

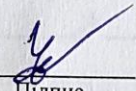
## КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

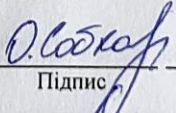
на тему Метод виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами

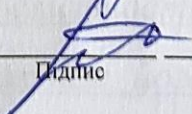
Галузь знань 12 – Інформаційні технології  
Шифр і назва галузі знань

Спеціальність 122 – Комп'ютерні науки  
Шифр і назва спеціальності

Освітня програма Комп'ютерні науки  
Назва освітньої програми

Виконав: студент групи КНс-21-1  Ілля ЧЕРВОНЧУК  
Курс, 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

21 червня 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 схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами»

2. Завдання видано студенту Іллі ЧЕРВОНЧУКУ

(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

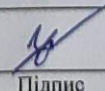
5. Дата видачі завдання студенту: « 16 » 02 2024 р.

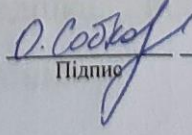
6. Зміст пояснювальної записки (перелік задач) та вихідні дані:

Мета роботи – покращення виявлення схильності до нетипової поведінки за текстовим дописом користувача за рахунок його аналізу нейромережевими засобами. Для досягнення мети слід вирішити такі задачі: виконати дослідження предметної області виявлення схильності до нетипової поведінки; створити метод виявлення схильності до нетипової поведінки за текстовим дописом; описати інформаційну структуру системи; спроектувати базу даних; створити відповідну програмну реалізацію на основі створеного методу; виконати тестування створеного програмного забезпечення та виконати дослідження ефективності створеного методу виявлення схильності до нетипової поведінки з використанням розробленого ПЗ.

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  Ілля ЧЕРВОНЧУК  
Курс, група виконавця Підпис Ім'я, ПРІЗВИЩЕ

Керівник: викладач каф. КН  Олена СОБКО  
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ

## Анотація

Тема кваліфікаційної роботи бакалавра: «Метод виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами».

Виконавець кваліфікаційної роботи бакалавра: студент групи КНс-21-1 Ілля ЧЕРВОНЧУК

Керівник кваліфікаційної роботи бакалавра: викладач кафедри КН Олена СОБКО

Кваліфікаційна робота бакалавра містить:

Пояснювальна записка				Кількість додатків
Сторінок	Рисунків	Таблиць	Джерел інформації	
66	23	12	29	4

Метою кваліфікаційної роботи бакалавра є покращення виявлення схильності до нетипової поведінки за текстовим дописом користувача за рахунок його аналізу нейромережевими засобами. Було розроблено метод виявлення схильності до нетипової поведінки за текстовим дописом нейромережевими засобами. Для розробки відповідної інформаційної системи було використано мову програмування Python, редактор програмного коду PyCharm, СКБД SQLite.

Розроблена система призначена для онлайн-сервісів з метою виявлення схильностей до нетипової поведінки за текстовими дописами. Напрямами практичного використання розробленої інформаційної системи визначено виконання автоматизованого аналізу текстових дописів користувачів з метою виявлення схильності до нетипової поведінки.

Ключові слова: нетипова поведінка, психічні розлади, рекурентні нейромережі, виявлення нетипової поведінки.

