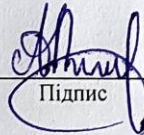
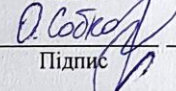
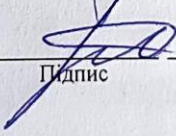
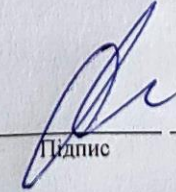


КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

на тему Метод виявлення стресового стану за дописами з соціальних
інтернет-мереж NLP-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

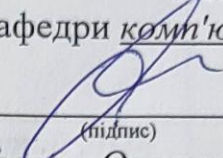
Виконав: студент групи КНс-21-1  Антон ЮРЧАК
Група виконавця Підпис Ім'я, ПРІЗВИЩЕ
Керівник: викладач каф. КН  Олена СОБКО
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ
Нормоконтроль: к.т.н., доц. каф. КН  Руслан БАГРІЙ
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ

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

20 06 2024 р.

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 стресового стану за дописами з соціальних інтернет-мереж NLP-засобами»

2. Завдання видано студенту Антону ЮРЧАКУ
(Ім'я, прізвище)

3. Керівник роботи викладач кафедри КН Олена СОБКО
(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 виявлення стресового стану за дописами з соціальних інтернет-мереж за рахунок аналізу дописів NLP-засобами. Для досягнення мети слід виконати такі задачі: здійснити дослідження предметної, виконати огляд теоретичних підходів щодо виявлення стресового стану за дописами; виконати аналіз наукових надбань та існуючих програмних предметної області; виділити нерозв'язані та розробити метод виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами; на основі розробленого методу виконати проектування інформаційної структури системи; виконати підготовку навчальних даних; здійснити вибір засобів розробки для інформаційної системи та здійснити програмну реалізацію; провести тестування розробленої програмної реалізації та дослідження ефективності.

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

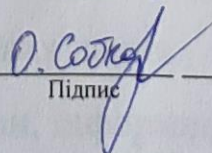
Група виконавця


Підпис

Антон ЮРЧАК
Ім'я, ПРІЗВИЩЕ

Керівник: викладач каф. КН

Науковий ступінь, посада


Підпис

Олена СОБКО
Ім'я, ПРІЗВИЩЕ

Анотація

Тема кваліфікаційної роботи бакалавра: «Метод виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами»

Виконавець кваліфікаційної роботи бакалавра: студент групи КНс-21-1 Антон ЮРЧАК

Керівник кваліфікаційної роботи бакалавра: викладач кафедри КН Олена СОБКО

Кваліфікаційна робота бакалавра містить:

Пояснювальна записка				Кількість додатків
Сторінок	Рисунків	Таблиць	Джерел інформації	
62	32	8	26	4


Кваліфікаційна робота бакалавра присвячена спрощенню виявлення стресового стану за дописами з соціальних інтернет-мереж за рахунок аналізу дописів NLP-засобами, у межах чого було розроблено метод виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами, а також відповідну програмну реалізацію, що використовує створений метод.

Для розробки інформаційної системи було використано мову програмування Python та середовище програмування PyCharm. Розроблена система призначена для психологів у якості додаткового інструменту відстеження стресових станів пацієнтів, а також для само діагностування користувачів соціальних мереж на наявність стресового стану.

Напрямами практичного використання розробленої інформаційної системи визначено автоматизоване виявлення стресового стану за текстовим повідомленням, яке може також використовуватись науковцями для автоматизованої розмітки масивів текстових даних.

Ключові слова: GRU, стресовий стан, інформаційна система, нейромережа.

Виконавець: студент групи КНс-21-1
Група виконавця


Підпис

Антон ЮРЧАК
Ім'я, ПРІЗВИЩЕ

Зміст

Перелік скорочень	4
Вступ.....	5
Розділ 1 Характеристика предметної області автоматизованого виявлення стресового стану за дописами.....	7
1.1 Аналіз інформаційних моделей для виявлення стресового стану	7
1.2 Огляд теоретичних підходів до розв’язку подібних задач	9
1.3 Аналіз існуючих програмних засобів та наукових рішень у галузі виявлення стресового стану за текстом.....	12
1.4 Мета, задачі та вимоги до реалізації інформаційної системи	17
Розділ 2 Метод виявлення стресового стану за дописами з соціальних інтернет-мереж	19
2.1 Схема та етапи методу виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами.....	19
2.2 Функціональна структура інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж	22
2.3 Розробка архітектури нейронної мережі	24
2.4 Проектна архітектура інформаційної системи виявлення стресового стану за дописами та взаємозв’язок компонентів	25
2.5 Підготовка робочих вхідних даних для інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж	27
2.6 Особливості використання спеціалізованих програмних компонентів інформаційної системи виявлення стресового стану	29
2.7 Висновки до розділу 2	32
Розділ 3 Експериментальне дослідження методу виявлення стресового стану за дописами.....	34
3.1 Визначення шляхів дослідження та засобів створення програмного забезпечення	34

3.2 Вибір засобів розробки інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж.....	35
3.3 Структура та функціональне призначення програмних складових інформаційної системи виявлення стресового стану за дописами	36
3.4 Особливості реалізації програмних складових інформаційної системи виявлення стресового стану за дописами	38
3.5 Тестування інформаційної системи та вимоги до розгортання	41
3.6 Аналіз функціональності інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами	49
3.7 Результати досліджень	54
3.8 Висновки до розділу 3	57
Загальні висновки.....	59
Перелік посилань.....	61
Додатки	

Перелік скорочень

Скорочення, термін, позначення	Пояснення
NLP	Natural Language Processing
SVM	Метод опорних векторів
RNN	Рекурентні нейронні мережі
CNN	Згорткові нейронні мережі
BERT	Bidirectional Encoder Representations from Transformers
GPT	Generative Pre-trained Transformer
ML	Машинне навчання
UGC	User Generated Content
BDI	Beck Depression Inventory
IOT	Інтернет речей
GSR	Датчик шкірно-гальванічної реакції
KNN	Метод К-найближчого сусіда
MLP	Багатошаровий перцептрон
DT	Дерево рішень
DL	Глибоке навчання
ПЗ	Програмне забезпечення
КРБ	Кваліфікаційна робота бакалавра
GRU	Gated Recurrent Unit
CSV	Comma-Separated Values
JSON	JavaScript Object Notation

Вступ

Кваліфікаційна робота бакалавра присвячена спрощенню виявлення стресового стану за дописами з соціальних інтернет-мереж за рахунок аналізу дописів NLP-засобами, у межах чого було розроблено метод виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами, а також відповідну програмну реалізацію, що використовує створений метод.

Актуальність. Люди проводять все більше часу в інтернеті, особливо в соціальних мережах. Це створює нові можливості для вивчення їхнього ментального стану за допомогою аналізу текстових даних, що може бути корисним для розуміння і покращення їхнього здоров'я.

У зв'язку з останніми подіями в Україні, можна стверджувати, що Інтернет поступово перетворюється на джерело загрози для інформаційної безпеки людей, суспільства та держави. Поширення сумнівного та необ'єктивного контенту разом із застосуванням технологій інформаційно-психологічного впливу на свідомість індивідів може сприяти збільшенню невдоволення діючою державною владою, національних конфліктів, соціальної напруженості та інших негативних явищ.

Стрес є серйозною проблемою сучасного суспільства, і його виявлення на ранній стадії може допомогти уникнути серйозних наслідків для здоров'я. Аналіз текстових даних з соціальних мереж може допомогти виявити ознаки стресу, навіть якщо людина не обговорює його явно.

Об'єкт дослідження – процес виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами.

Предмет дослідження – методи та технології машинного навчання для роботи з текстовою інформацією.

Мета кваліфікаційної роботи бакалавра – спрощення виявлення стресового стану за дописами з соціальних інтернет-мереж за рахунок аналізу дописів NLP-засобами.

Завдання кваліфікаційної роботи бакалавра – виконати дослідження предметної області виявлення стресового стану за дописами; в рамках дослідження предметної області виконати огляд теоретичних підходів щодо виявлення стресового стану за дописами; виконати аналіз наукових надбань та існуючих програмних рішень в області виявлення стресового стану за дописами; розробити метод виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами; на основі розробленого методу виконати проектування інформаційної структури системи; виконати підготовку навчальних даних для тренування класифікатора для виявлення стресового стану за дописами; здійснити вибір засобів розробки для інформаційної системи; здійснити програмну реалізацію інформаційної системи для виявлення стресового стану; провести тестування розробленої програмної реалізації; здійснити дослідження ефективності розробленого методу з використанням розробленої програмної реалізації.

Розділ 1 Характеристика предметної області автоматизованого виявлення стресового стану за дописами

1.1 Аналіз інформаційних моделей для виявлення стресового стану

Стрес представляє собою природну реакцію організму на зовнішні подразники, і ця реакція може виявлятися на психічному, фізичному та емоційному рівнях, допомагаючи нам адаптуватися до змін [1].

Стрес є широко поширеною проблемою, особливо в розвинених країнах, де до 70% населення перебуває в стані постійного стресу, а понад 90% досвідчують стрес кілька разів на місяць. Ці статистичні дані є тривожними, зважаючи на потенційно небезпечні наслідки стресу [2].

Переживання стресу вимагає значних енергетичних затрат від людини. Тому тривалий вплив стресових факторів може призводити до відчуття слабкості, апатії та втоми. Крім того, стрес пов'язаний з розвитком близько 80% усіх відомих захворювань.

Основні ознаки стресового стану можуть бути розділені на кілька аспектів [3]:

1. Фізичні прояви. Під впливом стресу збільшується частота серцевих скорочень, збільшується швидкість дихання, а також може підвищитися гострота відчуттів. Цей стан не дає організму відпочити, викликає тривогу та певний ступінь напруження.

2. Психічні прояви. Стрес можна описати як стан нервової напруги або занепокоєння, що виникає внаслідок складної ситуації або взаємодії з оточуючим середовищем.

3. Тривалість впливу. Стрес може бути гострим, коли він виникає внаслідок короткочасних ситуацій, або хронічним, коли він продовжується тривалий час.

Типи стресу [4]:

– Передстресовий стан. Тривога та нервові напруження перед очікуваною стресовою ситуацією.

– Еустрес. Корисний стрес, який мобілізує резерви організму для ефективного розв'язання проблеми або виклику.

– Дистрес. Шкідливий стрес, який може призвести до зниження функціональності та перешкоджати повноцінному життю.

– Емоційний стрес. Супроводжується емоціями, такими як тривога, страх або гнів.

Г. Сельє, канадський вчений і лікар, виявив, що кожен організм реагує на стрес подібним чином, і він поділив цей процес на три фази [5]:

1. Фаза тривожної реакції. На цій стадії організм мобілізує всі свої захисні механізми. Під впливом стресу підвищується функціональність органів і систем, що дозволяє посилити такі почуття, як увага, пам'ять та сприйняття. Індивід виявляє здатність швидко мислити і знаходити рішення для подолання стресу.

2. Фаза тривоги. На цій стадії організм адаптується до змін, намагаючись повернути все під контроль. Індивіду вдається звикнути до нових умов, проте тривалість цієї фази може призвести до виснаження.

3. Фаза опору. Якщо організм не може швидко адаптуватися до стресу, ця стадія може призвести до виснаження. Фізичні сили слабшають, а психічний стан може погіршуватися.

Стрес може має вплив на фізичне та психічне здоров'я, тому важливо розпізнавати його ознаки та реагувати на них вчасно.

Виявлення стресу за текстовими повідомленнями може є складним завданням, оскільки багато залежить від контексту, індивідуальних особливостей та вміння відчутти субтельні нюанси. Однак, можна виділити певні ознаки, на які можна звернути увагу.

Частота та тип помилок. Збільшена кількість помилок у тексті (наприклад, орфографічні помилки, повторення слів) може свідчити про те, що особа поспішає або відчуває невпевненість.

Зміна тону голосу. Різка зміна тону голосу або стилю письма може вказувати на зміну настрою або емоційний стрес.

Використання емоційних виразів. Подібно до зміни тону, використання більше емоційно забарвлених слів або виразів може бути показником емоційного напруження.

Збільшена або зменшена активність. Особа може бути більш або менш активною в обміні повідомленнями, залежно від ступеня стресу, який вона відчуває.

Прохання про допомогу або підтримку. Люди, які відчувають стрес, можуть шукати підтримку або виявляти потребу у допомозі у своїх повідомленнях.

Зниження обсягу повідомлень. Підвищений стрес може призводити до зниження кількості та обсягу повідомлень, оскільки особа може відчувати втомленість або неспроможність виражати свої думки чи почуття.

Отже, у сучасному світі, стрес є практично невід'ємною деструктивною складовою життя людини. За даними вчених, 70% населення перебуває в стані постійного стресу, а з урахування пандемії та різного роду агресії, рівень стресу продовжує зростати. Отже, є потреба в автоматизованому виявленні стресу, що може допомогти людям краще справлятися з проблемами та підвищить якість життя.

1.2 Огляд теоретичних підходів до розв'язку подібних задач

Для виявлення стресу у повідомленнях соціальних інтернет-сервісів необхідно виконати аналіз сучасних теоретичних підходів до вирішення таких задач. Одним із популярних підходів для досягнення цієї мети можна використовувати NLP-засоби (Natural Language Processing), які дозволяють автоматизовано аналізувати текстовий контент.

NLP – це галузь машинного навчання, спрямована на розуміння та обробку людської мови в її різноманітних формах. Вона охоплює як писемну, так і усну мову, дозволяючи комп'ютерам взаємодіяти з людьми через текстові повідомлення, аудіозаписи, мовлення та інші форми мовного виразу. NLP

використовується в широкому спектрі застосувань, включаючи автоматичне перекладання мов, розпізнавання мовлення, аналіз настрою в соціальних мережах, генерацію тексту, виявлення спаму та багато іншого. Вона відкриває безліч можливостей для автоматизації та поліпшення спілкування між комп'ютерами та людьми [6].

У сфері NLP дані, що використовуються для навчання моделей, складаються з обмеженого словникового запасу. Однак часто зустрічаються позасловникові терміни, які не входять в цей словниковий запас. При обробці нових текстів з використанням навченої моделі, для таких термінів зазвичай призначається загальний або стандартний заміник.

Корпус (в множині – Corpora) відіграє ключову роль у процесі навчання моделей NLP. Це колекція текстових даних, яка може включати в себе огляди фільмів, коментарі в Інтернеті, літературні тексти або будь-яку іншу текстову інформацію. Корпуси використовуються для навчання моделей на великому обсязі даних, щоб вони могли ефективно розпізнавати, аналізувати та генерувати тексти в майбутньому.

Серед засобів NLP для задачі виявлення стресу у повідомленнях соціальних інтернет-сервісів є два ключові підходи – підхід на основі простих класифікаторів та підхід з глибоким навчанням.

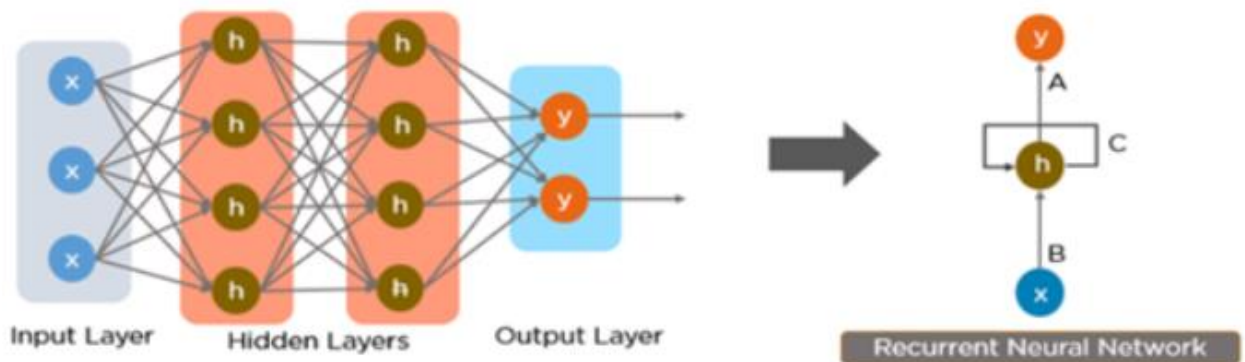
Підхід на основі простих класифікаторів, таких як наївний Байєсовський класифікатор і метод опорних векторів (SVM), використовується в NLP для різноманітних завдань, включаючи класифікацію тексту, аналіз настроїв, визначення тематики тощо.

Наївний Байєсівський класифікатор. Цей метод ґрунтується на теоремі Байєса та припущенні про незалежність між ознаками (словами) в документах. Він використовує ймовірність того, що документ належить до певної категорії, з урахуванням ймовірностей входження окремих слів у цю категорію. Наївний Байєсівський класифікатор є простим у реалізації та має швидкість роботи, але він може бути не дуже ефективним у складних завданнях або при наявності великої кількості взаємозв'язаних ознак.

Метод опорних векторів. Це метод, який розділяє простір ознак за допомогою гіперплощини так, щоб він максимально відокремлював приклади різних класів. У контексті NLP, SVM може використовуватися для розділення текстів на класи, такі як позитивний та негативний настрої, теми новин тощо. SVM є ефективним для роботи з великими наборами даних та забезпечує високу точність класифікації.

Підхід глибокого навчання в NLP полягає в застосуванні глибоких нейронних мереж для розуміння та генерації природної мови. Нижче наведено декілька нейромереж для реалізації цього підходу.

Рекурентні нейронні мережі. Цей тип мережі враховує контекст тексту, використовуючи попередні відомості при обробці наступних слів. RNN часто використовуються для завдань, таких як машинний переклад, аналіз настроїв та генерація тексту (рисунок 1.1).



Риснуок 1.1 – Рекурентна нейромережа [7]

Згорткові нейронні мережі. CNN ефективно використовуються для аналізу тексту, зокрема для класифікації та виявлення патернів у тексті (рисунок 1.2).

Трансформери. Це недавній прорив у глибокому навчанні, який став основою для багатьох передових моделей у сфері NLP, таких як BERT, GPT та інші. Вони здатні ефективно моделювати довгострокові залежності у тексті та генерувати високоякісний контент.

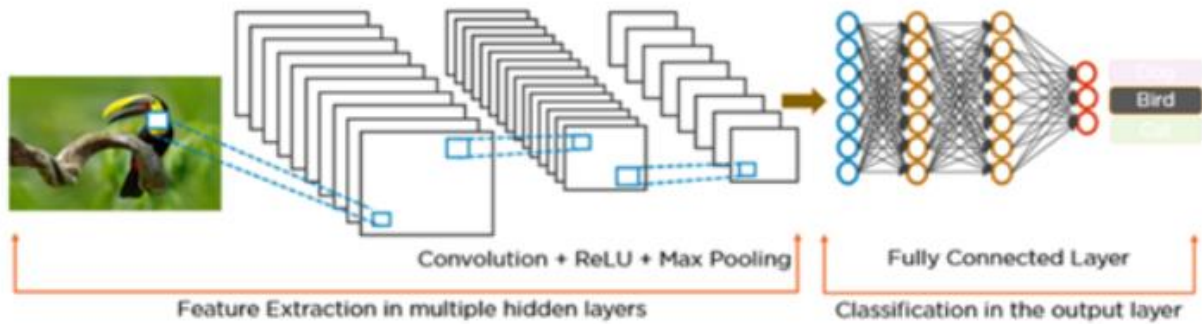


Рисунок 1.2 – Згорткова нейронна мережа [7]

Підхід глибокого навчання в NLP вимагає великих обсягів даних та обчислювальних ресурсів для навчання, але він зазвичай досягає найкращих результатів у багатьох завданнях обробки природної мови.

Отже, з розглянутих теоретичних підходів буде використано підхід глибокого навчання, оскільки така задача як виявлення стресу у повідомленні містить велику кількість складних ознак, а також приховані залежності.

1.3 Аналіз існуючих програмних засобів та наукових рішень у галузі виявлення стресового стану за текстом

Психологічний стрес становить значну загрозу для благополуччя людей, і своєчасно виявити стрес і впоратися з ним може бути доволі складно. Тому дана тема широко привокуює увагу науковців та розробників. Зі збільшенням використання платформ соціальних мереж люди звикли ділитися своїми повсякденними справами та спілкуватися з друзями в Інтернеті. Це дає можливість використовувати дані, отримані з цих онлайн-соціальних мереж для завчасного виявлення стресу. У дослідницькій роботі авторів [8] набір даних у режимі реального часу збирається для дослідження кореляції між рівнем стресу користувачів та їхньою соціальною взаємодією. Отриманий набір атрибутів, пов'язаних зі стресом, що включає текстові, візуальні та соціальні аспекти, служить індикатором рівня стресу в користувачів. У дослідженні пропонується нова модель, що може використовувати отримані атрибути, пов'язані зі стресом,

для покращення ефективності виявлення. Пропонована модель враховує стресові стани як користувачів, так і їхніх друзів у соцмережі. Аналізуючи взаємодію користувачів і спільний вміст, можна точно оцінити рівень стресу, а щоб зробити результати цього дослідження доступними для громадськості, створено вебсайт, який дозволяє користувачам оцінювати рівень стресу та досліджувати відповідні види діяльності. Використовуючи розроблену платформу, люди можуть отримати уявлення про власний рівень стресу та отримати доступ до ресурсів для ефективного керування стресом. За допомогою проведених експериментів та аналізу продемонстровано високу ефективність запропонованої моделі у виявленні стресу. Інтеграція даних соціальних мереж і атрибутів, пов'язаних зі стресом, дозволила значно підвищити точність виявлення стресу, забезпечуючи проактивну допомогу та підтримку особам, які переживають психологічний стрес.

У дослідженні [9] автори прагнуть розширити аналіз настроїв і емоцій для виявлення стресу людини на основі публікацій і коментарів, якими користувач поділився на платформах соціальних мереж. Вміст соціальних мереж здебільшого використовується для огляду, оцінки, впливу чи аналізу настроїв. Автори використовують великомасштабні набори даних із твітами, щоб виконати аналіз настроїв за допомогою алгоритмів машинного навчання та моделі глибокого навчання BERT для класифікації настроїв. Також запропоновано використання Latent Dirichlet Allocation, що є неконтрольованим методом машинного навчання для сканування групи документів, розпізнавання шаблонів слів і фраз у них, а також збору груп слів і подібних виразів, які найбільш точно ілюструють набір документів. Це допомагає передбачити, яка тема пов'язана з текстовими даними. За допомогою цих моделей можна виявляти емоції користувачів онлайн. Крім того, ці емоції можна використовувати для аналізу стресу чи депресії. Підсумовуючи, моделі ML і модель BERT мають дуже хороший рівень виявлення. Це дослідження корисне для благополуччя психічного здоров'я, а його результати оцінюються за допомогою різних

показників на макро- та мікрорівнях і вказують на те, що навчена модель визначає стан емоцій на основі соціальних взаємодій.

Дослідники Арья та Мішра представляють огляд застосування машинного навчання в секторі охорони здоров'я, їх обмеження, прогнозний аналіз і виклики в цій галузі, які потребують передових досліджень і технологій. Автори переглянули статті про виявлення психічного стресу за допомогою ML, які використовували сайти соціальних мереж, блоги, дискусійні форуми, техніку анкетування, набір клінічних даних, дані в реальному часі, технологію біосигналів (ЕКГ, ЕЕГ), бездротовий пристрій і суїцидальну схильність. Дослідження показує високий потенціал алгоритмів ML у психічному здоров'ї [10]. Автори Aldarwish та ін. використовували алгоритми машинного навчання SVM і Naive-Bayesian для прогнозування стресу з UGC-контенту, створеного користувачами, на сайтах соціальних мереж (Facebook, Twitter, Live Journal). У дослідженні були використані набори даних стресу соціальної взаємодії на основі настрою та негативізму, а також BDI-опитувальник із 6773 повідомлень, 2073 депресивних, 4700 недепресивних (текстових). Автори досягли точності 57% від SVM та 63% від Naive-Bayesian. Також авторами наголошувалось на виявленні стресу за допомогою методів великих даних [11].

