлення кібербулінгу в дописах соціальних інтернет-мереж нейромережевими засобами*

Автор: *студент групи КНс-21-1 Владислав Андрощук*

Спеціальність: *122 – Комп'ютерні науки*

Освітня програма: *освітньо-професійна*

Науковий керівник: *викладач кафедри КН Марина Молчанова*

Після аналізу звіту подібності зроблено такий висновок:

№	Висновок	Позначка про відповідність
1	Запозичення, виявлені в роботі, є законними і не є плагіатом. Робота приймається до захисту.	***відповідає***
2	Виявлені запозичення не є плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи Робота приймається до захисту, але має бути відкоригована. Відкоригований варіант має бути поданий на кафедру за 2 дні до захисту, разом із заявою щодо самостійності виконання письмової роботи та ідентичності друкованої та електронної версії роботи	
3	Виявлені запозичення не є плагіатом, але частково розміщені в розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. В зв'язку з цим мета роботи та поставлені завдання не були досягнені. Робота може бути допущена до захисту (наступного року) після того як буде відкоригована та допрацьована і успішно пройде повторну перевірку на академічний плагіат.	
4	Робота містить навмисні текстові спотворення, передбачувані спроби укриття запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	

Підтвердження:

Запозичення, виявлені в роботі Владислава Андрощука, не є плагіатом, оскільки: запозичення розміщені в розділі огляду існуючих підходів, не описують безпосередньо авторську роботу і не стосуються її результатів; усі запозичення фрагментарні; до запозичень входять фрагменти, що не мають авторства і містять поширені конструкції; серед запозичень знаходяться загальновідомі терміни та скорочення.

Обсяг запозичень, визначений системами виявлення збігів/ ідентичності/схожості, складає:

- за системою Anti-Plagiarism: 4%;

- за системою Unicheck: 11.2 %

Керівник роботи _____ *Марина МОЛЧАНОВА*

Гарант ОП _____ *Олександр МАЗУРЕЦЬ*

Завідувач кафедри КН _____ *Олександр БАРМАК*

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МОН УКРАЇНИ

Кафедра комп'ютерних наук

ВІДГУК НАУКОВОГО КЕРІВНИКА

на кваліфікаційну роботу бакалавра

студента *гр. КНс-21-1 Андрощука Владислава Івановича*

за темою *Метод виявлення кібербулінгу в дописах соціальних інтернет-мереж нейромережевими засобами*

1. Актуальність теми

У сучасному світі, де взаємодія через соціальні мережі є частиною повсякденного життя, виявлення кібербулінгу набуває великої актуальності. Швидке поширення ворожих та образливих повідомлень у цифровому просторі негативно впливає на психічне та емоційне здоров'я користувачів, тому розробка ефективних методів і технологій для виявлення та протидії кібербулінгу є надзвичайно важливою для забезпечення безпеки та добробуту в інтернет-середовищі.

2. Відповідність роботи предметній області Стандарту спеціальності 122 Комп'ютерні науки

За стандартом, а саме описом предметної області, об'єктом дослідження є процес виявлення кібербулінгу в дописах соціальних інтернет-мереж нейромережевими засобами. Метою кваліфікаційної роботи бакалавра є спрощення експертизи виявлення кібербулінгу за рахунок автоматизованого виявлення кібербулінгу в дописах соціальних інтернет-мереж. Мета роботи досягнута шляхом розробки методу виявлення кібербулінгу в дописах соціальних інтернет-мереж з використанням методів та засобів машинного навчання для роботи з текстовою інформацією. Отже, результати виконання кваліфікаційної роботи бакалавра відповідають стандарту бакалавра спеціальності 122 – Комп'ютерні науки.

3. Професійні та особистісні якості бакалавра

Під час виконання кваліфікаційної роботи бакалавра Андрощук Владислав Іванович показав хороші знання з предметної області. До виконання роботи студент підійшов відповідально, виконуючи усі етапи вчасно та якісно.

4. Ступінь самостійності під час виконання кваліфікаційної роботи

Отримані під час виконання кваліфікаційної роботи результати є наслідком особистої роботи студента, який самостійно виконував всі поставлені задачі.

5. Ступінь оволодіння методами дослідження

У процесі виконання кваліфікаційної роботи студент продемонстрував високий рівень знань та володіння необхідними інструментами, методами, методиками та технологіями, що відповідають спеціальності 122 – Комп'ютерні науки.