Виконавець: студент групи КНс-21-1

Курс, група виконавця

  
Підпис

Ілля ЧЕРВОНЧУК

Ім'я, ПРІЗВИЩЕ

## Зміст

Перелік скорочень .....	4
Вступ.....	5
Розділ 1 Характеристика предметної області виявлення схильності до нетипової поведінки за текстовим дописом .....	7
1.1 Аналіз інформаційних моделей щодо виявлення нетипової поведінки.....	7
1.2 Огляд теоретичних підходів щодо виявлення схильності до нетипової поведінки за текстовими дописами.....	13
1.3 Аналіз існуючих програмних рішень та наукових публікацій.....	15
1.4 Мета, задачі та вимоги до реалізації інформаційної системи .....	19
Розділ 2 Розробка методу виявлення схильності до нетипової поведінки за текстовим дописом.....	20
2.1 Схема методу виявлення схильності до нетипової поведінки за текстовим дописом нейромережевими засобами .....	20
2.2 Розробка архітектури нейронної мережі .....	22
2.3 Проектна архітектура інтелектуальної системи виявлення схильності до нетипової поведінки та взаємозв'язок компонентів.....	25
2.4 Проектування бази даних програмної інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом .....	27
2.5 Підготовка робочих вхідних даних для системи .....	32
2.6 Особливості використання спеціалізованих програмних компонентів для рекурентних нейромереж .....	35
2.7 Висновки до розділу 2 .....	37
Розділ 3 Програмна реалізація інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом .....	39
3.1 Визначення шляхів дослідження та засобів створення інтелектуальної системи виявлення схильності до нетипової поведінки .....	39
3.2 Вибір засобів розробки інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача.....	40

3.3 Структура та функціональне призначення програмних складових інтелектуальної системи виявлення схильності до нетипової поведінки.....	42
3.4 Особливості реалізації програмних складових інтелектуальної системи виявлення схильності до нетипової поведінки .....	45
3.5 Тестування інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом та вимоги до розгортання.....	51
3.6 Результати досліджень .....	54
3.7 Висновки до розділу 3 .....	58
Висновки .....	61
Перелік посилань.....	64
Додатки	

## Перелік скорочень

Скорочення, термін, позначення	Пояснення
ПЗ	Програмне забезпечення
ПТСР	Посттравматичний стресовий розлад
ШІ	Штучний інтелект
NLP	Natural Language Processing
SVC	Support Vector Classification
RNN	Рекурентна нейронна мережа
SVM	Support Vector Machine
RF	Random Forest
NB	Naive Bayes
BiLSTM	Bidirectional LSTM
LSTM	Long short-term memory
SMOTE	Synthetic Minority Over-sampling Technique
БД	База даних
ІС	Інформаційна система
ІТ	Інформаційні технології
КРБ	Кваліфікаційна робота бакалавра
СКБД	Система керування базами даних

## Вступ

Кваліфікаційна робота бакалавра присвячена покращенню виявлення схильності до нетипової поведінки за текстовим дописом користувача за рахунок його аналізу нейромережевими засобами, а також відповідної програмної реалізації у вигляді інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами, на базі розробленого методу.

**Актуальність.** З ростом популярності соціальних мереж люди віддають більше особистої інформації в онлайн-середовищі. Це створює унікальну можливість для застосування нейромережових методів для аналізу текстового контенту і виявлення потенційно нетипової поведінки.

На ряду з цим зростає інтерес до психологічних аспектів індивідуальної поведінки та її впливу на суспільство. Виявлення нетипової поведінки може допомогти в ранньому виявленні можливих проблем, таких як депресія, самогубство, насильницька поведінка тощо.

Компанії, що надають онлайн-сервіси, все більше зацікавлені в застосуванні таких методів для забезпечення безпеки своїх користувачів, виявлення шахраїв, боротьби з кіберзлочинністю та забезпечення загальної благополучності у віртуальному просторі.

Нейромережові методи стають все більш потужними і ефективними в аналізі тексту та виявленні складних закономірностей. Це відкриває нові можливості для створення точних моделей для виявлення нетипової поведінки на основі аналізу текстового контенту.

**Об'єкт дослідження** – процес виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами.

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

**Мета кваліфікаційної роботи бакалавра** – покращення виявлення схильності до нетипової поведінки за текстовим дописом користувача за рахунок його аналізу нейромережевими засобами.

**Завдання кваліфікаційної роботи бакалавра** – провести дослідження предметної області для задачі виявлення схильності до нетипової поведінки; виконати огляд теоретичних підходів, а також обрати підхід для нейромережевого виявлення схильності до нетипової поведінки; розробити метод виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами; описати інформаційну структуру системи для виявлення схильності до нетипової поведінки за текстовим дописом користувача; обрати набір даних для навчання нейромережі; створити відповідну програмну реалізацію на основі створеного методу; виконати тестування інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами; виконати дослідження ефективності створеного методу виявлення схильності до нетипової поведінки з використанням розробленої інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами.