Авторами [12] зазначено, що стрес є нормальним явищем у сучасному світі, і він змушує людей реагувати на різноманітні фактори, що призводить до фізіологічних і поведінкових змін. Якщо людина надто довго пам'ятає про стрес, це вплине на тіло. Багатьох захворювань, пов'язаних зі стресом, можна уникнути, якщо стрес виявити раніше. Коли людина перебуває у стані стресу, закономірність можна виявити за допомогою різних біосигналів, таких як теплові, електричні, імпедансні, акустичні, оптичні тощо, і за допомогою цих біосигналів можна визначити рівень стресу. У роботі використовується набір даних, отриманий за допомогою датчика Інтернету речей (ІОТ), що призвело до збору інформації про реальну життєву ситуацію, пов'язану з психічним здоров'ям людини. Щоб отримати шаблон для виявлення стресу, були зібрані дані з таких датчиків, як датчик шкірно-гальванічної реакції (GSR) і

електрокардіограма. Потім набір даних буде класифіковано за категоріями за допомогою багаторівневого перцептрона (MLP), дерева рішень (DT), К-найближчого сусіда (KNN), опорної векторної машини (SVM) і алгоритмів глибокого навчання (DL). Точність, влучність, запам'ятовуваність і F1-Score використовуються для оцінки ефективності даних. Нарешті, серед усіх класифікаторів машинного навчання найкращу продуктивність показало Дерево рішень, де точність DT становить 95%, точність 96%, запам'ятовування 96% і F1-оцінка 96%.

Окрім уваги вчених до питання виявлення стресу також є деякі напрацювання у сфері застосувань. Одним з таких застосувань є Worry Watch. Це мобільний сервіс, покликаний стежити за психологічним станом свого власника. Користувачам потрібно відзначати у додатку рівень занепокоєння, його причини, відстежувати історію виникнення тривожності та аналізувати рівень напруженості протягом певного періоду [13] (рисунок 1.3).

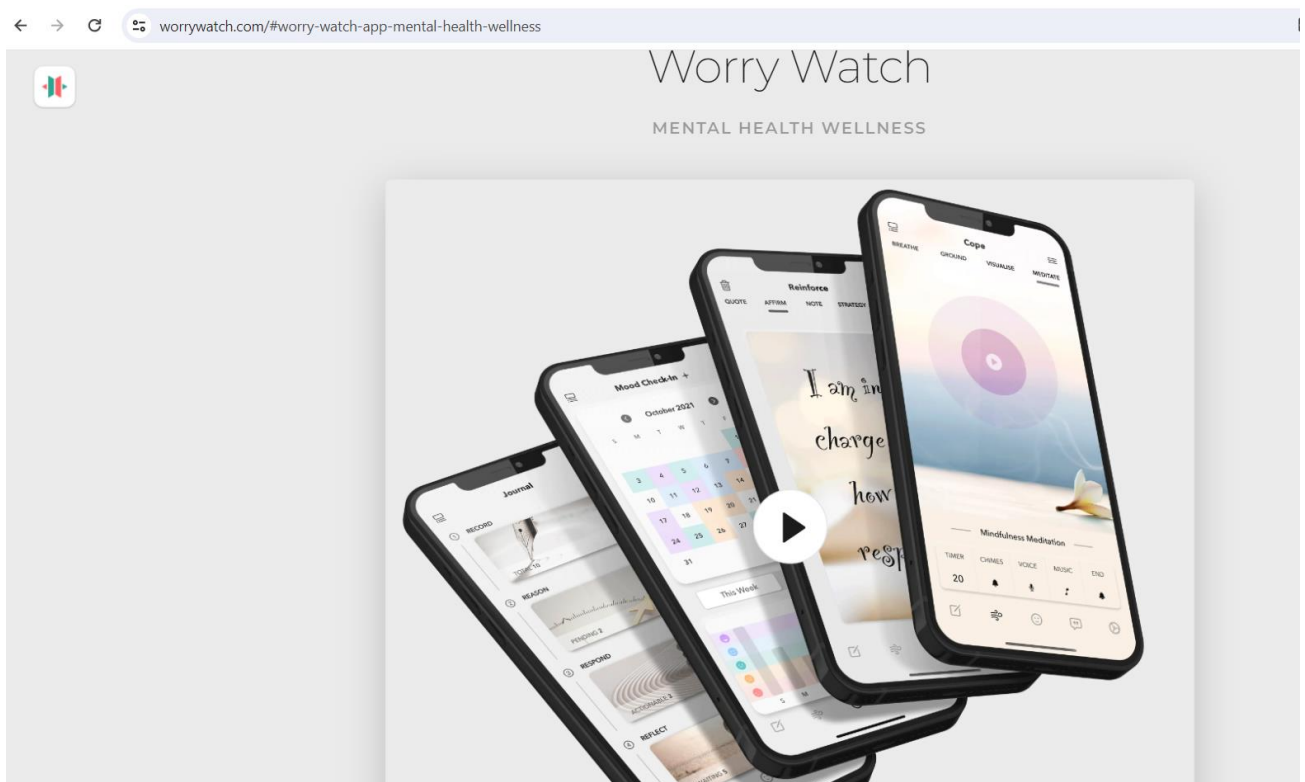


Рисунок 1.3 – Сайт Worry Watch [14]

Розробники вважають, що ця програма є хорошим способом стежити за своїм станом. Вона має допомогти тим людям, які часто відчують стрес, безпричинні страхи та тривогу, і хочуть з'ясувати, чому так відбувається. Програма допоможе відстежити такі моменти і, можливо, знайти причину переживань.

Ще одним за стосунком є StressPal, що є не лише сайтом, а й комплексною платформою, що пропонує різні інструменти та ресурси для подолання стресу та покращення психічного здоров'я [15].

Програми для самопомоги. StressPal Frontline, програма, розроблена спеціально для людей, які працюють у сфері першої допомоги та рятувальних служб. Вона допомагає їм краще справлятися зі стресом, пов'язаним з їхньою роботою. Вигляд сайту наведено на рисунку 1.4.

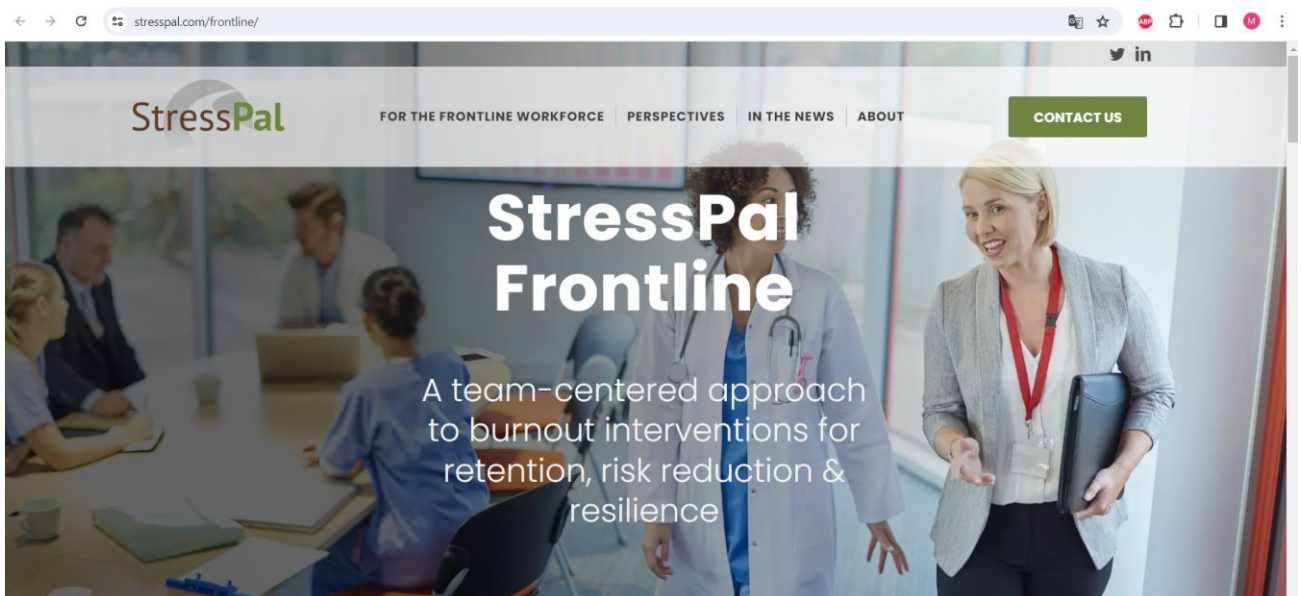


Рисунок 1.4 – Сайт StressPal [16]

StressPal Resilience. Програма, що допомагає людям розвивати стійкість до стресу та покращувати загальне психічне здоров'я.

StressPal Sleep. Програма, що допомагає людям з проблемами сну, пов'язаними зі стресом.

Також StressPal має мобільний додаток, доступний для iOS та Android, який пропонує:

- щоденник настрою, який дозволяє відстежувати ваш настрій протягом дня;
- бібліотеку ресурсів, що містить статті, відео та інші ресурси про подолання стресу;
- вправи на релаксацію, такі як медитація, дихальні вправи та інші вправи, що допомагають розслабитися.
- можливість спілкуватися з іншими користувачами, є форум, де можна ділитися своїм досвідом та отримувати підтримку від інших людей.

Отож, зважаючи на факт того, що стрес є практично невід'ємною деструктивною складовою життя людини. Переживання стресу вимагає значних енергетичних затрат від людини. Тому тривалий вплив стресових факторів може призводити до відчуття слабкості, апатії та втоми. Тому є потреба в автоматизованому виявленні стресу, що може допомогти людям краще справлятися з проблемами та підвищить якість життя. З розглянутих теоретичних підходів було прийнято рішення використовувати підхід глибокого навчання, оскільки така задача як виявлення стресу у повідомленні містить велику кількість складних ознак, а також приховані залежності.

Зважаючи на широке коло науковців, чия увага прикута до проблеми виявлення стресу за дописами, а також зважаючи на спроби існуючих програмних рішень, напрямок є актуальним. Проте, існуючих досліджень недостатньо для вирішення проблеми автоматизованого виявлення стресу, і більшість існуючого ПЗ працює тільки з англійською мовою. Тому подальша розробка є актуальною та доцільною.

1.4 Мета, задачі та вимоги до реалізації інформаційної системи

Метою кваліфікаційної роботи бакалавра є спрощення виявлення стресового стану за дописами з соціальних інтернет-мереж за рахунок аналізу дописів NLP-засобами, що досягається шляхом розробки методу виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами, а

також відповідної програмної реалізації, що буде використовувати створений метод.

Для досягнення мети потрібно виконати наступні задачі:

- виконати дослідження предметної області виявлення стресового стану за дописами;
- в рамках дослідження предметної області виконати огляд теоретичних підходів щодо виявлення стресового стану за дописами;
- виконати аналіз наукових надбань та існуючих програмних рішень в області виявлення стресового стану за дописами;
- розробити метод виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами;
- на основі розробленого методу виконати проектування інформаційної структури системи;
- виконати підготовку навчальних даних для тренування класифікатора для виявлення стресового стану за дописами;
- здійснити вибір засобів розробки для інформаційної системи;
- здійснити програмну реалізацію інформаційної системи для виявлення стресового стану;
- провести тестування розробленої програмної реалізації;
- здійснити дослідження ефективності розробленого методу з використанням розробленої програмної реалізації.

Розділ 2 Метод виявлення стресового стану за дописами з соціальних інтернет-мереж

2.1 Схема та етапи методу виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами

Метод виявлення стресового стану за дописами з соціальних інтернет-мереж призначений для автоматизованого аналізу текстових повідомлень, що розміщені у соціальних мережах, з метою виявлення стресу у дописах. Схема та етапи методу наведено на рисунку 2.1.

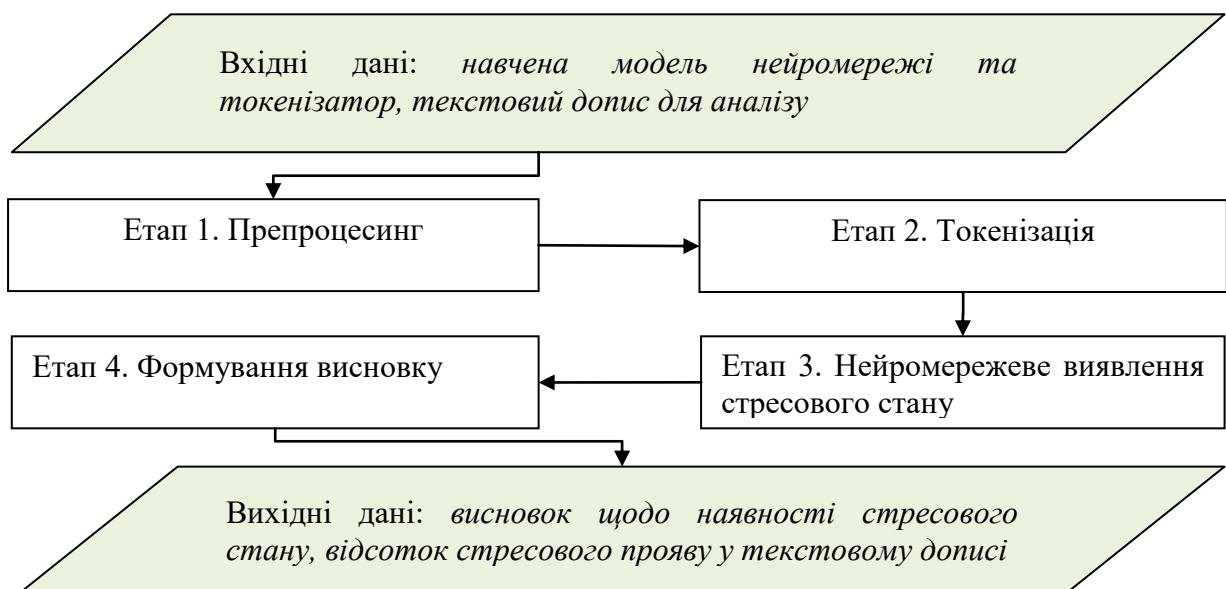


Рисунок 2.1 – Схема та етапи методу виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами

Вхідними даними методу є навчена модель нейромережі та токенізатор, що використовувався під час навчання, а також текстовий допис для аналізу. В рамках виконання кваліфікаційної роботи бакалавра буде використано рекурентну нейромережу архітектури GRU.

Вихідними даними методу є висновок щодо наявності стресового стану, а також відсоток стресового прояву у текстовому дописі.

Першим етапом методу виявлення стресового стану за дописами є препроцесинг, що включає в себе видалення смайлів, знаків пунктуації, стоп-слів тощо.

Наступним етапом є токенизація, яка здійснюється шляхом використання завантаженого у вхідних даних токенизатора, та обробкою ним тексту, що пройшов препроцесинг.

Третім етапом є неймережеве виявлення стресового стану, яке здійснюється шляхом обробки завантаженою навченою неймережевою моделлю числової послідовності з попереднього етапу.

Четвертим етапом є формування висновку щодо наявності стресового стану у текстовому дописі, що аналізується, а також відсоток його наявності.

Оскільки вхідними даними є навчена неймережева модель, то першочерговою задачею є формування таких моделей для виявлення стресового стану. Схема формування навченої моделі та токенизатора наведена на рисунку 2.2.

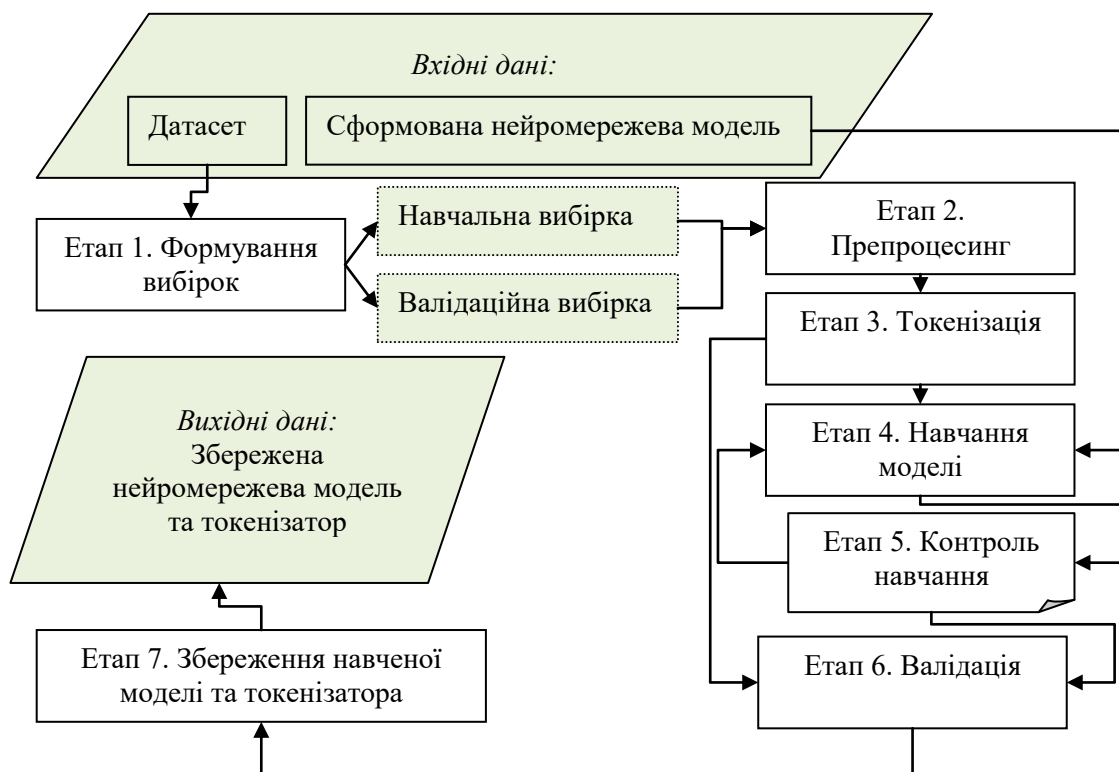


Рисунок 2.2 – Схема та етапи формування навченої неймережевої моделі та токенизатора

Вхідними даними для формування навченої нейромережевої моделі та токенизатора є датасет та сформована нейромережева модель. Вихідними даними є збережена нейромережева модель та токенизатор.

Першим етапом є формування вибірок: навчальної та валідаційної. Вибірki поділяються у співвідношенні 80 на 20 %, де 80 % – навчальна вибірка, 20 % – валідаційна.

Другим етапом є препроцесинг. Обидві вибірки (навчальна та валідаційна) проходять видалення смайлів, знаків пунктуації, стоп-слів тощо.

Наступним етапом є токенизація. Очищені навчальна та валідаційна вибірки проходять процес токенизації, після чого наявні тексти перетворюються у числові послідовності.

Четвертим етапом є навчання нейромережевої моделі архітектури GRU. Навчання моделі здійснюється на основі вхідної нейромережевої архітектури з використанням навчальної вибірки. Цей етап тісно пов'язаний з етапом контролю навчання, який слідує за показниками точності та функції втрат на кожній епосі навчання.

Шостим етапом є етап валідації, коли вже натренована модель пробує класифікувати зразки з валідаційної вибірки, на основі чого формуються оцінки нейромережевої моделі за метриками Accuracy, Precision, Recall.

Наступним етапом є збереження навченої моделі та токенизатора, якщо метрики мають значення вище 80 %. У випадку, якщо значення метрик незадовільні, процес навчання починається спочатку зі зміною архітектури нейромережі у вхідних даних.

Отже, наведено схему та основні етапи методу виявлення стресового стану за дописами з соціальних інтернет-мереж, що перетворює вхідні дані у вигляді навченої моделі нейромережі та токенизатора і текстового допису для аналізу у вихідні дані у вигляді висновку наявності стресового стану та відсотку стресового прояву у текстовому дописі.

2.2 Функціональна структура інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж

Функціональна структура інформаційної системи логічно розбита на 3 підсистеми та головне меню. Інформаційна системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами складається із таких підсистем: «Підсистема налаштування та навчання нейромережі», «Підсистема виявлення стресового стану за дописами» та «Підсистема візуальної аналітики». Взаємодія підсистем наведена на рисунку 2.3.

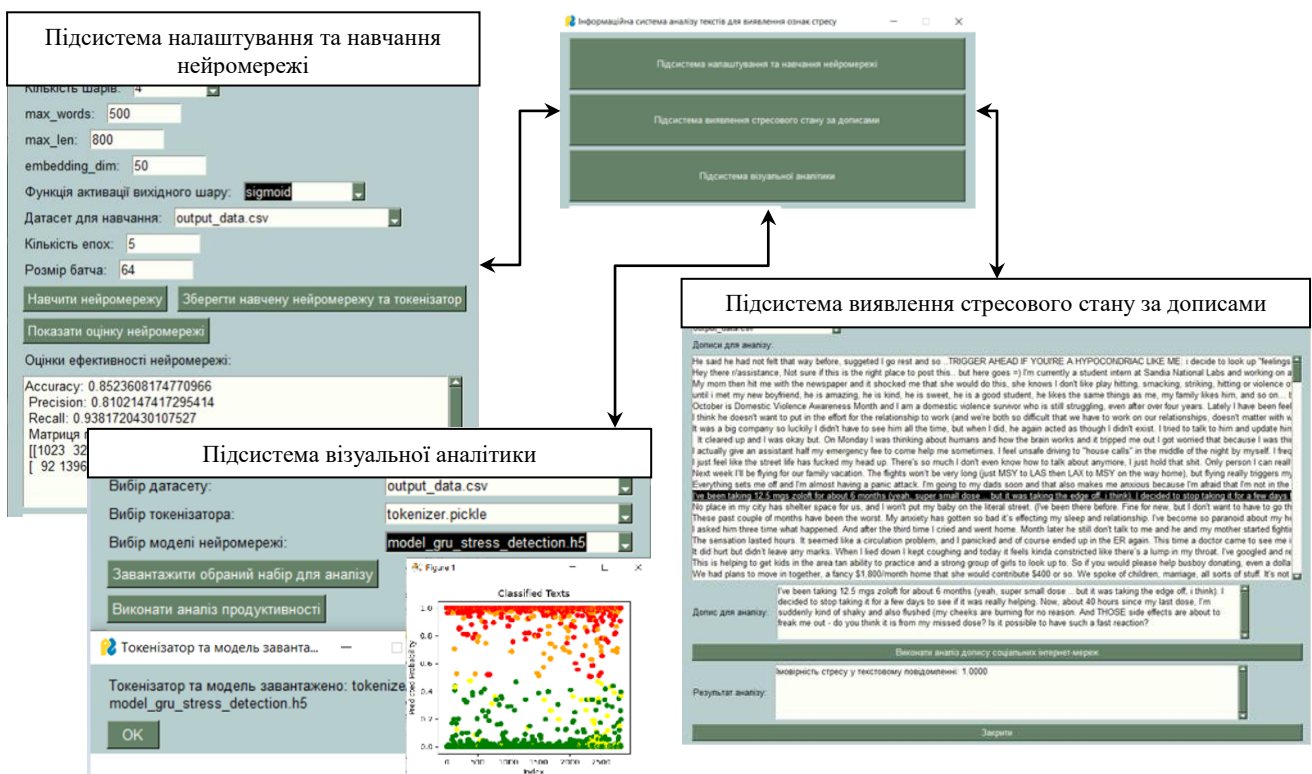


Рисунок 2.3 – Взаємодія підсистем інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами

Підсистема налаштування та навчання нейромережі призначена для навчання та збереження навчених моделей нейромереж, а також їх токенизаторів за користувацькими параметрами. Дана підсистема виконує такі функції:

- уведення кількості епох навчання;
- уведення розміру батча;

- вибір датасету для вибору навчальної та валідаційної вибірок;
- вибір кількості шарів в архітектурі;
- уведення розмірності словника;
- уведення максимальної довжини послідовності токенів;
- навчання нейромережі за користувацькими параметрами;
- проведення аналізу продуктивності за метриками навченої моделі;
- збереження навченої нейромережевої моделі та токенизатора.