6. Повнота та якість розкриття теми роботи

Тема дослідження була обґрунтована і детально розкрита, що свідчить про глибоке розуміння автором поставлених задач. Робота включала аналіз актуальності теми на основі літературних джерел і підходів, що підтверджує необхідність подальших досліджень в цій сфері. Поставлені завдання були чітко сформульовані і успішно виконані під час виконання кваліфікаційної роботи бакалавра.

7. Логічність, послідовність, аргументованість, літературна грамотність викладення матеріалу

Текст записки має логічний та послідовний виклад. Усі твердження підкріплені сучасними літературними джерелами. Етапи розробки методу та інформаційної системи детально пояснено та грамотно описано.

8. Можливість практичного застосування кваліфікаційної роботи бакалавра, окремих її частин

Розроблений метод виявлення кібербулінгу в дописах соціальних інтернет-мереж нейромережевими засобами, може бути імплементованим в автоматизовані системи соціальних мереж для захисту користувачів від негативного впливу.

9. Висновок про можливість допуску кваліфікаційної роботи бакалавра до захисту, на яку оцінку заслуговує робота

Враховуючи цілком достатній рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту. Рекомендована оцінка ***«добре»*** .

Керівник _____ *викладач каф. КН Марина МОЛЧАНОВА*

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МОН УКРАЇНИ

Кафедра комп'ютерних наук

РЕЦЕНЗІЯ

на кваліфікаційну роботу бакалавра

студента *гр. КНс-21-1 Андрощука Владислава Івановича*

за темою: *Метод виявлення кібербулінгу в дописах соціальних інтернет-мереж нейромережевими засобами*

1. Актуальність обраної теми

У сучасному світі, де соціальні мережі стали невід'ємною частиною життя, проблема кібербулінгу стає все більш актуальною. Зростання кількості користувачів сприяє поширенню ворожих та образливих повідомлень, які можуть завдати серйозної шкоди психічному та емоційному здоров'ю. Нейромережеві моделі здатні розпізнавати ознаки кібербулінгу, такі як агресивна мова та загрози, дозволяючи автоматично виявляти ці прояви без потреби в ручному моніторингу.

2. Повнота розкриття мети та завдань роботи

У кваліфікаційній роботі бакалавра мета та завдання були повністю розкриті. Автор детально висвітлив процес розробки та програмну реалізацію методу виявлення ознак кібербулінгу в текстових повідомленнях користувачів за допомогою нейромережевих засобів.

3. Зміст кожного розділу роботи

Кожен розділ бакалаврської роботи містить інформацію, що відповідає темі, починаючи з теоретичних аспектів виявлення ознак кібербулінгу в текстових дописах користувачів за допомогою нейромережевих технологій і закінчуючи практичною реалізацією. У першому розділі наведено характеристику предметної області виявлення кібербулінгу в дописах соціальних інтернет-мереж. У другому розділ здійснено проєктування інформаційної системи, а також спроєктовано метод виявлення кібербулінгу в дописах соціальних інтернет-мереж. У третьому розділі виконано дослідження методу виявлення кібербулінгу в дописах соціальних інтернет-мереж.

4. Оцінка розробленої інформаційної системи, її практична цінність

Розроблена інформаційна система виявлення кібербулінгу за текстовим дописом соціальних інтернет-мереж показала хороші результати дослідження, тому має значний потенціал для застосування та подальшого впровадження у вебсистеми.

5. Якість оформлення кваліфікаційної роботи бакалавра

Оформлення бакалаврської роботи виконане на високому рівні та включає всі необхідні компоненти, такі як розділи, таблиці, графіки і посилання на джерела. Робота відзначається чіткою структурою, логічною послідовністю та науковим стилем викладу, що забезпечує легкість сприйняття і дозволяє чітко оцінити проведені дослідження та отримані результати.

6. Недоліки кваліфікаційної роботи бакалавра

Кваліфікаційна робота бакалавра виконана на високому рівні, проте розроблена система виявлення кібербулінгу за текстовим дописом соціальних інтернет-мереж виконана з англомовним інтерфейсом та працює тільки з англійськими текстами. В тексті записки присутні несуттєві граматичні помилки

7. Загальний висновок (допускається чи не допускається до захисту), та оцінка на яку заслуговує кваліфікаційна робота.

Враховуючи високий рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту. Рекомендована оцінка ***«добре»***.

Рецензент д.т.н., проф. Мисенко С.М.