## **Розділ 1 Характеристика предметної області виявлення схильності до нетипової поведінки за текстовим дописом**

### **1.1 Аналіз інформаційних моделей щодо виявлення нетипової поведінки**

В сучасному світі відзначається значущий вплив стресу на життя людини. Фізіологічний аспект стресу включає активацію гормональних систем, таких як кортизол та адреналін, що може призводити до негативного впливу на серцево-судинну систему та збільшення ризику серцевих захворювань [1].

У психологічному плані стрес може призводити до розвитку психічних проблем, таких як депресія, тривожність та порушення сну, що впливає на загальну якість життя та здатність справлятися з щоденними завданнями. Соціальний вимір стресу проявляється у взаємовідносинах, спричиняючи конфлікти в родині та на роботі, а також може сприяти відчуттю соціальної ізоляції та погіршенню міжособистісних стосунків. Наукові дослідження також підтверджують, що тривалий стрес може збільшувати схильність до хвороб, викликаючи імунологічні порушення та запальні процеси в організмі [2].

Стрес – це реакція організму на внутрішні або зовнішні негативні впливи, які перевищують його здатність адаптації. Пошук точних причин стресу є складною задачею, оскільки вони можуть бути індивідуальними та залежати від конкретної ситуації. Стрес може впливати на поведінку людини, включаючи можливість виникнення нетипової поведінки. Стрес є реакцією організму на зовнішні або внутрішні небезпеки або труднощі, і впливає на різні аспекти фізіології та психіки [3].

Одна з реакцій на стрес – це зміна поведінки в намаганні подолати або вирішити ситуацію. Наприклад, люди можуть виявляти нетипові звички, звертатися до незвичайних стратегій управління стресом або проявляти несподівані реакції на зовнішні подразники.

Індивідуальна реакція на стрес є великою мірою особистісною і може виявлятися в різних формах. Деякі люди можуть виявляти нетипову активність

або відмовлятися від звичних звичок, тоді як інші можуть приховувати емоції або реагувати агресивно [4].

Нетипова поведінка людини – це те, що відхиляється від загально прийнятих норм, стандартів чи очікувань, і може викликати захоплення, подив чи непокій в соціальному контексті. Це може бути пов'язано з різними сферами життя, такими як соціальна взаємодія, робоча діяльність, особисті стосунки чи взаємодія з оточуючим середовищем [5].

Нетипова поведінка може виявлятися в різних формах, таких як нестандартний стиль одягу, незвичайний спосіб висловлювання, рішення непопулярних або незвичайних проблем, або вираження нетрадиційних переконань та цінностей. Зазвичай, оцінка, що є "нетиповою", може бути суб'єктивною і залежати від конкретного культурного, соціального та індивідуального контексту.

Нетипова поведінка може бути джерелом творчості та новаторства, але також може викликати неприйняття від оточуючих. У загальному розумінні, це поняття не завжди має негативне чи позитивне значення, а залежить від конкретного контексту та інтерпретації.

Окрім стресу нетипова поведінка людини може бути обумовлена різноманітними факторами, і часто це є комплексною взаємодією різних елементів. Наприклад, мають вплив такі фактори [6]:

- культурні та соціальні впливи часто визначають, що вважається «типовою» або «нетиповою» поведінкою в конкретному суспільстві;
- різноманітні психічні розлади, такі як депресія, тривожність, шизофренія чи розлади особистості;
- хвороби, психофізіологічні аномалії чи вроджені властивості можуть впливати на нормальну функцію мозку і вести до нетипової поведінки;
- травматичні випадки чи негативні досвіди можуть викликати реакції, що виявляються у нетиповій поведінці;
- деякі властивості нетипової поведінки можуть мати генетичні причини або бути обумовлені біологічними аспектами функціонування мозку;

– вживання наркотиків чи алкоголю може впливати на мозкову активність та призводити до нетипової поведінки.