Підсистема виявлення стресового стану за дописами є головною підсистемою, яка використовує як вхідні дані навчені попередньою підсистемою класифікатори. Підсистема виявлення стресового стану виконує такі функції:

- вибір датасету для аналізу;
- вибір текстового допису з обраного датасету;
- деталізація обраного допису для аналізу;
- написання текстового допису «вручну»;
- визначення наявності стресового стану в текстовому повідомленні.

Підсистема візуальної аналітики є допоміжною підсистемою навчання нейромереж, адже вона дозволяє розширити функціональні можливості щодо дослідження ефективності навчених класифікаторів. Функціями даної підсистеми є:

- вибір набору даних для аналізу;
- вибір токенизатора для аналізу;
- вибір нейромережевої моделі для аналізу;
- завантаження їх для подальшої перевірки;
- виконання аналізу продуктивності.

Дана підсистема сформує ряд графіків, які покликані візуалізувати помилкову класифікацію стресового стану.

Отже, наведено функціональну структуру інформаційної системи, що складається із трьох основних підсистем: «Підсистема налаштування та навчання нейромережі», «Підсистема виявлення стресового стану за дописами» та «Підсистема візуальної аналітики», які відповідають за описаний функціонал.

2.3 Розробка архітектури нейронної мережі

Архітектура нейронної мережі є важливим аспектом проектування інтелектуальної інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами. У рамках виконання КРБ буде використано рекурентну нейронну мережу архітектури GRU із динамічною кількістю шарів. Архітектура нейромережі наведена на рисунку 2.4.

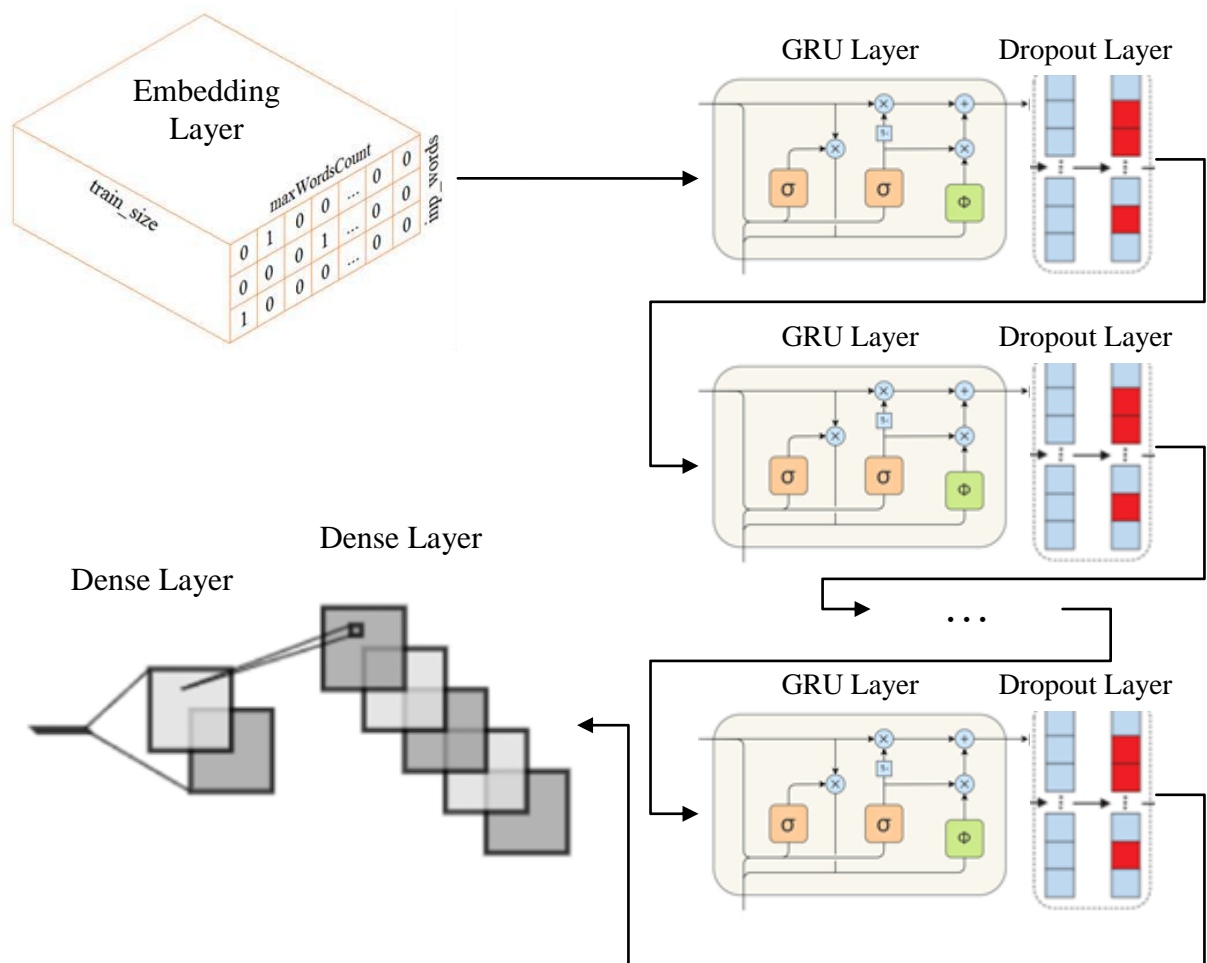


Рисунок 2.4 – Архітектура нейронної мережі

Обов'язковими шарами запропонованої нейромережі є шар вбудування, мінімум один шар GRU, мінімум один шар Dropout та два повнозв'язних шари Dense.

Шар вбудування (**EmbeddingLayer**) має вхідними даними кількість слів в словнику та розмірність векторного представлення кожного токена у тексті.

Його ключовою задачею є для перетворення послідовностей токенів у векторні представлення фіксованої довжини. Основна мета цього шару – вивчити розподілені представлення слів, таким чином, щоб семантично схожі слова були розташовані у просторі близько одне до одного.

Наступними складовими архітектури йде зв'язка шарів GRU та шару Dropout. GRU є типом рекурентного шару, який дозволяє моделі враховувати залежності в послідовностях та здійснювати передачу інформації через часові кроки, у свою чергу Dropout дозволяє запобігти перенавчанню. Кількість таких зв'язок визначається експериментальним шляхом в залежності від показників метрик. Проте у рамках виконання кваліфікаційної роботи бакалавра кількість таких складових не перевищує п'яти.

Останніми шарами йдуть два повнозв'язні шари (Dense Layer). У передостанньому повнозв'язному шарі використовується функція активації ReLu, щоб нейрон міг передавати тільки додатні значення. Це допомагає уникнути проблеми від зникнення градієнту та поліпшує швидкість збіжності під час навчання.

Останній повнозв'язний шар має один нейрон, та динамічну функцію активації (обирається експериментальним шляхом, проте за замовчуванням використовується сигмоїд).

Отже, наведено архітектуру рекурентної нейронної мережі GRU, що буде використана для інтелектуальної інформаційної системи аналізу текстів для виявлення ознак стресу.

2.4 Проектна архітектура інформаційної системи виявлення стресового стану за дописами та взаємозв'язок компонентів

Проектна архітектура інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами та взаємозв'язок компонентів наведено на рисунку 2.5. Інтелектуальна інформаційна система складатиметься із таких підсистем: «Підсистема формування датасету»,

«Підсистема налаштування та навчання нейромережі», «Підсистема виявлення стресового стану за дописами» та «Підсистема візуальної аналітики».

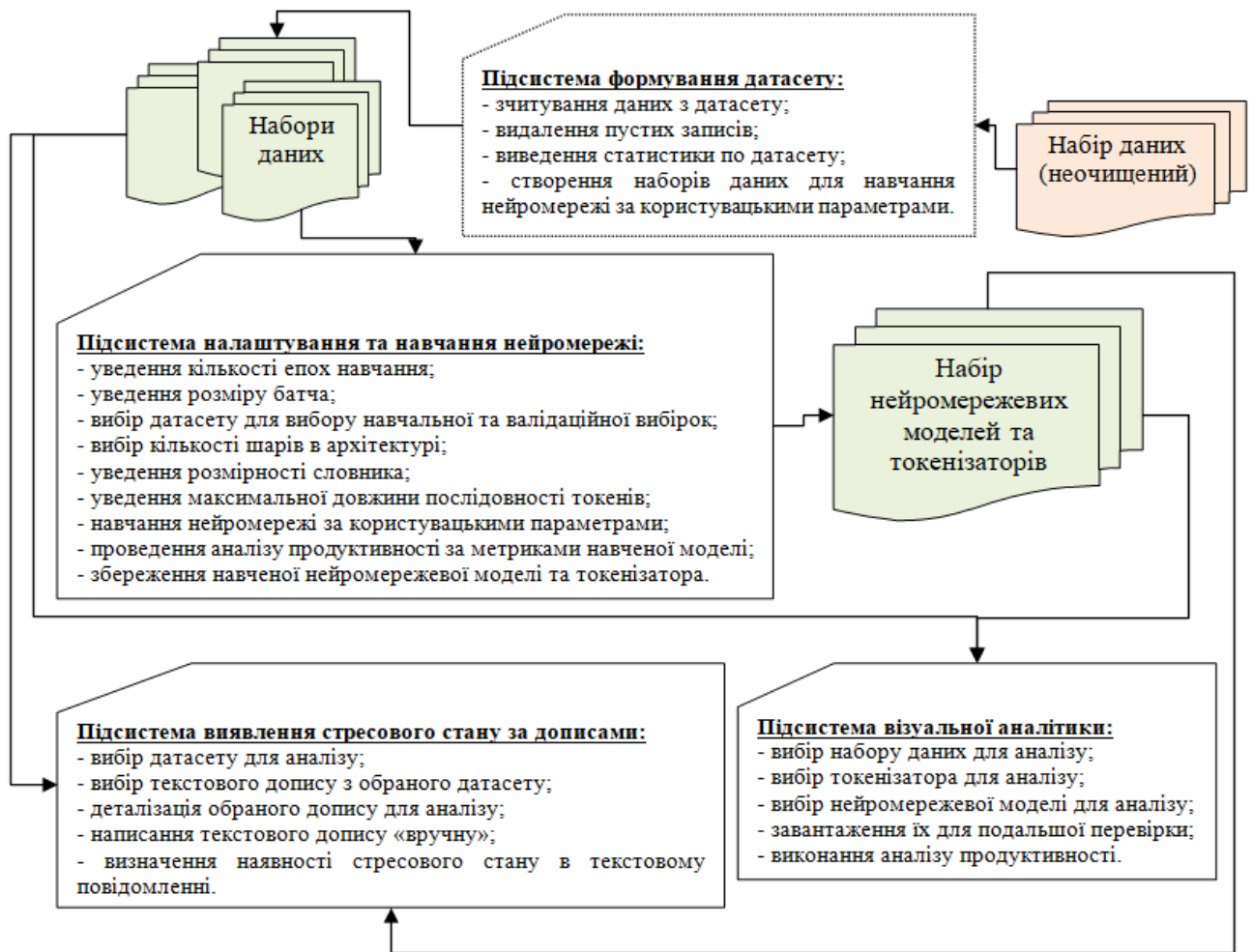


Рисунок 2.5 – Проектна архітектура інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами

Підсистема формування датасету не має графічного інтерфейсу користувача, і призначена для перетворення отриманих неочищених наборів даних в набори даних, які в подальшому будуть використовуватись підсистемами налаштування та навчання нейромережі, виявлення стресового стану за дописами та візуальної аналітики.

Підсистема налаштування та навчання нейромережі за продуктованими наборами даних з попередньої підсистеми відповідає за навчання та збереження нейромережевих класифікаторів. Збережені токенизатори та класифікатори

використовуються підсистемами виявлення стресового стану за дописами та візуальної аналітики.

Отже, наведено проектну архітектуру інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами, що складається із таких підсистем: «Підсистема формування датасету», «Підсистема налаштування та навчання нейромережі», «Підсистема виявлення стресового стану за дописами» та «Підсистема візуальної аналітики» та їх взаємозв'язок.

2.5 Підготовка робочих вхідних даних для інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж

У якості вхідних даних для інтелектуальної інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами у буде використано набір даних із платформи Kaggle з назвою «Human Stress Prediction» [17]. Виявлення психологічного стресу через аналіз текстових даних є складною задачею у сучасній дослідницькій сфері. Різні слова, вживані в публікаціях, можуть вказувати на стан емоційного напруження у людини. З метою розробки та навчання моделей машинного навчання для виявлення психологічного стресу, сформовано набір даних, що містять текстові записи з субреддів, присвячених психічному здоров'ю. Приклад даних наведено на рисунку 2.6.

Дані датасету включають різноманітні проблеми, пов'язані з психічним здоров'ям, які люди обговорюють у онлайн-спільнотах. Кожна публікація в такому наборі даних маркована числом: 0, що означає відсутність стресу, та 1, що вказує на наявність психологічного напруження.

Human Stress Prediction

56

New Notebook

Download

Data Card Code (25) Discussion (1) Suggestions (0)

Stress.csv (1.42 MB) Download Fullscreen Close

Detail Compact Column 7 of 7 columns


subreddit	post_id	sentence_range	text	label	confidence
subreddit is a specific community or forum	unique_id	sentence index	text use for stress detection	0 and 1, 0 means no stress and 1 means stress	confidence on text
ptsd 21%	2343 unique values	(0, 5) 12%	2820 unique values		0.43
relationships 19%		[0, 5] 12%			
Other (1702) 60%		Other (2167) 76%			
ptsd	8601tu	(15, 20)	He said he had not felt that way before, suggested I go rest and so ..TRIGGER AHEAD IF YOU'RE A HYPO...	1	0.8
assistance	81brx9	(0, 5)	Hey there r/assistance, Not sure if this is the right place to post this.. but here goes => I'm curr...	0	1

Рисунок 2.6 – Вигляд датасету «Human Stress Prediction»

Оскільки датасет містить надлишкову інформацію, дані були попередньо оброблені, а саме на основі даного набору даних було сформовано альтернативний, який містив тільки дві колонки: текст та мітку наявності стресу. Також набір даних було відсортовано, і відкинуто порожні записи. В початковому наборі даних було понад 5000 записів, з яких після очистки залишилось близько 4000. Було відкинуто тексти, що мають довжину менше 100 символів, та тексти, що мають довжину понад 1500 символів. Результуючий датасет складається із 2100 записів категорії «стрес» та 1900 записів, що належать категорії «не стрес» (рисунок 2.7).

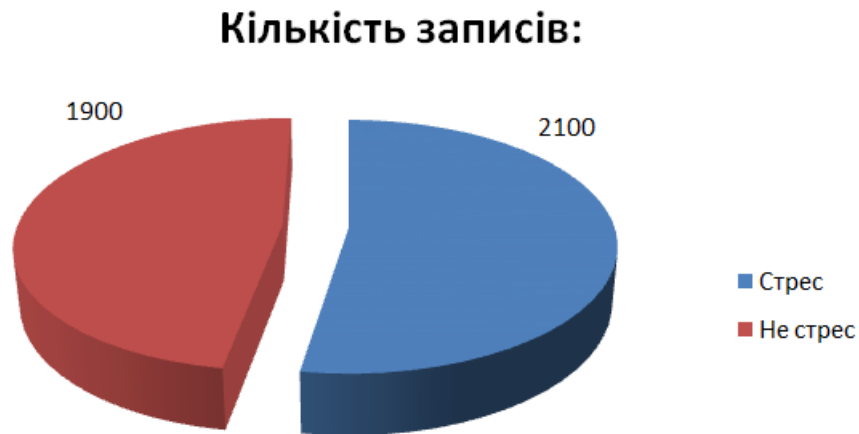


Рисунок 2.7 – Розподіл даних за класами

Отже, у якості набору даних для навчання нейромережових класифікаторів та вхідних даних інформаційної системи виявлення стресового стану за дописами буде використано модифікований набір даних «Human Stress Prediction», що складається із 2100 записів категорії «стрес» та 1900 записів, що належать категорії «не стрес».

2.6 Особливості використання спеціалізованих програмних компонентів інформаційної системи виявлення стресового стану

У рамках виконання кваліфікаційної роботи бакалавра буде використано ряд спеціалізованих програмних компонентів, які спростять процес розробки інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами.

Модуль «Csv» у Python забезпечує функціонал для роботи з CSV-файлами. Це інструмент для читання та запису даних у форматі CSV, що є одним із найпоширеніших форматів для обміну даними [18]. Він дозволяє взаємодіяти зі структурованими даними, представленими у вигляді таблиць, де кожен рядок відповідає одному запису, а кожний стовпець – одному полю. Цей модуль буде використано для роботи підсистеми формування датасету, так як завантажений з Kaggle набір даних має формат .csv. Також даний модуль буде використано у підсистемах, що взаємодіють з датасетами для зчитування даних.

Бібліотека «TensorFlow» є відкритою бібліотекою для машинного навчання, розроблена компанією Google. Вона надає різноманітні інструменти для розробки та навчання моделей штучних нейронних мереж [19]. Однією з основних особливостей TensorFlow є його графова структура обчислень, де операції представляються у вигляді вузлів, а дані – у вигляді ребер графа. Це дозволяє виконувати паралельні обчислення та оптимізувати ефективність виконання складних моделей.

Одним з найбільших переваг TensorFlow є його широкий спектр можливостей для машинного навчання. Він містить в собі різноманітні модулі для роботи з нейронними мережами, зокрема підтримку згорткових, рекурентних та трансформерних мереж. TensorFlow також включає інструменти для обробки даних, візуалізації результатів та оптимізації моделей.

Завдяки своїй популярності та активному співтовариству розробників, TensorFlow є однією з провідних бібліотек для машинного навчання у світі. Вона широко використовується у великих корпораціях, дослідницьких лабораторіях та стартапах для розв'язання різноманітних завдань у галузі штучного інтелекту та аналізу даних. TensorFlow постійно оновлюється та розширюється, що робить його потужним інструментом для розробників у сфері машинного навчання. В рамках виконання кваліфікаційної роботи бакалавра буде використано для навчання рекурентних нейромереж.

«Pandas» є потужною бібліотекою для аналізу даних та маніпуляцій з ними у середовищі Python [20]. Вона надає інструменти для роботи з табличними даними, такими як датафрейми, що дозволяють легко і ефективно виконувати операції з даними, такі як фільтрація, сортування, групування та агрегація. Однією з ключових переваг «Pandas» є його простота використання та широкий спектр функціоналу, що робить його популярним інструментом у сфері аналізу даних.

Однією з найважливіших особливостей «Pandas» є його здатність працювати з різними джерелами даних. Він може імпортувати дані з різних форматів, таких як CSV, Excel, SQL, JSON та інші, і перетворювати їх у

датафрейми для подальшого аналізу. Крім того, Pandas надає інструменти для збереження даних у різних форматах, що дозволяє обмінюватися даними між різними програмами та середовищами. В рамках кваліфікаційної роботи бакалавра, дана бібліотека буде використовуватись для обробки навчальної вибірки для навчання нейромережі.

Бібліотека «sklearn.metrics» входить до складу бібліотеки «scikit-learn» і надає широкий спектр метрик для оцінки якості моделей машинного навчання. Ці метрики використовуються для вимірювання того, наскільки добре модель прогнозує реальні дані. Оцінка якості моделі є важливою частиною розробки і валідації моделей, а «sklearn.metrics» надає набір інструментів для цього [21].

Модуль включає в себе різні типи метрик, такі як метрики класифікації, регресії, кластеризації та векторної кількісної оцінки. Для задач класифікації, доступні метрики, такі як точність (accuracy), матриця помилок (confusion matrix), влучність (precision), відгук (recall), F1-мера (F1-score) та інші. Цей модуль дозволяє оцінювати різні аспекти моделі, використовуючи вбудовані функції для обчислення метрик. Використання цих метрик в роботі дозволить визначити, наскільки добре модель вирішує конкретну задачу та допомагає вибрати найбільш підходящу модель для конкретного завдання.

«matplotlib.pyplot» є підмодуль бібліотеки «Matplotlib», який надає інтерфейс для створення графіків та візуалізації даних у середовищі Python. Він є одним з найпоширеніших інструментів для візуалізації даних та графічного представлення результатів аналізу [22].

За допомогою «matplotlib.pyplot» можна створювати різноманітні типи графіків, такі як лінійні графіки, точкові діаграми, стовпчикові графіки, гістограми, кругові діаграми та багато інших. Цей інструмент надає розгалужені можливості для налаштування вигляду графіків, включаючи колір, розмір, стиль ліній, маркери точок та інше. Даний модуль буде використано для побудови графіків підсистеми візуальної аналітики.

«PySimpleGUI» є бібліотекою для створення графічного інтерфейсу користувача в середовищі Python. Вона відома своєю простотою використання та

зручністю для швидкого створення інтерактивних програм з візуальним інтерфейсом [23].

Однією з особливостей «PySimpleGUI» є його простий та інтуїтивно зрозумілий синтаксис, що дозволяє швидко створювати різноманітні вікна, кнопки, поля введення, списки та інші елементи інтерфейсу. Вона має вбудовану підтримку різних тем оформлення, що дозволяє легко налаштувати вигляд програми під свої потреби.

«PySimpleGUI» є багатоплатформовою бібліотекою, тобто програми, що створені з його використанням, можуть працювати на різних операційних системах, таких як Windows, macOS та Linux, без змін у коді. Дана бібліотека буде використана для розробки інтерфейсу користувача.

Отже, для розробки інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами буде використано такий набір програмних розширень: модуль «Csv» буде використано у підсистемах, що взаємодіють з датасетами для зчитування даних, «TensorFlow» для навчання рекурентних неймереж, «Pandas» для обробки навчальної вибірки для навчання неймережі, «sklearn.metrics» для оцінки якості навчання неймережевої моделі, «matplotlib.pyplot» для швидкого створення інтерактивних програм з візуальним інтерфейсом.

2.7 Висновки до розділу 2

Створено метод виявлення стресового стану за дописами з соціальних інтернет-мереж, що призначений для автоматизованого аналізу текстових повідомлень, які розміщені у соціальних мережах, з метою виявлення стресу у дописах.

Наведено функціональну структуру інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами, що складається із трьох основних підсистем із графічним інтерфейсом: «Підсистема налаштування та навчання неймережі», «Підсистема виявлення стресового

стану за дописами» та «Підсистема візуальної аналітики», які відповідають за описаний функціонал.

Наведено проектну архітектуру інформаційної системи, що складається із чотирьох підсистем та описано їх взаємозв'язок.

Наведено архітектуру рекурентної нейронної мережі GRU, що буде використана для інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами.

Обрано набір даних для навчання нейромережевих класифікаторів та вхідних даних інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами. Буде використано модифікований набір даних «Human Stress Prediction», що складається із 2100 записів категорії «стрес» та 1900 записів, що належать категорії «не стрес».