Соціокультурний контекст визначає стандарти та очікування, які суспільство накладає на людей, визначаючи, що вважається "типовим" чи "нетиповим" в поведінці. Наприклад, різні культури мають свої унікальні стандарти поведінки – те, що може вважатися прийнятним в одному суспільстві, може бути відхиленням в іншому. Економічні обставини також можуть впливати на поведінку. Наприклад, у багатьох суспільствах може вважатися нетиповим, якщо людина з більш низьким економічним статусом виявляє споживацькі звички, які асоціюються з більш високим статусом. А в культурах індивідуалістичного характеру може бути підкреслено важливість самовираження і незалежності, тоді як у колективістичних культурах більший акцент може бути зроблений на соціальній згуртованості. Поведінка, яка виокремлює індивіда від групи, може вважатися нетиповою. Субкультури, які представляють собою групи людей з власними унікальними цінностями, нормами та стандартами, можуть визначати свої власні форми та норми поведінки, які можуть виявлятися нетиповими в ширшому соціокультурному контексті. Наприклад, панк-субкультура має агресивний стиль одягу, татуювання та пірсинг, виразний протест проти соціальних та політичних уявлень, антиавторитарна атмосфера. А гот-субкультура має темний стиль одягу, макіяж із важкою фокусом на експресивності, цікавість околицями життя та смерті, індивідуалізм та відчуття трагічної естетики [7].

Психічні розлади представляють собою різноманітні порушення психічного здоров'я, які можуть впливати на сприйняття, мислення, настрої і поведінку особи. Ці розлади можуть виникати як внаслідок генетичних факторів, так і через вплив зовнішніх чинників, таких як стрес, травми чи соціокультурні умови.

Люди, що мають психічні розлади, можуть виявляти нетипову поведінку в залежності від характеристик свого конкретного розладу. Особи, які страждають від депресії, можуть демонструвати нетипову поведінку, таку як

відсутність інтересу до раніше улюблених занять, суїцидальні думки, відсутність енергії та інші симптоми, які відрізняються від звичайної активності. У людей із шизофренією може виникати нетипова поведінка, така як агресія, роздратованість, розсіяність думок, або виражена соціальна ізоляція. Вони також можуть переживати галюцинації та боязні, що також може впливати на їхню поведінку. Посттравматичний стресовий розлад є психічним розладом, який може виникати після особливо страшного чи травматичного досвіду. Цей розлад впливає на різні аспекти психічного здоров'я та може викликати нетипову поведінку. Люди із ПТСР можуть реагувати на стресові події нетиповим чином, наприклад, включати агресивну поведінку, виражати неврівноважені емоції, використовувати унікальні стратегії управління стресом та ізолюватися від інших [8].

Травматичні випадки та негативні досвіди можуть значно впливати на поведінку людини, надаючи їй нетипові реакції на оточуючий світ. Ці досвіди можуть включати в себе різні види травм: фізичні травми, емоційні або психічні травми, травми від війни, сексуальні травми, втрату близьких, насильство чи інші події, що порушують психічний стан людини. Такі випадки можуть призводити до нетипової поведінки внаслідок різних механізмів захисту та адаптації. Наприклад, людина може виявляти агресивну поведінку внаслідок неконтрольованої злості або страху, роздратованість може виявлятися через конфлікти в колективі, а уникання соціальних ситуацій може бути реакцією на соціальну анклавність, викликану травматичними досвідами. Такі негативні досвіди можуть створювати глибокі сліди в психіці, викликаючи різні симптоми, включаючи тривожність, депресію, порушення сну, гіпервигнаність, ірраціональний страх або відчуття безпорадності. Люди можуть виявляти нетипову поведінку, намагаючись подолати ці емоційні стани або виражаючи їх у специфічний спосіб [9].