Обрано такий набір програмних розширень: модуль «Csv» буде використано у підсистемах, що взаємодіють з датасетами для зчитування даних, «TensorFlow» для навчання рекурентних нейромереж, «Pandas» для обробки навчальної вибірки для навчання нейромережі, «sklearn.metrics» для оцінки якості навчання нейромережевої моделі, «matplotlib.pyplot» для швидкого створення інтерактивних програм з візуальним інтерфейсом.

За розробленим методом виявлення стресового стану за дописами з соціальних інтернет-мереж необхідно розробити застосунок, за допомогою якого провести дослідження ефективності. Також необхідно окремо провести функціональне тестування створеної інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами.

Розділ 3 Експериментальне дослідження методу виявлення стресового стану за дописами

3.1 Визначення шляхів дослідження та засобів створення програмного забезпечення

Для запропонованого методу виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами необхідно провести дослідження ефективності. Оскільки в своїй основі розроблений метод має рекурентну нейромережу архітектури GRU, є потреба дослідити вплив параметрів навчання нейромережі, таких як кількості епох навчання, розмір батча, кількість шарів в архітектурі, розмірність словника тощо. Оцінюватись нейромережеві моделі будуть з використанням метрик: Accuracy, Precision, Recall та матриця плутанини.

Також окрім дослідження ефективності самих нейромережевих моделей необхідно дослідити функціонал застосунку, що реалізовує розроблений метод, а саме:

- навчання нейромережі за користувацькими параметрами;
- проведення аналізу продуктивності за метриками навченої моделі;
- збереження навченої нейромережевої моделі та токенизатора;
- визначення наявності стресового стану в текстовому повідомленні;
- виконання аналізу продуктивності шляхом застосування візуальної аналітики.

Для перевірки коректності усіх цих функцій необхідно створити відповідні тест-кейси.

Отже, окремо буде досліджено вплив параметрів навчання на якість нейромережевих моделей, та проведено тестування створеного програмного забезпечення засобами тест-кейсів.

3.2 Вибір засобів розробки інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж

У якості засобів розробки інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж буде використано мову програмування Python та інтегрована середовище розробки PyCharm.

Python було обрано, оскільки враховуючи широкі можливості цієї мови, вона ідеально підходить для реалізації різноманітних завдань, включаючи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засоб. Python є однією з найбільш універсальних та популярних мов програмування, що підтверджується його першим місцем у рейтингу TIOBE [24]. Цю мову використовують розробники, аналітики, тестувальники та системні адміністратори завдяки її простоті та розширеним можливостям. Крім того, Python є необхідним інструментом для фахівців з машинного навчання та аналітики даних, що підкреслює його значення в сучасній індустрії. Python має простий і зрозумілий синтаксис, що сприяє швидкому розробленню програмного забезпечення та його подальшій підтримці, а також має широкі можливості для навчання нейромережевих моделей.

У роботі над проектом обрано інтегроване середовище розробки PyCharm, яке надійно підтримує Python і забезпечує широкі можливості для зручного програмування [25]. PyCharm включає в себе різноманітні інструменти, такі як підказки під час роботи з кодом, автодоповнення, інтегровану систему керування версіями та інші, що значно полегшує процес розробки.

Однією із переваг використання PyCharm є його здатність ефективно працювати з великими проектами, а інтуїтивний інтерфейс дозволяє зручно організувати код, робити швидкі зміни та відлагоджувати програми.

Отже, для розробки інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж буде використано мову програмування Python, яка має широкі можливості в галузі розробок програмного забезпечення

пов'язаного зі штучним інтелектом та інтегрована середовище розробки PyCharm.

3.3 Структура та функціональне призначення програмних складових інформаційної системи виявлення стресового стану за дописами

Після етапу проектування інформаційної системи необхідно створити структуру та описати функціональне призначення програмних складових системи. Діаграма класів інформаційної системи наведена на рисунку 3.1.

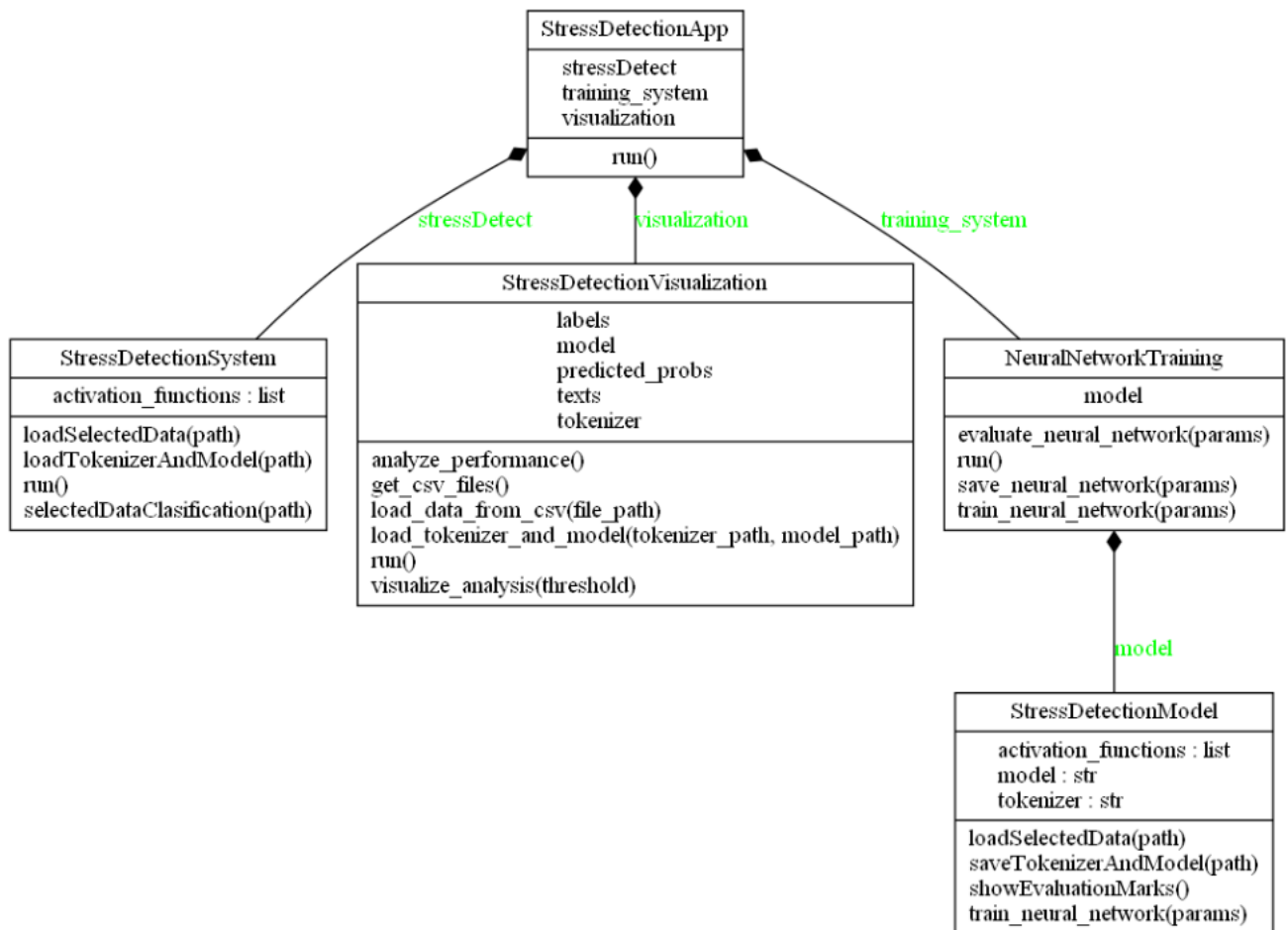


Рисунок 3.1 – Діаграма класів інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж

Клас `NeuralNetworkTraining` призначений для реалізації графічної підсистеми навчання нейромережі. Тісно пов'язаний із класом

StressDetectionModel, у якому міститься алгоритмічна складова щодо етапів навчання нейромережі та токенизатора. Метод `train_neural_network()` реалізовує логіку для навчання нейромережі. Метод `loadSelectedData()` призначений для завантаження з файлової системи набору навчальних даних. Метод `saveTokenizerAndModel()` призначений для збереження навченої нейромережевої моделі та токенизатора у файловій системі. Метод `showEvaluationMarks()` призначений для виведення статистики навчання. У свою чергу методи класу `NeuralNetworkTraining` на кшталт `train_neural_network()` призначені для виклику по натисненню на відповідний контрол на формі користувача, та викликатимуть всередині логіку класу `StressDetectionModel`.

Клас `StressDetectionSystem` призначений для реалізації підсистеми виявлення стресового стану за дописами. Метод `loadTokenizerAndModel()` призначений для завантаження моделі та токенизатора, якими буде здійснюватись виявлення стресового стану. Метод `loadSelectedData()` призначений для завантаження даних на форму користувача з обраного датасету. Метод `selectedDataClasification()` призначений для виявлення стресу у обраному користувачем повідомленні.

Клас `StressDetectionVisualization` призначений для реалізації підсистеми візуальної аналітики. Метод `get_csv_files()` призначений для виведення у випадяючому списку усіх наявних датасетів у папці проєкту. Метод `load_tokenizer_and_model` призначений для завантаження для подальшої роботи обраного токенизатора та нейромережевої моделі. Метод `visualize_analysis()` призначений для виведення графіків які демонструють коректність класифікації даних нейромережею.

Отже, таким чином наведено структуру та описано функціональне призначення програмних складових інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами.

3.4 Особливості реалізації програмних складових інформаційної системи виявлення стресового стану за дописами

За спроектованою функціональною та проектною архітектур інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами, було виконано програмну реалізацію, яка буде ключовою для дослідження ефективності розробленого методу виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами. Підсистема навчання та налаштування параметрів нейромережі має графічний інтерфейс (рисунок 3.2), який дозволяє змінювати користувацькі параметри для навчання.

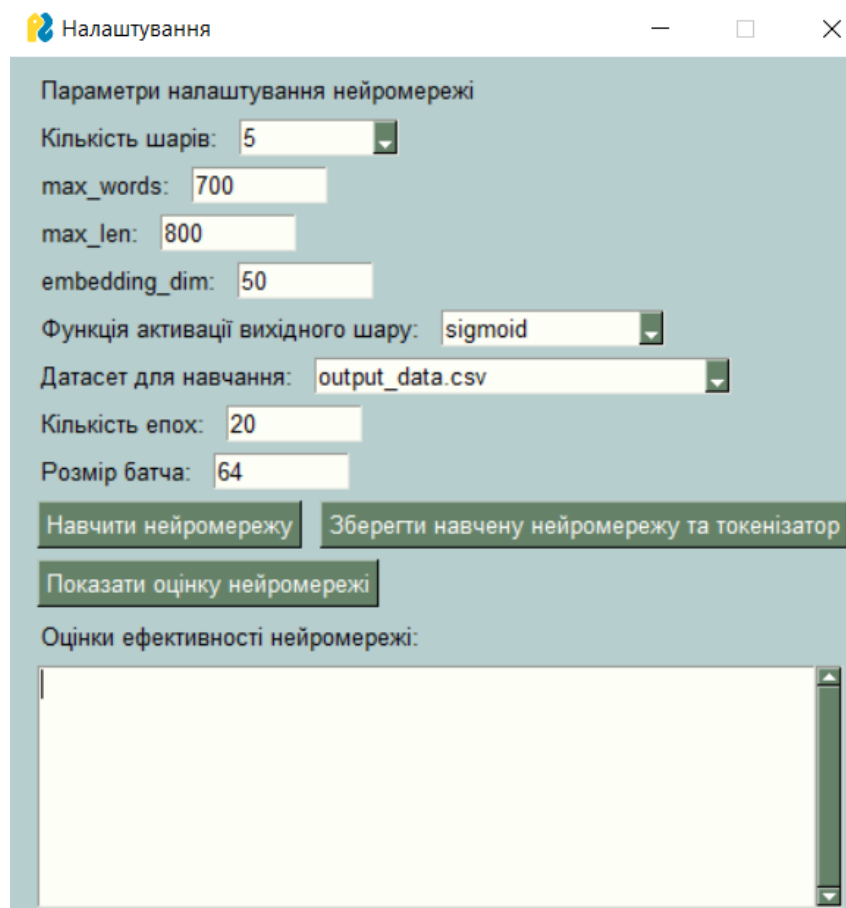


Рисунок 3.2 – Інтерфейс підсистеми навчання нейромережі

Дана підсистема окрім звичних параметрів навчання типу кількості епох та розміру батча також дозволяє змінювати кількість шарів GRU, що може в

рамках програмної реалізації варіюватись від 1 до 7. Шар GRU йде в доповненні з шаром Dropout, щоб уникнути перенавчання.

Оцінка моделі для виявлення стресового стану за допомогою навченої нейронної мережі здійснюється через завантаження щойно навченої і збереженої моделі за допомогою методу `load_model()`. Разом з моделлю завантажуються і токенизатор, який використовувався для попередньої обробки даних перед подачею на вхід нейронну мережу.

Дані для оцінки моделі завантажуються з файлу для валідації, де кожен рядок містить текстовий запис та відповідний мітки класу. Текстові дані перетворюються в послідовності за допомогою токенизатора, а потім досягається необхідна довжина за допомогою доповнення нулями (`padding`) до максимальної довжини, яка визначена з виходу першого шару нейронної мережі `max_len = model.layers[0].output_dim`.

Оцінка здійснюється за допомогою обчислення таких метрик як `accuracy`, `precision`, `recall` і `confusion matrix` з використанням вбудованих функцій від бібліотеки `scikit-learn`. Отримані метрики формуються у вигляді рядка, який відображається у вікні інтерфейсу програми (рисунок 3.3).

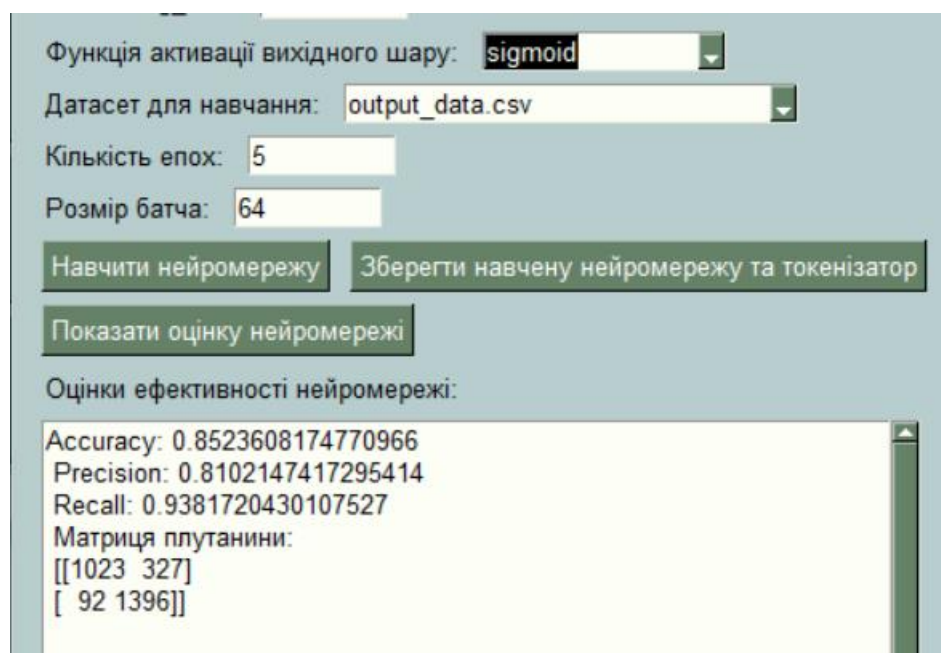


Рисунок 3.3 – Оцінка ефективності нейромережі

Підсистема виявлення стресового стану за дописами функціонує завдяки завантаженню моделі нейромережі та токенизатора, які використовуються для аналізу користувачького посту. Приклад роботи підсистеми виявлення стресового стану наведено на рисунку 3.4.

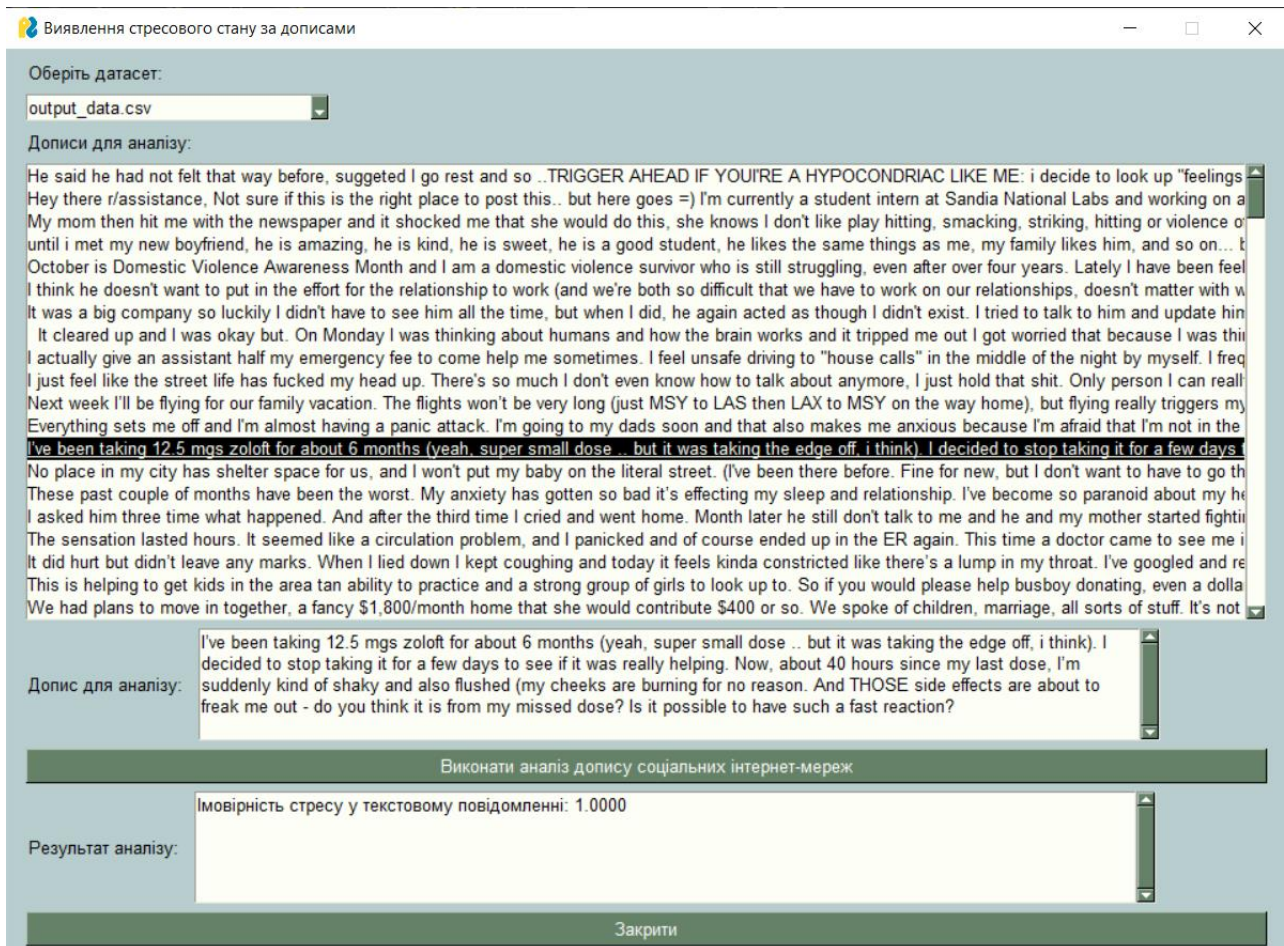


Рисунок 3.4 – Приклад ідентифікації стресового стану

Підсистема візуальної аналітики реалізована з використанням бібліотеки Matplotlib, і основний її функціонал це ілюстрація можливості класифікації текстових даних нейромережевою моделлю. Створюється DataFrame, що містить інформацію про індекс тексту, передбачену ймовірність, істинну мітку та передбачену мітку.

Графіки розділені на два підграфіка, перший графік відображає тексти, які були правильно класифіковані. Зелений колір відповідає правильно

класифікованим текстам без стресу (true label = 0), червоний – правильно класифікованим текстам з наявністю стресу (true label = 1).

Другий графік відображає тексти, які були неправильно класифіковані. Жовтий колір відповідає неправильно класифікованим текстам з наявним стресом (true label = 1, але predicted label = 0), помаранчевий – неправильно класифікованим текстам без стресу (true label = 0, але predicted label = 1).

Графіки наведені на рисунку 3.5.

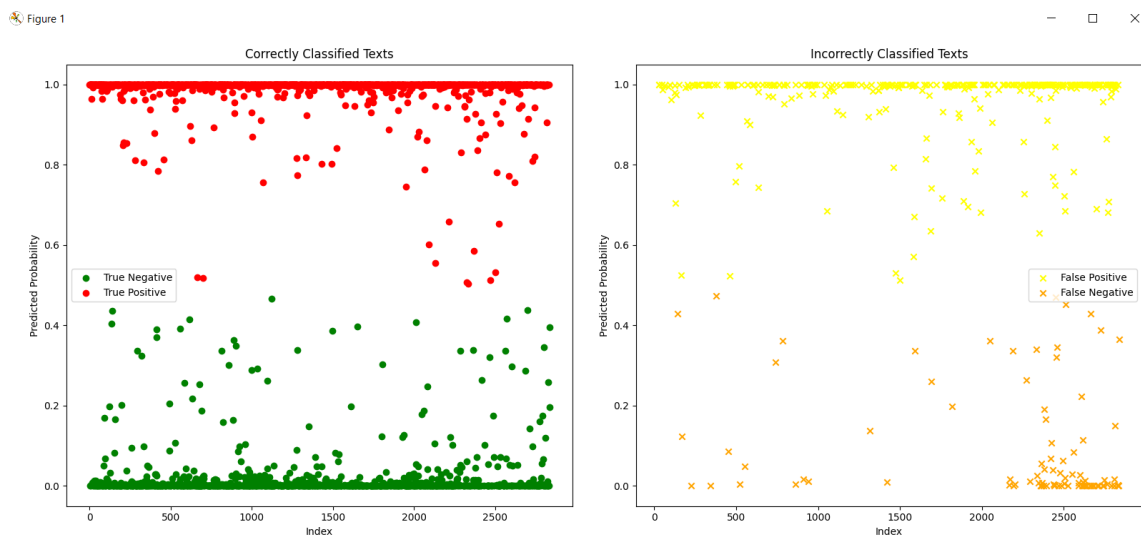


Рисунок 3.5 – Ілюстрація коректної та некоректної ідентифікації стресового стану

Отже, таким чином було здійснено програмну реалізацію методу виявлення стресового стану за дописами з соціальних інтернет-мереж призначений у вигляді інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами, що своїм призначенням має автоматизацію аналізу текстових повідомлень, що розміщені у соціальних мережах, з метою виявлення стресу у дописах.

3.5 Тестування інформаційної системи та вимоги до розгортання

Після створення інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами є потреба протестувати її

та перевірити на предмет наявності непрацюючих функцій. Перевірка функціональності буде виконана засобами тест-кейсів.

Таблиця 3.1 – Тест-кейс 00001

Тест-кейс ID: 00001	Приоритет: 1	Створено: 10.04.2024, Антон ЮРЧАК
Назва: перевірка виконання навчання нейромережі за користувацькими параметрами		
Кроки		Очікуваний результат
<p>1. Відкрити застосунок.</p> <p>2. Перейти на підсистему налаштування та навчання нейромережі, натиснувши однойменну кнопку в головному меню.</p> <p>3. Установити такі параметри: к-сть шарів GRU: 2, max_words: 500, max_len: 20, embedding_dim: 150, функція активації: сигмоїд, к-сть епох: 20, розмір батча 64.</p> <p>4. Натиснути на кнопку «Навчити нейромережу»</p>		<p>Відкрилась форма головного меню</p> <p>Відкрилась підсистема налаштування та навчання нейромережі</p> <p>Розпочато навчання нейромережі, по завершенню якого висвітиться повідомлення про успішне завершення навчання</p>
Результат виконання тест-кейсу: перевірку пройдено успішно.		

Відповідно, буде створено 5 тест-кейсів для перевірки таких основних функцій:

– перевірка виконання навчання нейромережі за користувацькими параметрами;

- перевірка проведення аналізу продуктивності за метриками навченої моделі;
- перевірка збереження навченої нейромережевої моделі та токенизатора;
- перевірка визначення наявності стресового стану в текстовому повідомленні;
- перевірка виконання аналізу продуктивності шляхом застосування візуальної аналітики.

Першим тестовим випадком буде перевірка виконання навчання нейромережі за користувацькими параметрами, кроки тестового випадку наведено в таблиці 3.1.

Тест пройдено успішно, результат виконання тест-кейсу 00001 наведено на рисунку 3.6. Як видно з рисунку, після завершення процесу навчання користувач побачить відповідне повідомлення, і навчена модель в подальшому може бути збережена та досліджена її ефективність.

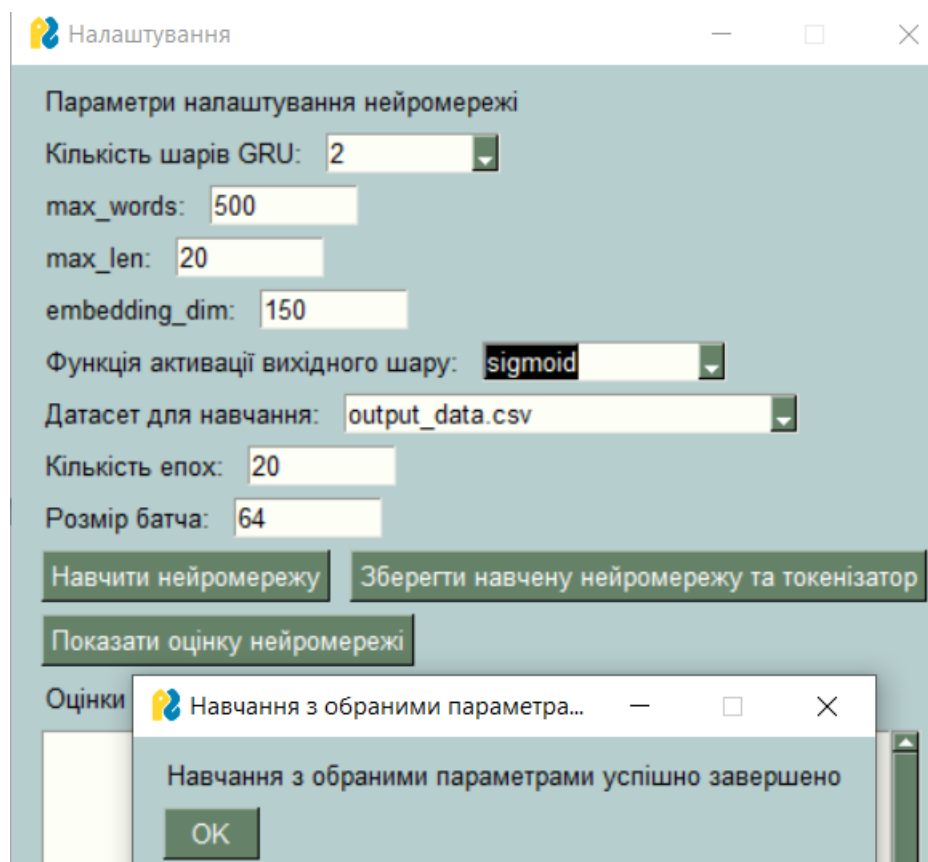


Рисунок 3.6 – Виконання тест-кейсу 00001

Наступним тестовим випадком буде перевірка виконання проведення аналізу продуктивності за метриками навченої моделі, кроки тестового випадку наведено в таблиці 3.2.

Таблиця 3.2 – Тест-кейс 00002

Тест-кейс ID: 00002	Приоритет: 1	Створено: 11.04.2024, Антон ЮРЧАК
Назва: перевірка проведення аналізу продуктивності за метриками навченої моделі		
Кроки		Очікуваний результат
<p>1. Відкрити застосунок.</p> <p>2. Перейти на підсистему налаштування та навчання нейромережі, натиснувши однойменну кнопку в головному меню.</p> <p>3. Установити такі параметри: к-сть шарів GRU: 2, max_words: 500, max_len: 20, embedding_dim: 150, ф-ція активації: сигмоїд, к-сть епох: 20, розмір батча 64.</p> <p>4. Натиснути на кнопку «Навчити нейромережу».</p> <p>5. Натиснути кнопку «Зберегти навчену нейромережу та токенизатор».</p> <p>6. Натиснути кнопку «Показати оцінку нейромережі».</p>		<p>Відкрилась форма головного меню</p> <p>Відкрилась підсистема налаштування та навчання нейромережі</p> <p>Розпочато навчання нейромережі, по завершенню якого висвітиться повідомлення про успішне завершення навчання.</p> <p>Повідомлення про збереження навченої моделі.</p> <p>Виведення оцінок нейромережі у вигляді значень метрик.</p>
Результат виконання тест-кейсу: перевірку пройдено успішно.		

Даний тест-кейс є прямим продовженням попереднього тестового випадку, адже для дослідження ефективності нейромережевої моделі вона повинна пройти етап навчання, а також вона повинна бути збережена разом з токенизатором.

Тест пройдено успішно, результат виконання тест-кейсу 00002 наведено на рисунку 3.7. Після виконання тесту користувач повинен побачити в полі «Оцінки ефективності нейромережі» значення метрик, за якими можна оцінити, на скільки добре нейромережа справляється із поставленою задачею.

The screenshot shows a web-based interface for training a neural network. It includes several input fields and buttons. The parameters are as follows:

- embedding_dim: 150
- Функція активації вихідного шару: sigmoid
- Датасет для навчання: output_data.csv
- Кількість епох: 20
- Розмір батча: 64

Below the parameters are three buttons: "Навчити нейромережу", "Зберегти навчену нейромережу та токенизатор", and "Показати оцінку нейромережі". The "Показати оцінку нейромережі" button is active, displaying the following performance metrics:

```

Оцінки ефективності нейромережі:
Accuracy: 0.8623608174780966
Precision: 0.8202157417297414
Recall: 0.9381720430107527
Матриця плутанини:
[[1024 326]
 [ 92 1396]]
  
```

Рисунок 3.7 – Результат успішного виконання тест-кейсу 00002

Наступним тестовим випадком буде перевірка збереження навченої нейромережевої моделі та токенизатора, кроки тестового випадку наведено в таблиці 3.3.

Після виконання кроків тест-кейсу 00003, можна побачити результат успішного проходження, адже навчена нейромережева модель збережена на жорсткому диску разом з токенизатором (рисунок 3.8).

Таблиця 3.3 – Тест-кейс 00003

Тест-кейс ID: 00003	Пріоритет: 1	Створено: 11.04.2024, Антон ЮРЧАК
Назва: перевірка збереження навченої нейромережевої моделі та токенизатора		
Кроки		Очікуваний результат
1. Відкрити застосунок.		Відкрилась форма головного меню
2. Перейти на підсистему налаштування та навчання нейромережі, натиснувши однойменну кнопку в головному меню.		Відкрилась підсистема налаштування та навчання нейромережі
3. Установити користувацькі параметри.		Розпочато навчання нейромережі, по завершенню якого висвітиться
4. Натиснути на кнопку «Навчити нейромережу».		повідомлення про успішне завершення навчання.
5. Натиснути кнопку «Зберегти навчену нейромережу та токенизатор».		Повідомлення про збереження навченої моделі.
Результат виконання тест-кейсу: перевірку пройдено успішно.		

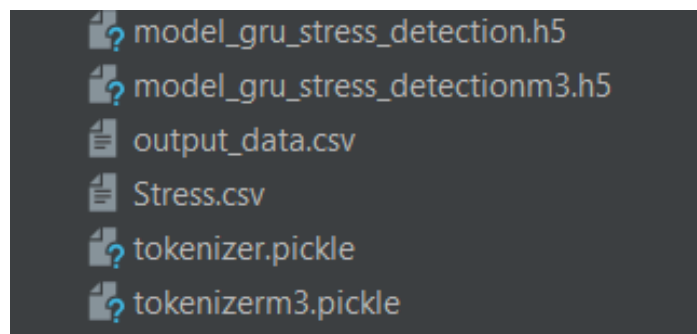


Рисунок 3.8 – Збережені токенизатори та навчені моделі

Наступним тестовим випадком буде перевірка визначення наявності стресового стану в текстовому повідомленні, кроки тестового випадку наведено в таблиці 3.4.

Таблиця 3.4 – Тест-кейс 00004

Тест-кейс ID: 00002	Пріоритет: 1	Створено: 11.04.2024, Антон ЮРЧАК
Назва: перевірка визначення наявності стресового стану в текстовому повідомленні		
Кроки	Очікуваний результат	
<ol style="list-style-type: none"> 1. Відкрити застосунок. 2. Перейти на підсистему виявлення стресового стану за дописами. 3. Обрати опис з переліку наявних. 4. Натиснути на кнопку «Виконати аналіз допису з соціальних інтернет-мереж». 	<p>Відкрилась форма головного меню</p> <p>Відкрилась підсистема виявлення стресового стану за дописами.</p> <p>Обраний допис відобразився в полі «Допис для аналізу».</p> <p>В полі «Результат аналізу» виведено результат щодо наявності стресового стану у тексті.</p>	
Результат виконання тест-кейсу: перевірку пройдено успішно.		

Після виконання кроків з таблиці 3.4, користувач може переконатись в успішному проходженні тесту, результат коректного виконання наведено на рисунку 3.9.

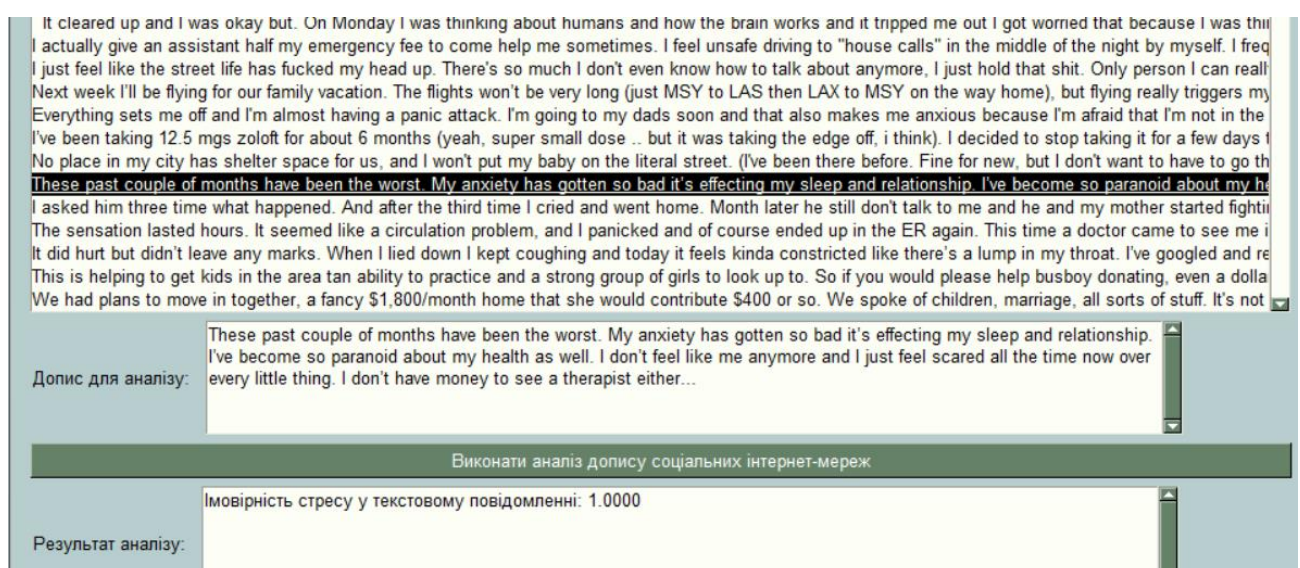


Рисунок 3.9 – Виконання тест-кейсу 00004

Кроки останнього тестового випадку наведено в таблиці 3.5.

Таблиця 3.5 – Тест-кейс 00005

Тест-кейс ID: 00002	Пріоритет: 1	Створено: 15.04.2024, Антон ЮРЧАК
Назва: перевірка виконання аналізу продуктивності шляхом застосування візуальної аналітики		
Кроки		Очікуваний результат
<ol style="list-style-type: none"> 1. Відкрити застосунок. 2. Перейти на підсистему візуальної аналітики. 3. Обрати з переліку наявних такенізатор та нейромережеву модель для аналізу. 4. Натиснути на кнопку «Завантажити обраний набір даних для аналізу». 5. Натиснути кнопку «Виконати аналіз продуктивності». 		<p>Відкрилась форма головного меню</p> <p>Відкрилась підсистема візуальної аналітики.</p> <p>Повідомлення про завантаження обраних моделей.</p> <p>Відобразились графіки аналізу продуктивності</p>
Результат виконання тест-кейсу: перевірку пройдено успішно.		

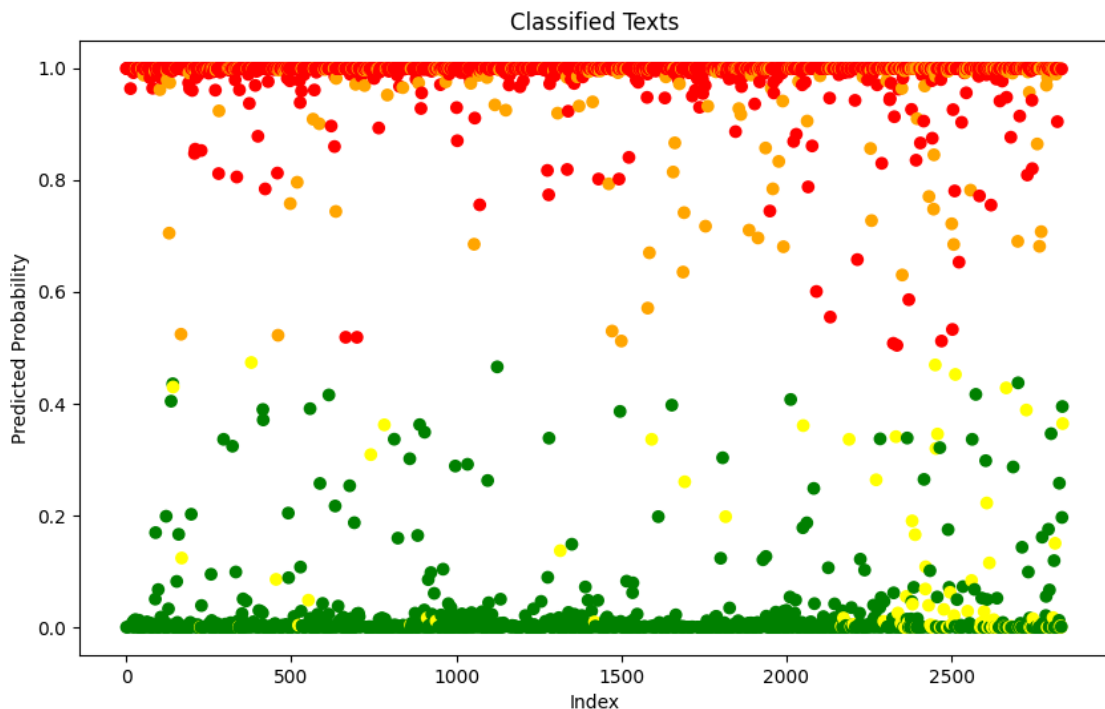


Рисунок 3.10 – Фрагмент виконання візуальної аналітики

Останнім тестовим випадком здійснюється перевірка виконання аналізу продуктивності шляхом застосування візуальної аналітики однойменної підсистеми візуальної аналітики. Результат успішного виконання тест-кейсу 00005 наведено на рисунку 3.10.

Отже, таким чином було здійснено тестування інформаційної системи, яке показало що весь заявлений функціонал працює коректно.

3.6 Аналіз функціональності інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами

Інформаційна система для виявлення стресу у текстових повідомленнях представлена трьома графічними підсистемами та головним меню, з якого можна здійснювати переходи між ними. Вигляд головного меню наведено на рисунку 3.11.

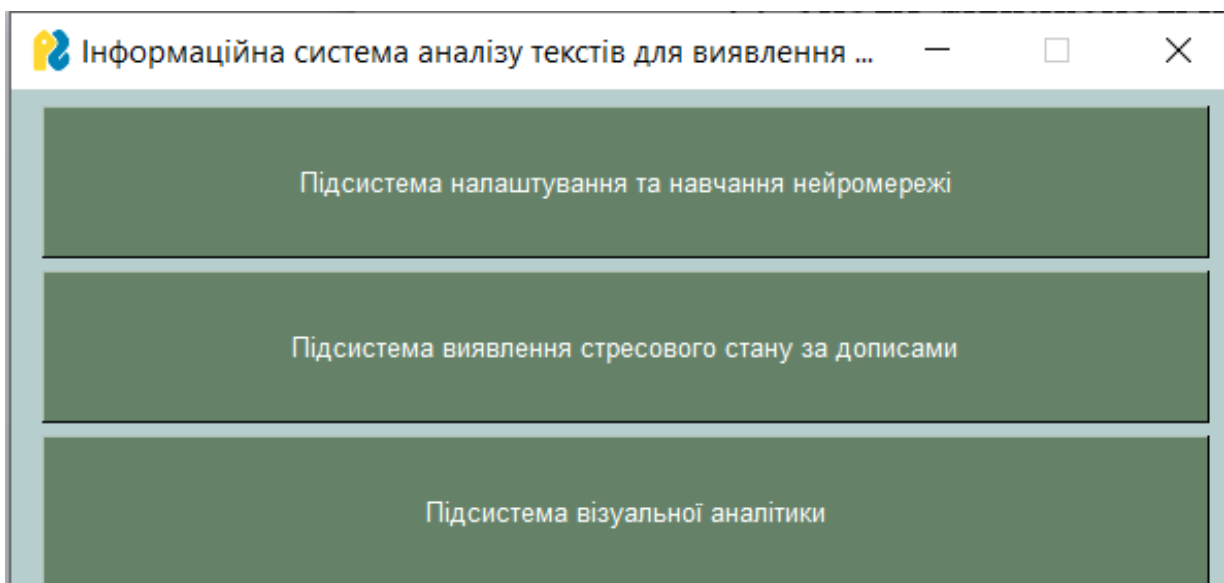


Рисунок 3.11 – Вигляд головного меню інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами

Для переходу на підсистему налаштування та навчання нейромереж необхідно натиснути кнопку «Підсистема налаштування та навчання

нейромережі». Вигляд вікна підсистеми навчання нейромереж наведено на рисунку 3.12.

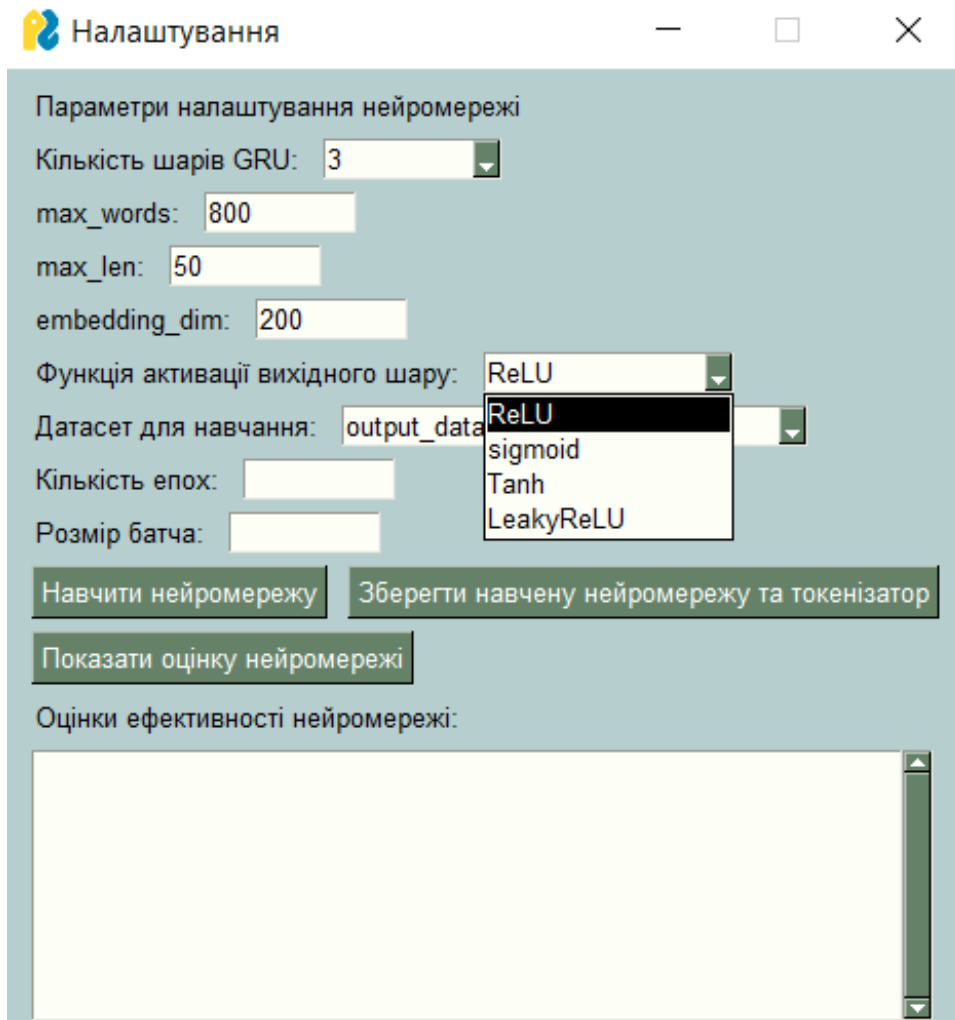


Рисунок 3.12 – Підсистема налаштування та навчання нейромережі

У межах даної підсистеми можна задавати різні користувацькі параметри щодо налаштування процесу навчання нейромережевих моделей. Для навчання нової нейромережевої моделі необхідно заповнити всі користувацькі поля та натиснути кнопку «Навчити нейромережу». Для збереження навченої нейромережевої моделі та її токенизатора необхідно натиснути кнопку «Зберегти навчену нейромережу та токенизатор», після чого користувач побачить повідомлення про збереження (рисунок 3.13).

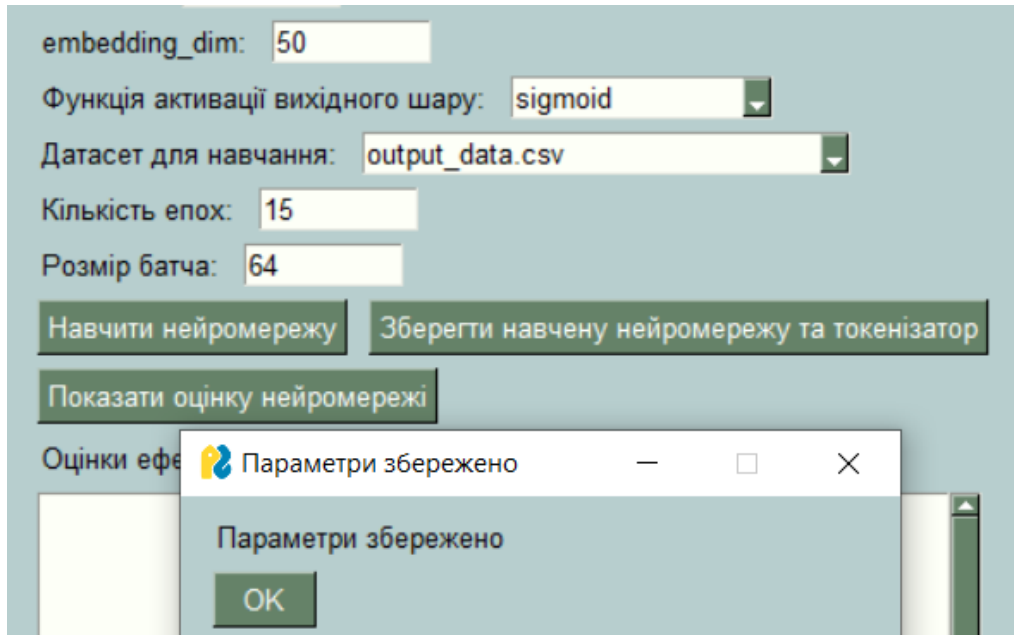


Рисунок 3.13 – Збереження нейромережі та токенизатора інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж

Для оцінювання ефективності навченої моделі необхідно натиснути кнопку «Показати оцінку нейромережі» (рисунок 3.14). Після чого користувач побачить оцінку навченої моделі за метриками.

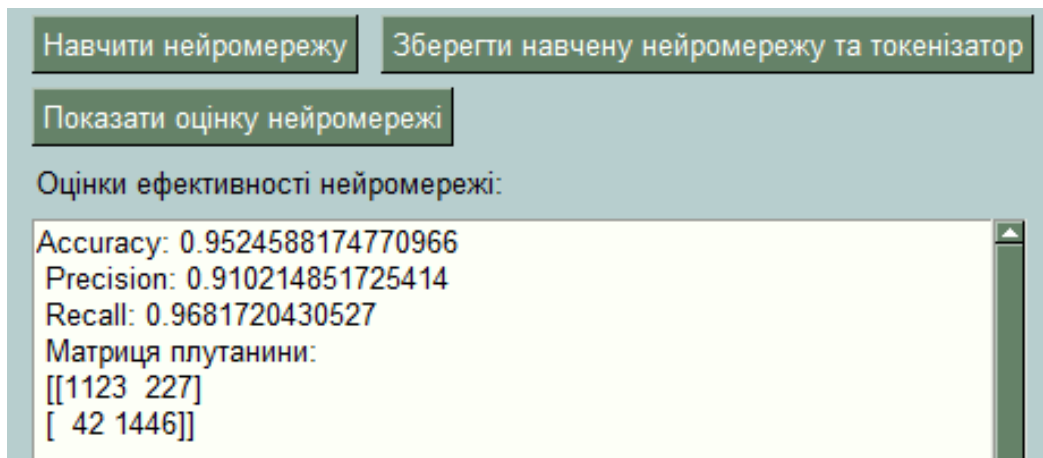


Рисунок 3.14 – Оцінка нейромережі метриками

Для переходу на підсистему виявлення стресового стану необхідно у головному меню натиснути кнопку «Підсистема виявлення стресового стану за дописами». Головний екран підсистеми наведено на рисунку 3.15.

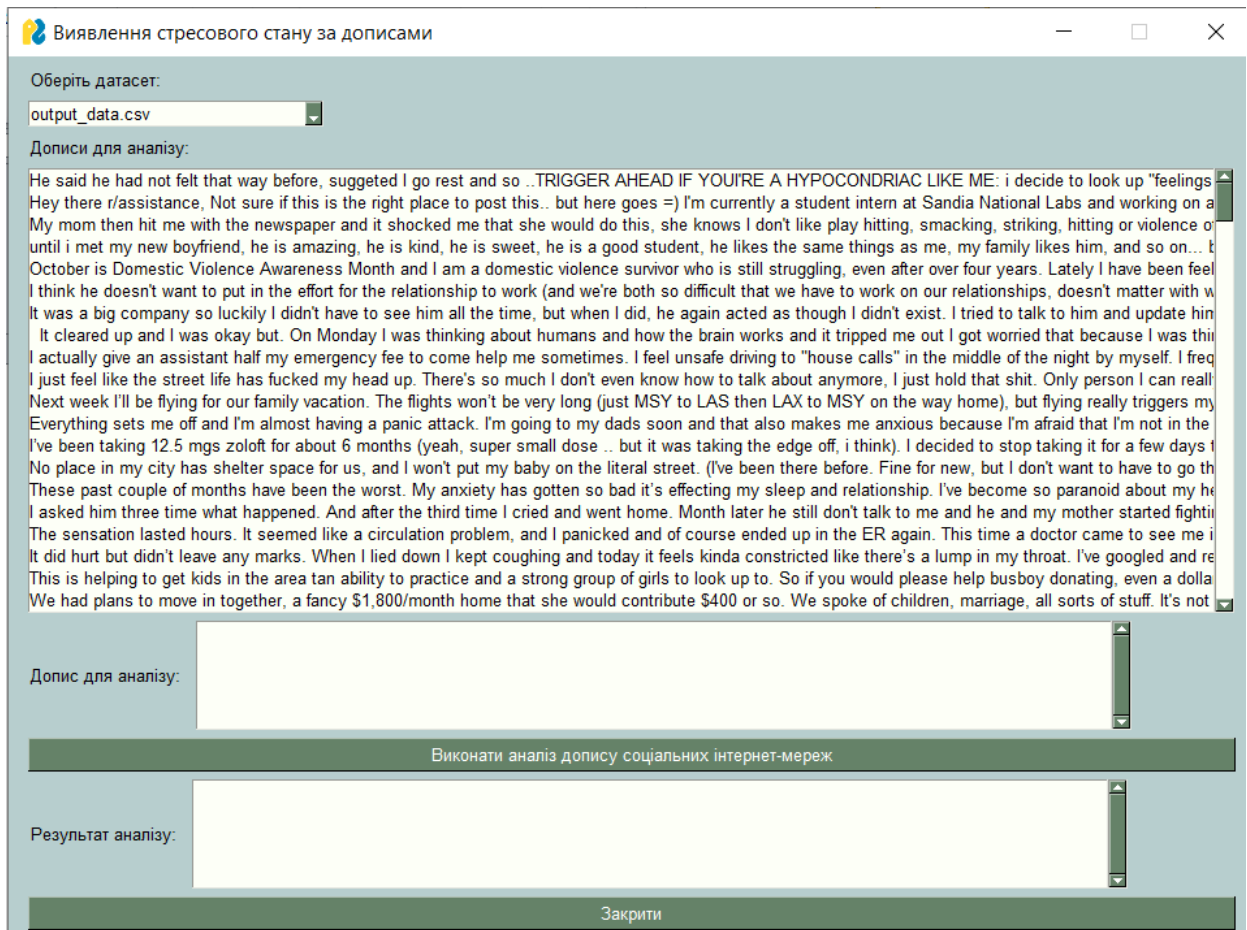


Рисунок 3.15 – Підсистема виявлення стресового стану за дописом

Можна виконувати як аналіз обраного з переліку допису, так і набрати допис «вручну» у текстовому полі «Допис для аналізу». Якщо обрати допис із переліку, його текст буде відображено у полі «Допис для аналізу» (рисунок 3.16).

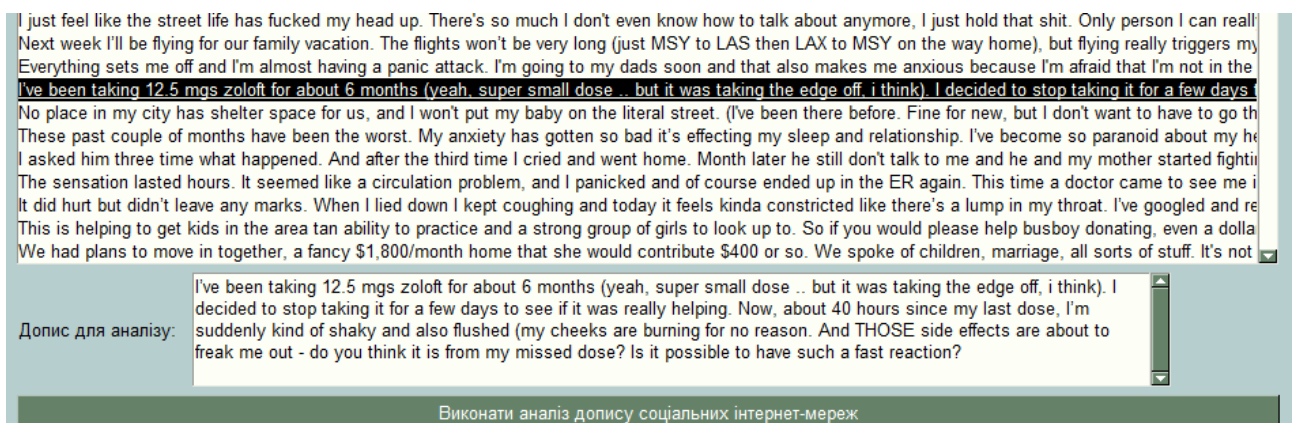


Рисунок 3.16 – Вибір допису для аналізу

Для визначення наявності стресового стану у дописі потрібно натиснути на кнопку «Виконати аналіз допису соціальних інтернет-мереж» (рисунок 3.17).

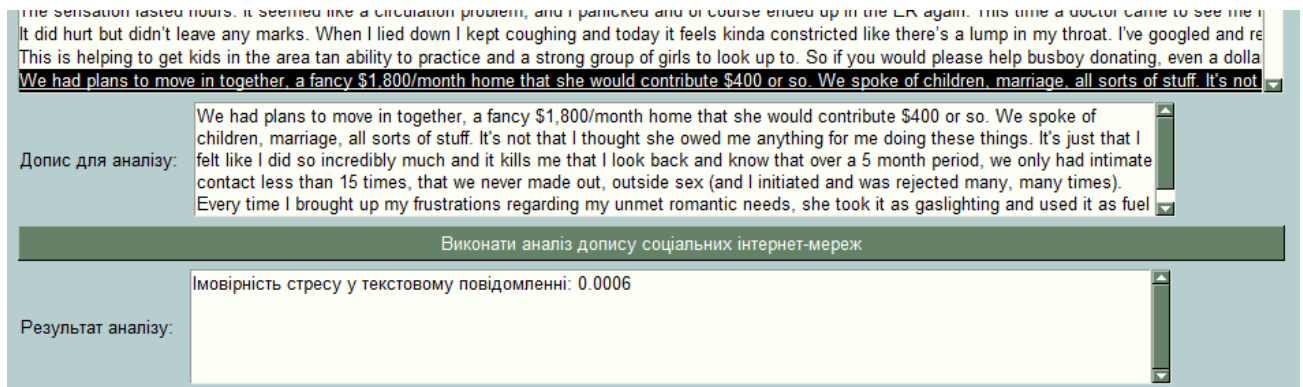


Рисунок 3.17 – Приклад виявлення стресу у повідомленні

Для використання підсистеми візуальної аналітики у головному меню необхідно натиснути кнопку «Підсистема візуальної аналітики» (рисунок 3.18).

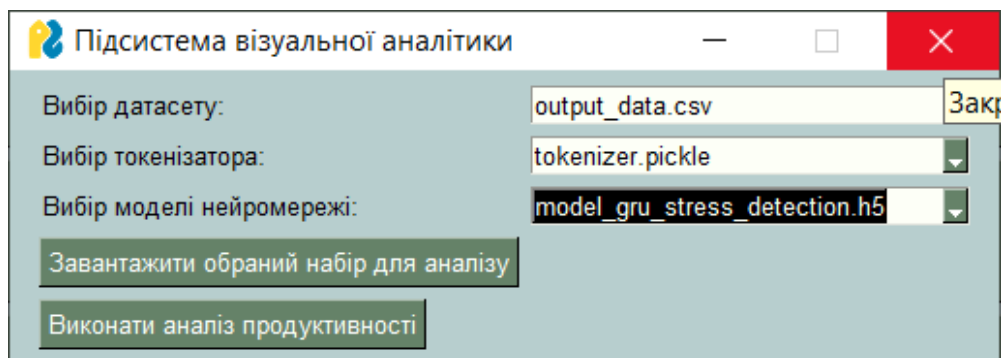


Рисунок 3.18 – Підсистема візуальної аналітики

Для того щоб скористатись нею спершу необхідно із випадючого списку обрати розмічений датасет для тесту, а також токенизатор та неймережеву модель, та натиснути кнопку «Завантажити обраний набір даних». Для виконання аналізу продуктивності необхідно натиснути кнопку «Виконати аналіз продуктивності», після чого по чергово будуть виводитись діаграми (рисунок 3.19).

Figure 1

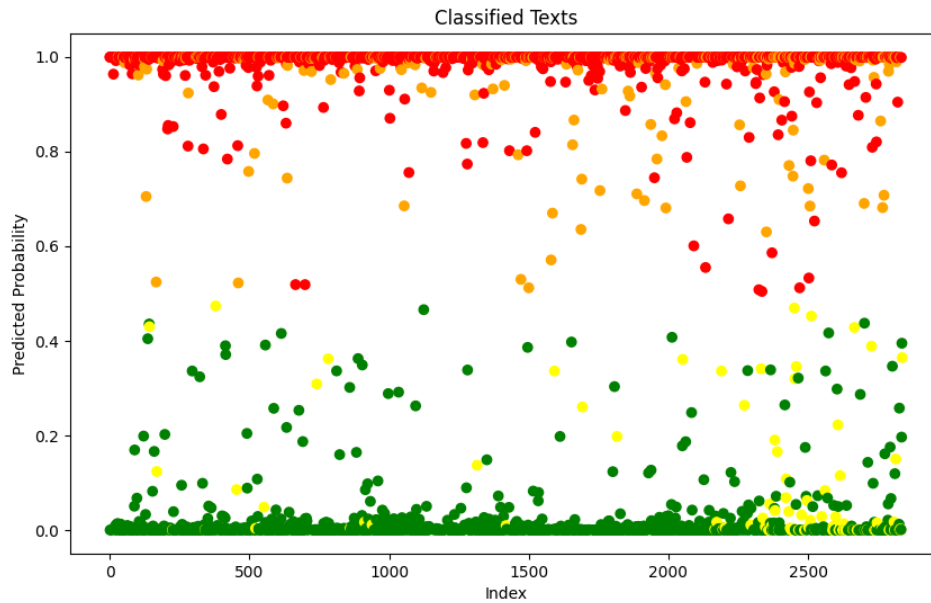
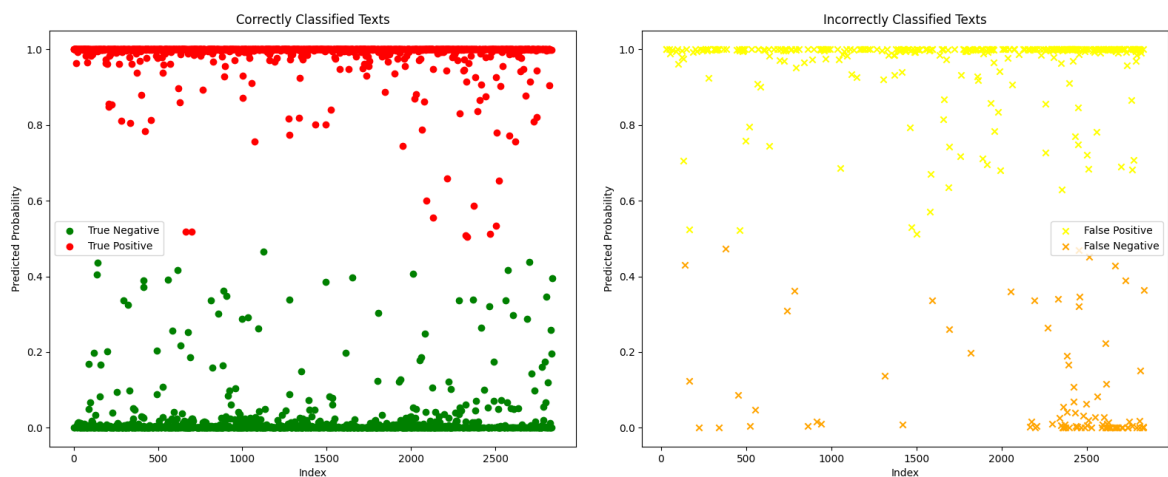


Figure 1



Риснуок 3.19 – Приклади діаграм візуальної аналітики

Отже, таким чином було описано функціональні можливості розробленої інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами.

3.7 Результати досліджень

Оскільки запропонований метод використовує нейромережу, як вже було вище зазначено, є потреба дослідити вплив параметрів навчання нейромережі, таких як кількості епох навчання, розмір батча, кількість шарів GRU в

архітектурі, розмірність словника. Оцінюватись нейромережеві моделі будуть з використанням метрик: Accuracy, Precision, Recall та матриця плутанини.

У якості валідаційного набору буде використано англomовних розмічених набір із 1000 записів, де 500 з них містять ознаки стресу, та 500 таких, що не містять ознак стресу.

В таблиці 3.6 наведено експеримент щодо впливу розміру словника та максимальної довжини послідовності tokenів. Кількість епох 20, розмір батча 32.

Таблиця 3.6 – Вплив параметрів токенизатора

<i>Параметри:</i>			<i>Метрики:</i>		
<i>max_words</i>	<i>max_len</i>	<i>embedding_dim</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
500	50	150	0.83	0.79	0.84
1000	150	200	0.86	0.81	0.86
1500	200	200	0.87	0.82	0.88
2000	300	250	0.91	0.91	0.92
2500	500	250	0.92	0.92	0.93
3000	550	300	0.94	0.95	0.94
3500	600	350	0.93	0.94	0.93

Результати таблиці 3.6 для наочності проілюстровані на графіку (рисунок 3.20).

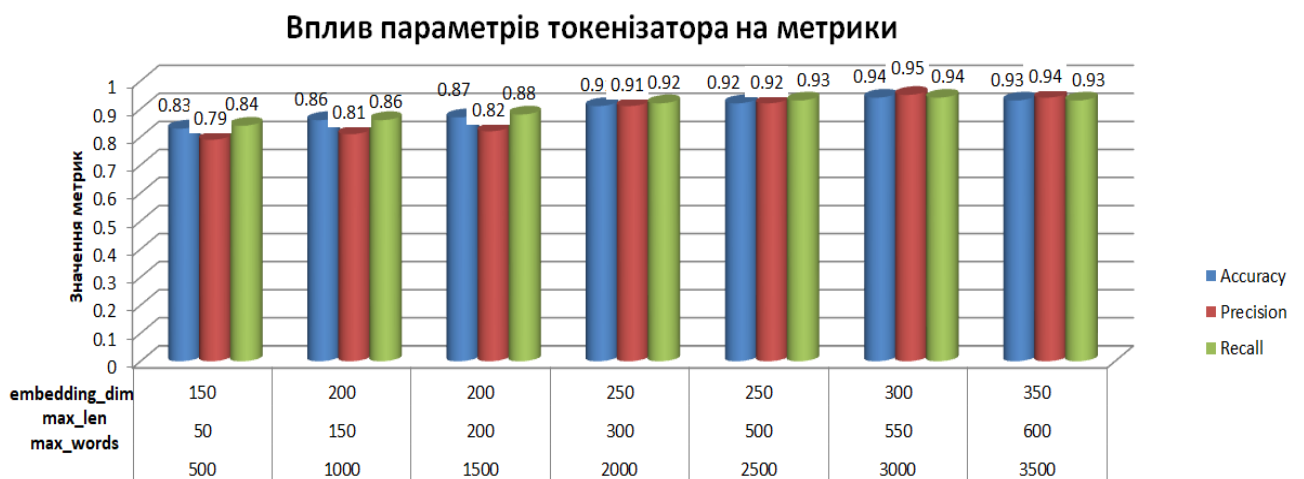


Рисунок 3.20 – Вплив параметрів токенизатора

Як видно з результатів таблиці 3.6, та рисунка 3.20, параметрами токенизатора для подальших досліджень доцільно обрати $\text{max_words} = 3000$, $\text{max_len} = 550$, та $\text{embedding_dim} = 300$. При таких параметрах матриця сплутувань набуває такого вигляду, як в таблиці 3.7

Таблиця 3.7 – Матриця плутанини

	Містить прояви стресу	Не містить проявів стресу
Містить прояви стресу	475	25
Не містить проявів стресу	30	470

Після досліджень параметрів токенизатора, є потреба дослідити вплив параметрів навчання нейромережі, таких як кількості епох навчання, розмір батча, кількість шарів GRU. Таблиця експерименту із значеннями метрик наведена в таблиці 3.8.

Таблиця 3.8 – Вплив параметрів нейромережі

Параметри:			Метрики:		
<i>batch</i>	<i>epoch</i>	<i>GRU count</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
16	10	1	0.94	0.94	0.95
32	10	2	0.93	0.93	0.94
64	10	3	0.95	0.95	0.94
16	20	4	0.94	0.95	0.94
32	20	5	0.94	0.95	0.94
64	20	6	0.93	0.94	0.93

Дані також проілюстровано на рисунку 3.21. Як видно з таблиці 3.8 та рисунку 3.21, збільшення епох не дає приросту в значеннях метрик, аналогічно як і збільшення шарів GRU. Найкращим варіантом з досліджених для виявлення стресу нейромережева модель навчена при $\text{batch} = 64$, $\text{epoch} = 10$, $\text{GRU count} = 3$.



Рисунок 3.21 – Вплив параметрів нейромережі

Дана програмна система може бути використана для науковців та дослідників природної мови, а також у якості самоконтролю рівня стресу за текстовим описом.

Отже, було досліджено вплив параметрів навчання нейромережі, таких як кількості епох навчання, розмір батча, кількість шарів GRU в архітектурі, а також вплив параметрів токенизатора. Оцінювались нейромережеві моделі з використанням метрик Accuracy, Precision, Recall та матриці плутанини. З розглянутих варіантів вдалось досягти значень Accuracy=0.95, Precision=0.95, та Recall=0.94 при batch = 64, epoch = 10, GRU count = 3 та таких параметрах токенизатора: max_words =3000, max_len=550, та embedding_dim=300.

3.8 Висновки до розділу 3

Для дослідження ефективності запропонованого методу виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами було обрано шлях дослідження впливу параметрів навчання та токенизатора на якість нейромережевих моделей, що оцінюється метриками Accuracy, Precision, Recall та матриці плутанини. Також було встановлено необхідність проведення тестування створеного програмного забезпечення засобами тест-кейсів.

У якості засобів розробки для інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами буде використано мову програмування Python, яка має широкі можливості в галузі розробок програмного забезпечення пов'язаного зі штучним інтелектом та інтегрована середовище розробки PyCharm.

Наведено структуру у вигляді діаграми класів та описано функціональне призначення програмних складових інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами, згідно до якої здійснено програмну реалізацію.

Здійснено тестування інформаційної системи засобами тест-кейсів, яке показало, що весь заявлений функціонал працює коректно згідно до поставлених завдань.

Описано функціональні можливості розробленого застосунку для виявлення стресових проявів у текстових повідомленнях на базі нейромережевого підходу.

Проведено дослідження ефективності, де досліджено вплив параметрів навчання нейромережі, таких як кількості епох навчання, розмір батча, кількість шарів GRU в архітектурі, а також вплив параметрів токенизатора. Оцінювались нейромережеві моделі з використанням метрик Accuracy, Precision, Recall та матриці плутанини.

Загальні висновки

Кваліфікаційна робота бакалавра була присвячена спрощенню виявлення стресового стану за дописами з соціальних інтернет-мереж за рахунок аналізу дописів NLP-засобами, у межах чого було розроблено метод виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами, а також відповідну програмну реалізацію, що використовує створений метод.

Для досягнення мети були поставлені та виконані наступні задачі:

- дослідження предметної області виявлення стресового стану за дописами;
- в рамках дослідження предметної області виконано огляд теоретичних підходів щодо виявлення стресового стану за дописами, було прийнято рішення використовувати підхід глибокого навчання, оскільки така задача як виявлення стресу у повідомленні містить велику кількість складних ознак, а також приховані залежності;
- виконано аналіз наукових надбань та існуючих програмних рішень в області виявлення стресового стану за дописами;
- розроблено метод виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами, наведено його схему та описано основні етапи;
- на основі розробленого методу виконано проектування інформаційної структури системи, описано її функціональну та проектну складові;
- виконано підготовку навчальних даних для тренування класифікатора для виявлення стресового стану за дописами – використано модифікований набір даних «Human Stress Prediction», що складається із 2100 записів категорії «стрес» та 1900 записів, що належать категорії «не стрес»;
- здійснити вибір засобів розробки для інформаційної системи - використано мову програмування Python та інтегрована середовище розробки PyCharm;

- здійснити програмну реалізацію інформаційної системи для виявлення стресового стану;

- проведено тестування розробленої програмної реалізації, яке показало що весь заявлений функціонал працює коректно;

- здійснено дослідження ефективності розробленого методу з використанням розробленої програмної реалізації – досліджено вплив параметрів навчання нейромережі, таких як кількості епох навчання, розмір батча, кількість шарів GRU в архітектурі, а також вплив параметрів токенизатора.

Отримані значення дослідження ефективності свідчать про спроможність створеного методу вирішувати задачу виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами з точністю 0.95.

Результат виконання кваліфікаційної роботи бакалавра повністю відповідає поставленим завданням та меті. Основні наукові та практичні результати доповідалися в доповіді «Practical Implementation of Neural Network Method for Stress Features Detection by Social Internet Networks Posts» на II Міжнародній науковій і теоретичній конференції «Scientific Review of the Actual Events, Achievements and Problems» 31 травня 2024 року (Berlin, Federal Republic of Germany, International Center of Scientific Research), за темою кваліфікаційної роботи бакалавра автором виконано наукову публікацію [26].

Щодо шляхів вдосконалення запропонованого підходу – можна дослідити підхід на моделях-трансформерах, а також інші модифікації рекурентних нейронних мереж.

Перелік посилань

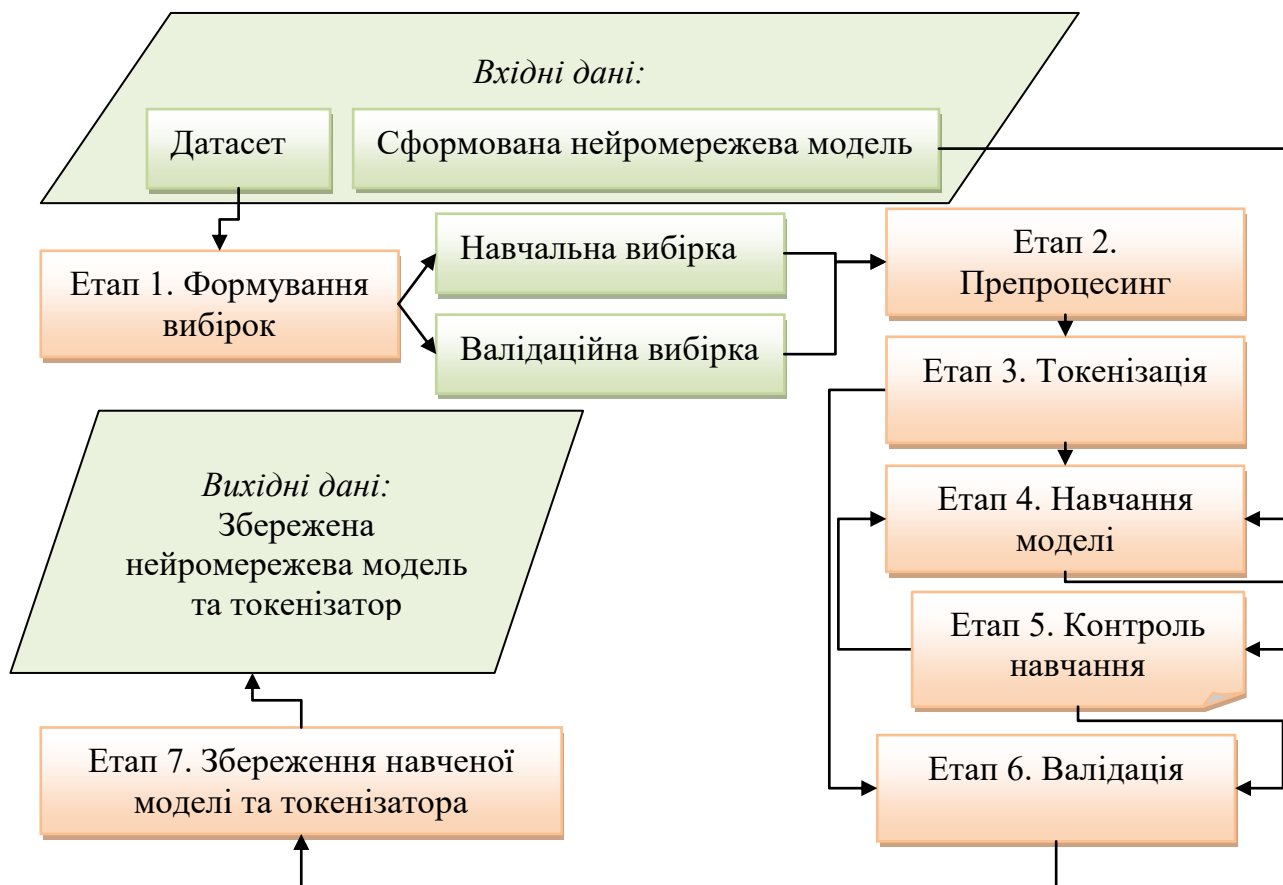
1. Що таке стрес та як він проявляється. URL: <https://marta.mva.gov.ua/articles/shcho-take-stres-ta-iak-vin-proiavlaietsia>
2. Стрес. URL: <https://www.ufmm.org.ua/psihichni-zahvoryuvannya/stres/>
3. Що таке стрес та як з ним боротися. URL: <https://phc.org.ua/news/scho-take-stres-ta-yak-z-nim-borotisya>
4. Що таке Стрес: види, ознаки, чим викликаний, як лікувати. URL: <https://vnk1.kiev.ua/disease/shho-take-stres-vidi-oznaki-chim-viklikanij-yak-likuvati/>.
5. Повна класифікація стресу: стадії та фази розвитку, види та різновиди. URL: <https://dln.com.ua/povna-klasifikaciya-stresu-stadi%D1%97-ta-fazi-rozvitku-vidi-ta-riznovidi>.
6. Словник NLP. URL: <https://medium.com/stinopys/%D1%81%D0%BB%D0%BE%D0%B2%D0%BD%D0%B8%D0%BA-nlp-b0fab1027551>
7. What is the main difference between RNN and CNN | NLP | RNN VS CNN. URL: <https://ashutoshtripathi.com/2021/07/12/the-main-difference-between-rnn-vs-cnn-nlp/>.
8. Application of Factor Graph Model in Stress Detection Using Social Network Messages. URL: <https://ebooks.iospress.nl/doi/10.3233/SSW230030>.
9. Stress detection using natural language processing and machine learning over social interactions. URL: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00575-6>
10. Arya V, Mishra AK. Machine learning approaches to mental stress detection: a review. Ann Optimization Theory Pract. 2021;31(4):55–67..
11. Aldarwish MM, Ahmad HF. Predicting Depression Levels Using Social Media Posts, 2017 IEEE 13th International Symposium on Autonomous Decentralized System (ISADS), Bangkok. 2017. pp. 277–280. <https://doi.org/10.1109/ISADS.2017.41>
12. Stress Detection using Machine Learning and Deep Learning. URL: <https://iopscience.iop.org/article/10.1088/1742-6596/1997/1/012019/meta>

13. Worry Watch provides CBT-based exercises and resources to help with anxiety. URL: <https://www.verywellmind.com/i-tried-worry-watch-review-7768309>.
14. Worry Watch. URL: <https://worrywatch.com/#worry-watch-app-mental-health-wellness>.
15. StressPal. URL: <https://stresspal.com/>.
16. StressPal Frontline. URL: <https://stresspal.com/frontline/>
17. Kaggle. Human Stress Prediction. URL: <https://www.kaggle.com/datasets/kreeshrajani/human-stress-prediction>
18. csv – Читання та запис файлу CSV. URL: <https://docs.python.org/uk/3/library/csv.html>
19. Wikipedia. TensorFlow. URL: <https://en.wikipedia.org/wiki/TensorFlow>
20. Wikipedia. Pandas (software). URL: [https://en.wikipedia.org/wiki/Pandas_\(software\)](https://en.wikipedia.org/wiki/Pandas_(software))
21. The sklearn.metrics module. URL: <https://medium.com/@bouimouass.o/the-sklearn-metrics-c568e0abcf03>
22. Wikipedia. Matplotlib. URL: <https://uk.wikipedia.org/wiki/Matplotlib>
23. Python – PySimpleGUI that you can do. URL: <https://medium.com/@darshanvasu3048/python-pysimplegui-that-you-can-do-9e94f8cdc607>
24. Мова програмування Python для машинного навчання та штучного інтелекту. URL: <https://goit.global/ua/articles/mova-prohramuvannia-python-dlia-mashynnoho-navchannia-ta-shtuchnoho-intelektu/>
25. Wikipedia. PyCharm. URL: <https://uk.wikipedia.org/wiki/PyCharm>
26. Mazurets O.V., Sobko O.V., Molchanova M.O., Zalutska O.O., Yurchak A.V. Practical Implementation of Neural Network Method for Stress Features Detection by Social Internet Networks Posts. Global Science: Prospects and Innovations. Proceedings of the II International Scientific and Theoretical Conference «Scientific Review of the Actual Events, Achievements and Problems». May 31, 2024. Berlin, Federal Republic of Germany: International Center of Scientific Research. 2024. Pp. 160-167.

ДОДАТКИ

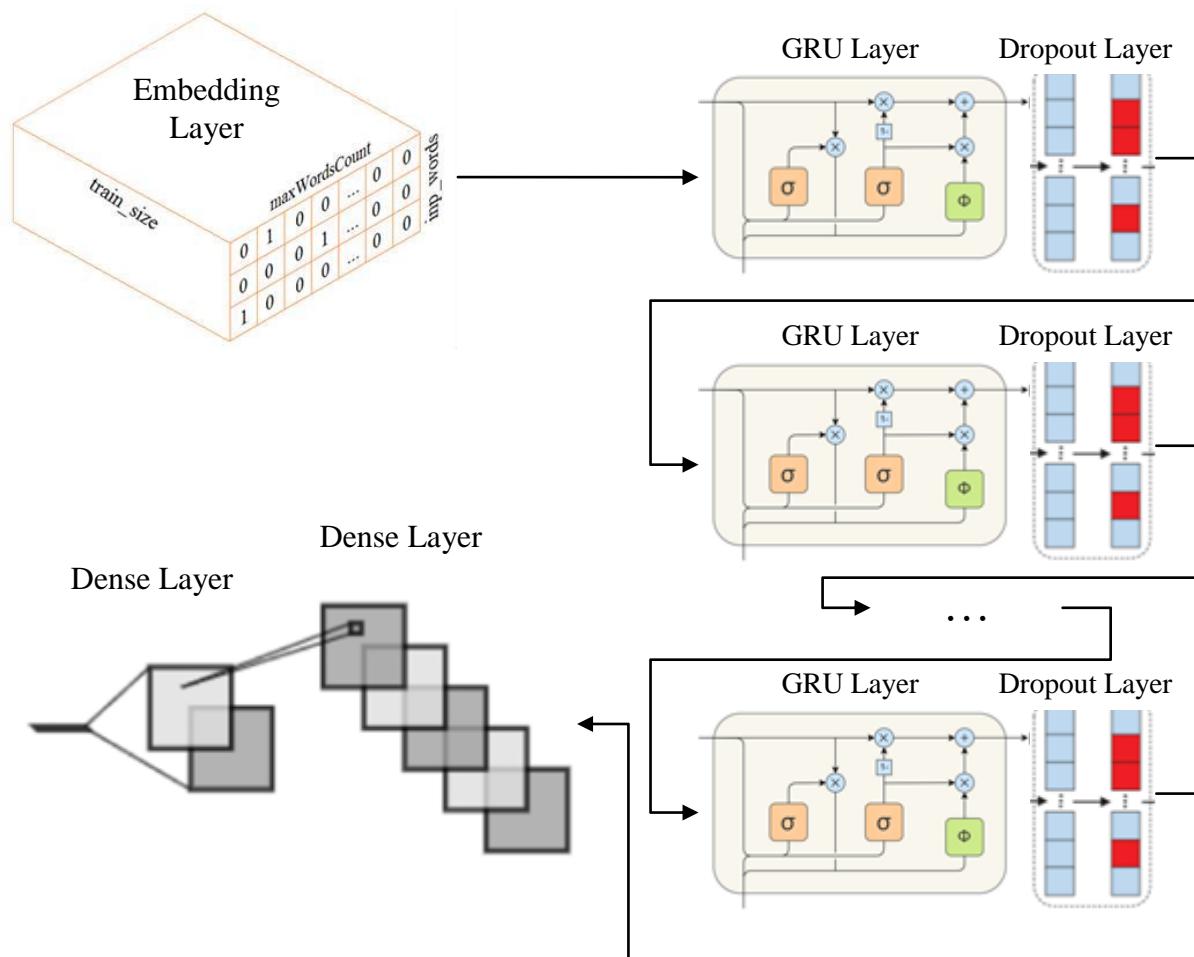
Додаток А

Схема та етапи формування навченої нейромережевої моделі та токенизатора



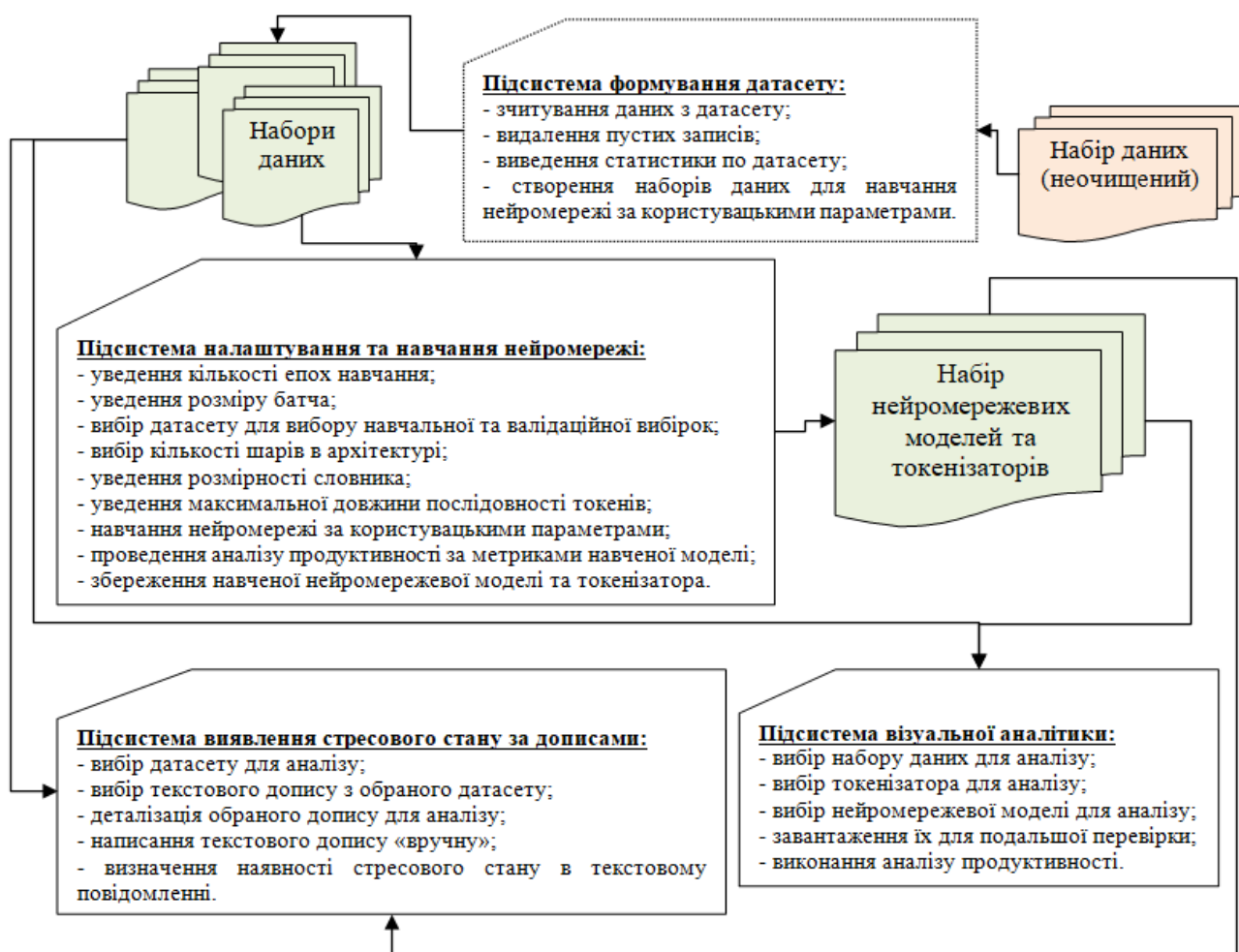
Додаток Б

Архітектура нейронної мережі для виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами



Додаток В

Проектна архітектура інформаційної системи виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами



Додаток Г

Презентаційний матеріал

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

МЕТОД ВИЯВЛЕННЯ СТРЕСОВОГО СТАНУ ЗА ДОПИСАМИ З СОЦІАЛЬНИХ ІНТЕРНЕТ-МЕРЕЖ NLP-ЗАСОБАМИ



Виконав:
студент групи КНс-21-1
Антон ЮРЧАК



Керівник:
викладач кафедри КН
Олена СОБКО

Актуальність

Люди проводять все більше часу в інтернеті, особливо в соціальних мережах. Це створює нові можливості для вивчення їхнього ментального стану за допомогою аналізу текстових даних, що може бути корисним для розуміння і покращення їхнього здоров'я.

У зв'язку з останніми подіями в Україні, можна стверджувати, що Інтернет поступово перетворюється на джерело загрози для інформаційної безпеки людей, суспільства та держави. Поширення сумнівного та необ'єктивного контенту разом із застосуванням технологій інформаційно-психологічного впливу на свідомість індивідів може сприяти збільшенню невдоволення діючою державною владою, національних конфліктів, соціальної напруженості та інших негативних явищ.

Стрес є серйозною проблемою сучасного суспільства, і його виявлення на ранній стадії може допомогти уникнути серйозних наслідків для здоров'я. Автоматизований аналіз текстових даних з соціальних мереж може допомогти виявити ознаки стресу, навіть якщо людина не виражає його явно.

Мета і задачі роботи

Метою кваліфікаційної роботи бакалавра є спрощення виявлення стресового стану за дописами з соціальних інтернет-мереж за рахунок аналізу дописів NLP-засобами.

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

- виконати дослідження предметної області виявлення стресового стану за дописами;
- в рамках дослідження предметної області виконати огляд теоретичних підходів щодо виявлення стресового стану за дописами;
- виконати аналіз наукових надбань та існуючих програмних рішень в області виявлення стресового стану за дописами;
- розробити метод виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами;
- на основі розробленого методу виконати проєктування інформаційної структури системи;
- виконати підготовку навчальних даних для тренування класифікатора для виявлення стресового стану за дописами;
- здійснити вибір засобів розробки для інформаційної системи;
- здійснити програмну реалізацію інформаційної системи для виявлення стресового стану;
- провести тестування розробленої програмної реалізації;
- здійснити дослідження ефективності розробленого методу з використанням розробленої програмної реалізації.

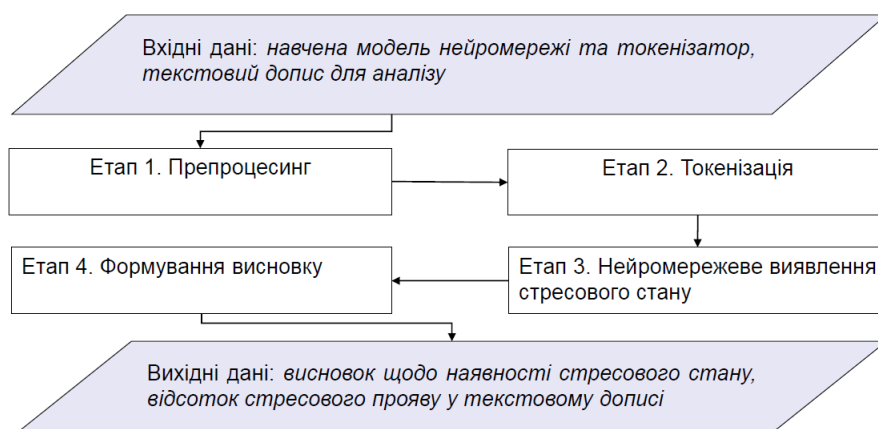
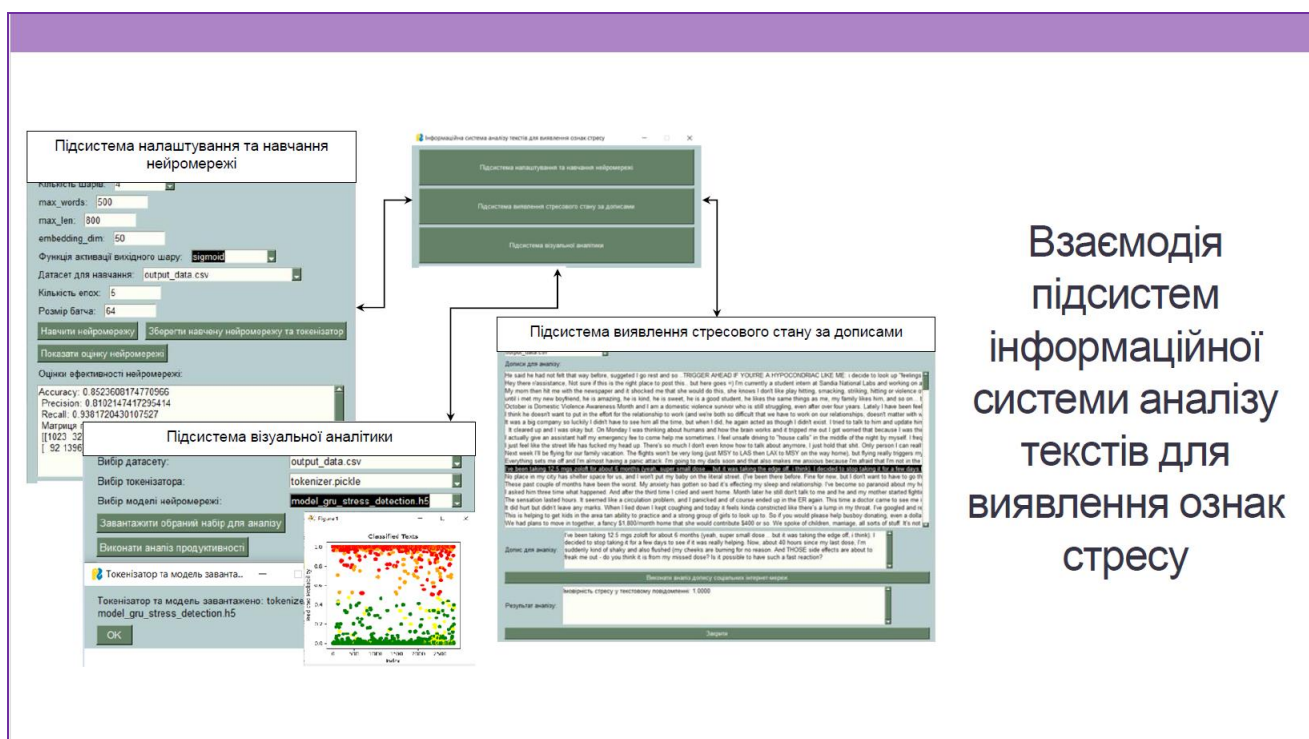
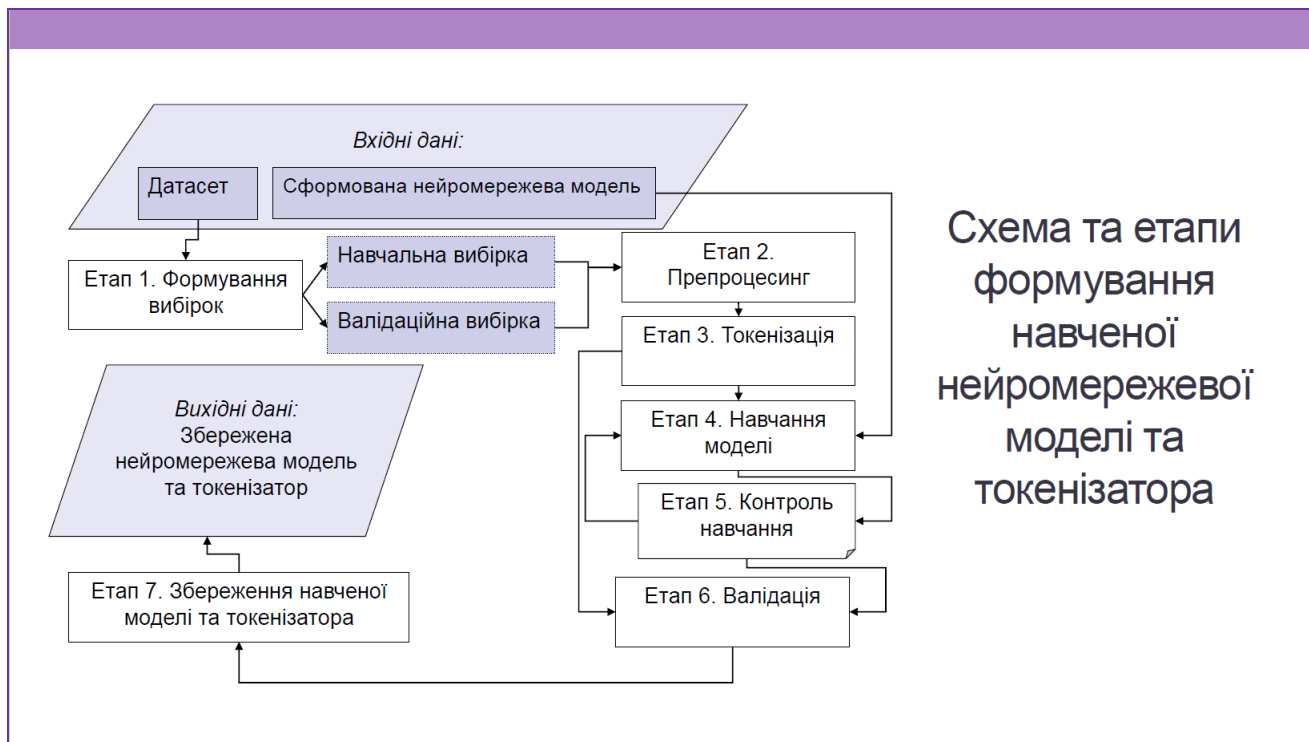
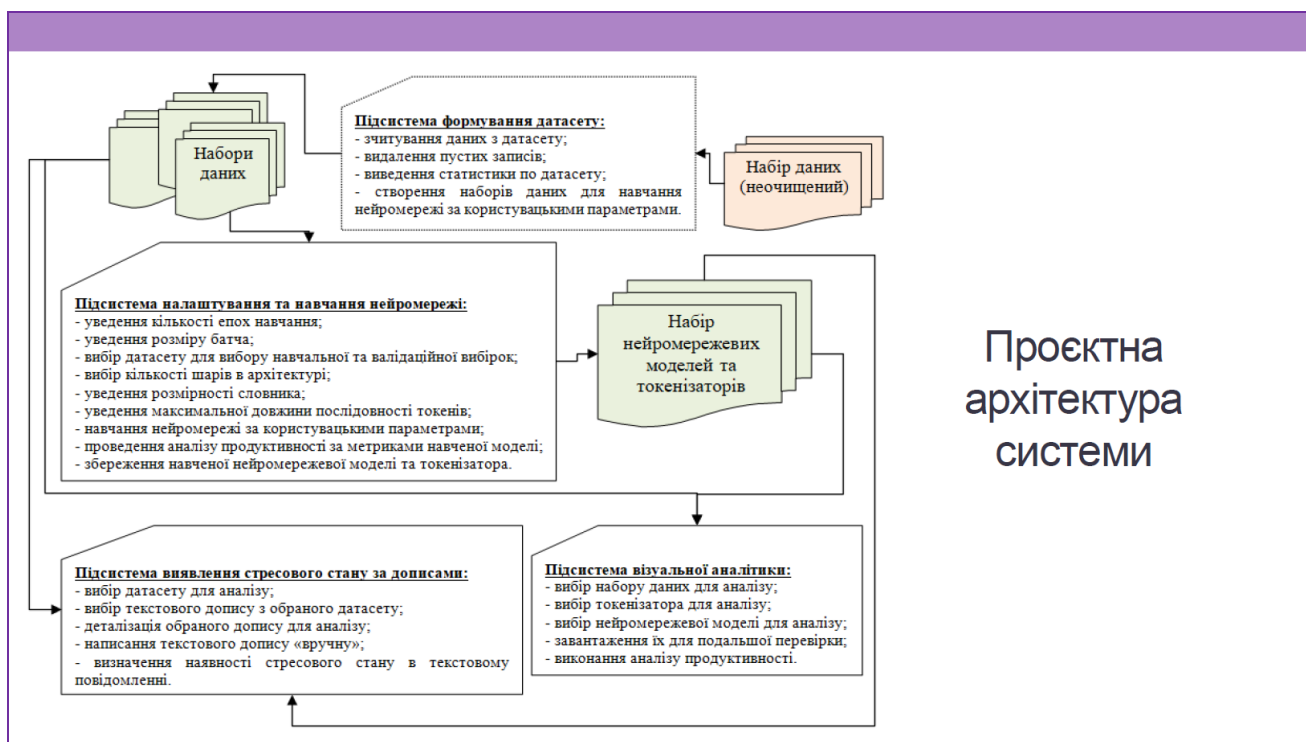
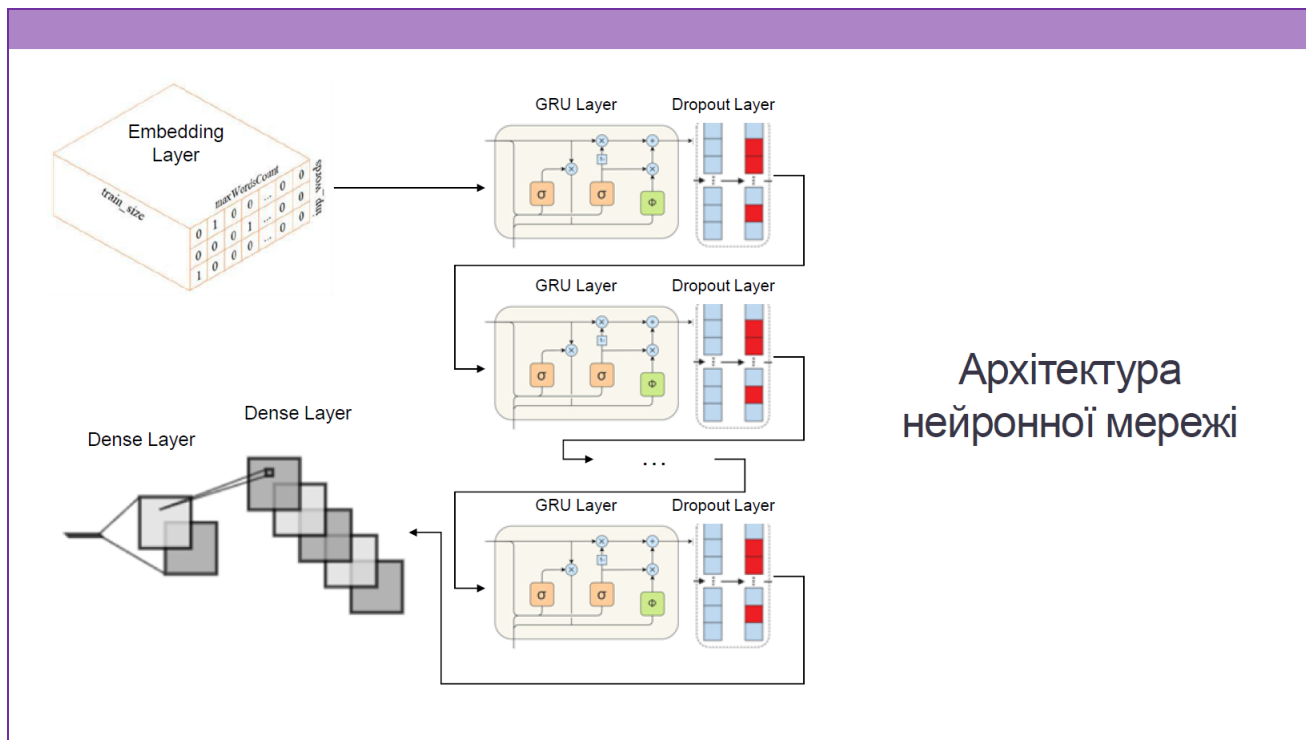


Схема та етапи методу виявлення стресового стану за дописами





Набір даних дослідження

Human Stress Prediction

Date Card Code (25) Discussion (1) Suggestions (0)

Stress.csv (1.42 MB)

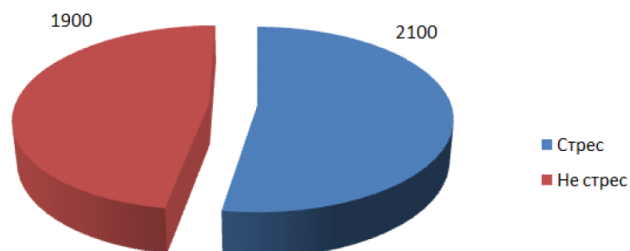
Detail Compact Column

7 of 7 columns

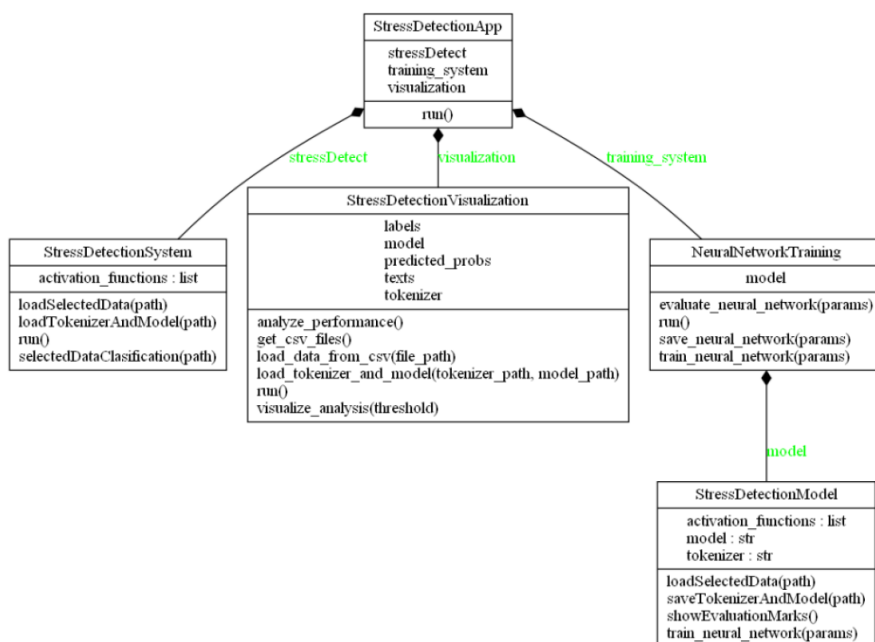
#	subreddit	#	post_id	#	sentence_range	#	text	#	label	#	confidence
	subreddit is a specific community or forum		unique_id		sentence index		text use for stress detection		0 and 1, 0 means no stress and 1 means stress		confidence on text
	ptsd	21%	2343 unique values		(0, 5]	12%			2820 unique values		
	relationships	19%			(0, 5]	12%					
	Other (1702)	60%			Other (2187)	76%					
	ptsd		8981tu		(15, 29)		He said he had not felt that way before, suggested I go rest and so ... TRIGGER AHEAD IF YOU'RE A HTPD...		1		8.8
	assistance		8lbrv9		(0, 5)		Hey there r/assistance, Not sure if this is the right place to post this... but here goes => I'm curr...		0		1

Вигляд датасету «Human Stress Prediction»

Кількість записів:



Розподіл даних за класами



Діаграма класів застосунку

Результати досліджень

	Містить прояви стресу	Не містить проявів стресу
Містить прояви стресу	475	25
Не містить проявів стресу	30	470

Матриця плутанини



Вплив параметрів нейромережі

Висновки

Було досягнуто мету кваліфікаційної роботи бакалавра – спрощення виявлення стресового стану за дописами з соціальних інтернет-мереж за рахунок аналізу дописів NLP-засобами.

Для досягнення мети були виконані наступні задачі:

- досліджено предметної області виявлення стресового стану за дописами;
- в рамках дослідження предметної області виконано огляд теоретичних підходів щодо виявлення стресового стану за дописами, було прийнято рішення використовувати підхід глибокого навчання, оскільки така задача як виявлення стресу у повідомленні містить велику кількість складних ознак, а також приховані залежності;
- виконано аналіз наукових надбань та існуючих програмних рішень в області виявлення стресового стану за дописами;
- розроблено метод виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами, наведено його схему та описано основні етапи;
- на основі розробленого методу виконано проектування інформаційної структури системи, описано її функціональну та проектну складові;
- виконано підготовку навчальних даних для тренування класифікатора для виявлення стресового стану за дописами
- здійснено вибір засобів розробки для інформаційної системи;
- виконано програмну реалізацію інформаційної системи для виявлення стресового стану та проведено тестування;
- здійснено дослідження ефективності розробленого методу з використанням розробленої програмної реалізації.

Ім'я користувача:
Кафедра КН

ID перевірки:
1016370658

Дата перевірки:
18.06.2024 08:11:26 EEST

Тип перевірки:
Doc vs Internet + Library

Дата звіту:
18.06.2024 11:15:28 EEST

ID користувача:
100005671

Назва документа: КНС-21-1 Юрчак_ЗАПИСКА

Кількість сторінок: 66 Кількість слів: 10618 Кількість символів: 85843 Розмір файлу: 2.76 MB ID файлу: 1016177779

Виявлено модифікації тексту (можуть впливати на відсоток схожості)

10.2% Схожість

Найбільша схожість: 3.98% з джерелом з Бібліотеки (ID файлу: 1016177778)

6.4% Джерела з Інтернету	401	Сторінка 68
7.4% Джерела з Бібліотеки	170	Сторінка 72

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.

Підозріле форматування 17 сторінок

Anti-Plagiarism v-15.257

Максимальне співпадіння з одним документом 4.0%

Словники перевірки: en_US, ru_RU, ua_UA. Помилки в документах: 11%

ID: 131270 Назва: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами Додано в БД: 2024-06-18 Автора: Антон ЮРЧАК Керівники: Олена СОБКО Консультанти: Опоненти:	Документ		Сумарний збіг по Базі Даних	
	Символи	Лексеми	Символи	Лексеми
	67029	975	4807 (7%)	70 (7%)

Джерело плагіату

ID	Опис	Наявність плагіату в документі	
		Символи	Лексеми

**РІШЕННЯ ЕКСПЕРТНОЇ 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 стресового стану за дописами з соціальних інтернет-мереж NLP-засобами

Автор: студент гр. КНс-21-1 Юрчак Антон Володимирович

Спеціальність: 122 – Комп'ютерні науки

Освітня програма: освітньо-професійна

Науковий керівник: викладач кафедри КН Олена Собко

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.	<i>відповідає</i>
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 Антона Юрчака, не є плагіатом, оскільки: запозичення розміщені в розділі огляду існуючих підходів, не описують безпосередньо авторську роботу і не стосуються її результатів; усі запозичення фрагментарні; серед запозичень знаходяться загальновідомі терміни, скорочення.

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

- за системою Anti-Plagiarism: 4%;

- за системою Unichек: 10,2.

Керівник роботи

О. Собко

Олена СОБКО

Гарант ОП

А. Мазурець

Олександр МАЗУРЕЦЬ

Завідувач кафедри КН

А. Бармак

Олександр БАРМАК



**ВІДГУК НАУКОВОГО КЕРІВНИКА
на кваліфікаційну роботу бакалавра**

студента гр. КНс-21-1 Антона Юрчака

за темою Метод виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами

1. Актуальність теми

Стрес є серйозною проблемою сучасного суспільства, яка негативно впливає на фізичне і психічне здоров'я людей. Виявлення стресу на ранніх стадіях може допомогти запобігти його важким наслідкам. Використання аналізу текстових даних з соціальних мереж для виявлення ознак стресу стає все більш актуальним, оскільки люди часто висловлюють свої емоції і думки в онлайн-просторі. Завдяки таким технологіям можна своєчасно ідентифікувати осіб, які можуть потребувати допомоги, навіть якщо вони не обговорюють свої проблеми явно, що дозволяє вчасно надати необхідну підтримку і запобігти подальшому погіршенню їхнього стану.

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

За стандартом, а саме описом предметної області, об'єктом дослідження є процес виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами. Метою роботи є спрощення виявлення стресового стану за дописами з соціальних інтернет-мереж за рахунок аналізу дописів NLP-засобами. При вирішенні поставленої задачі використано методи та технології машинного навчання для виявлення стресового стану. Отже, результати кваліфікаційної роботи бакалавра повністю відповідають стандарту бакалавра спеціальності 122 – Комп'ютерні науки.

3. Професійні та особистісні якості бакалавра

При виконанні кваліфікаційної роботи бакалавра Юрчак Антон Володимирович проявив себе як дисциплінований студент з високим рівнем самостійності. Також показав достатні вміння та навички в розробці програмного забезпечення заданої теми, що дозволило отримати відмінні результати.

4. Ступінь самостійності під час виконання кваліфікаційної роботи

Студент самостійно виконував усі завдання під час виконання кваліфікаційної роботи, тому результати роботи є особистим надбанням студента.

5. Ступінь оволодіння методами дослідження

Під час роботи над кваліфікаційною роботою Юрчак Антон Володимирович показав високий рівень володіння методами дослідження в області комп'ютерних наук.

6. Повнота та якість розкриття теми роботи

Автор кваліфікаційної роботи бакалавра повністю розкрив мету та завдання, здійснив аналіз сучасного стану предметної області та огляд існуючих рішень і підходів. Це дозволило розробити метод виявлення стресового стану за допомогою NLP-засобів на основі дописів із соціальних мереж та провести дослідження його ефективності.

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

Кваліфікаційна робота бакалавра відзначається чіткою структурою, що сприяє послідовному та логічному викладенню матеріалу з наведеним аргументуванням. Також слід відзначити високий рівень наукової грамотності автора.

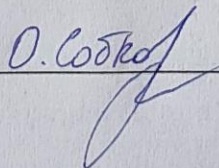
8. Можливість практичного застосування кваліфікаційної роботи бакалавра, окремих її частин

Розроблена інформаційна система призначена для використання психологами як додатковий інструмент для відстеження стресових станів пацієнтів, а також для самодіагностики користувачів соціальних мереж на наявність стресу. Вона має практичне застосування для автоматизованого виявлення стресового стану за текстовими повідомленнями, що також може бути корисним для науковців при автоматизованій розмітці великих обсягів текстових даних.

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

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

Керівник _____



викладач кафедри КН Олена Собко



РЕЦЕНЗІЯ

на кваліфікаційну роботу бакалавра

студента гр. КНс-21-1 Юрчака Антона Володимировича

за темою: Метод виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами

1. Актуальність обраної теми

Стрес є серйозною проблемою сучасного суспільства, і його виявлення на ранній стадії може допомогти уникнути серйозних наслідків для здоров'я. Аналіз текстових даних з соціальних мереж може допомогти виявити ознаки стресу, навіть якщо людина не обговорює його явно.

2. Повнота розкриття мети та завдань роботи

Автор кваліфікаційної роботи бакалавра повністю розкрив мету та завдання. В роботі проаналізовано сучасний стан предметної області, виконано огляд існуючих рішень та підходів, що дало змогу розробити метод виявлення стресового стану за дописами з соціальних інтернет-мереж NLP-засобами, а також провести дослідження ефективності.

3. Зміст кожного розділу роботи

Усі розділи роботи мають актуальну інформацію, що стосуються теми. У першому розділі надано характеристику предметної області виявлення стресового стану за дописами, обрано до реалізації підхід глибокого навчання, оскільки задача виявлення стресу у повідомленні містить велику кількість складних ознак, а також приховані залежності. Поставлено мету та задачі роботи. У другому розділі описано проєктування інформаційної системи аналізу текстів для виявлення ознак стресу, створено відповідний метод. У третьому розділі проведено експериментальне дослідження методу виявлення стресового стану за дописами з соціальних інтернет-мереж.

4. Оцінка розробленої інформаційної системи, її практична цінність

Розроблена інформаційна система аналізу текстів для виявлення ознак стресу має високий потенціал застосування. Автоматизоване виявлення стресового стану може бути використане у психологів у якості додаткового інструменту відстеження стресових станів пацієнтів, а також для само діагностування користувачів соціальних мереж на наявність стресового стану.

5. Якість оформлення кваліфікаційної роботи бакалавра

Автор якісно оформив кваліфікаційну роботу бакалавра. Робота містить необхідні розділи, ілюстративні матеріали, такі як таблиці та графіки, що допомагають краще розуміти представлену інформацію. Використана література підтверджує обґрунтованість висновків, а послідовне викладення матеріалу забезпечує легке сприйняття тексту.

6. Недоліки кваліфікаційної роботи бакалавра

Суттєвих недоліків кваліфікаційна робота бакалавра немає. Було б доцільно додати ще ряд параметрів нейромережі для підсистеми навчання нейромережі з метою подальших досліджень. А для підсистеми виявлення стресового стану за дописом додати можливість досліджувати вміст текстових файлів. По тексту пояснювальної записки виявлено кілька пунктуаційних помилок. Втім наведене вище не впливає на загальну якість роботи та отримані результати.

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

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

Рецензент

Г.Т.Н., проф.

Лисенко С.М.