Соціальні мережі стали не лише майданчиком для спілкування та обміну інформацією, але і простором, де може виявлятися нетипова поведінка з багатьох причин. Люди можуть виражати себе та взаємодіяти з іншими у цьому

віртуальному світі інакше, ніж у реальному житті. У соціальних мережах користувачі діляться своїм життям, думками, емоціями та враженнями. Вони стали не тільки засобом спілкування, але й ареною для вираження своєї особистості та поглядів на світ, а також вираження своїх переживань щодо подій, які мали місце в реальному житті. Соціальні мережі також можуть впливати на самопочуття та самооцінку, особливо через порівняння з іншими користувачами.

Соціальні мережі є площиною для нетипової поведінки, адже анонімність і віддаленість дозволяють деяким людям виражати свої думки та вчинки без традиційних обмежень. Відсутність фізичного контакту та віддаленість від реального світу роблять людей менш відповідальними за свої вчинки [10].

Соціальні мережі стали своєрідним віртуальним простором, де люди охоче діляться своїм внутрішнім станом, проблемами та переживаннями з різних причин. Люди, особливо в сучасному віртуальному світі, відчувають потребу у взаємодії та спілкуванні, а соціальні мережі надають можливість побачити та відчути взаємодію з іншими, навіть якщо вони далеко – ділитися своїми переживаннями стає засобом підтримки та взаємодопомоги. За допомогою соціальних мереж люди можуть шукати визнання та підтримку від свого оточення. Поділ профілю особистого життя може бути спрямованим на здобуття підтримки, порад або навіть просто визнання важливості свого існування. Для багатьох людей ділитися своїми проблемами та переживаннями у віртуальному просторі схоже на процес психотерапії – отримання відгуків, порад або просто вірного слухача може мати позитивний вплив на психічне здоров'я [11].

Загалом про схильність до нетипової поведінки шляхом аналізу текстових дописів можуть говорити наступні ознаки [12]:

- Зміни в емоційному тоні. Помітні зміни в емоційному тоні чи стилі публікацій, які суперечать звичайному стилю користувача.
- Агресивний контент. Часті чи надмірно агресивні коментарі, особливо ті, що спрямовані на інших користувачів.

- Депресивний контент. Текст допису, що містить депресивні ознаки, на кшталт думок про самоушкодження.
- Постійні конфлікти. Користувачі постійно вступають в конфлікти чи спростовують думки інших в дописах.
- Публікації неперевіреної інформації. Поширення неперевіреної чи спекулятивної інформації.
- Надмірна самовіддача чи скромність. Виразна самовіддача чи, навпаки, надмірна скромність у публікаціях може викликати увагу і свідчити про нетипову поведінку.
- Загрози чи нецензурна лексика. Публікації, які містять загрози або нецензурну лексику.
- Публікації з негативним впливом на інших. Публікації включають в себе контент, який може масштабно впливати негативно на інших.

Задача виявлення схильності до нетипової поведінки за текстовим дописом користувача виявляється важливою та актуальною з декількох причин. По-перше, інтернет та соціальні мережі стали важливою складовою сучасного спілкування, де люди виражають свої думки, почуття та інтереси. В такому віртуальному просторі можуть виникати ситуації, коли користувачі проявляють нетипову поведінку, що може впливати на їхню безпеку та комфорт, а також створювати негативний вплив на оточуючих [13].

Автоматизоване знаходження ознак нетипової поведінки в текстових дописах є важливим завданням інформаційних технологій, що спрямоване на забезпечення безпеки та етичного спілкування в цифровому середовищі. Нетипова поведінка може включати в себе агресивні висловлювання, депресивні та суїцидальні думки, спроби провокації, розповсюдження неправдивої інформації чи образливі коментарі. Тому автоматизоване виявлення нетипової поведінки важливе для вчасного надання підтримки особам, які можуть переживати емоційний стрес чи інші труднощі. Забезпечення підтримки може допомогти зменшити ризик розвитку конфліктів та сприяти покращенню

емоційного стану учасників віртуального спілкування, тому це є актуальним напрямом інформаційних технологій.

## **1.2 Огляд теоретичних підходів щодо виявлення схильності до нетипової поведінки за текстовими дописами**

Виявлення схильності до нетипової поведінки за текстовим дописом користувача є задачею штучного інтелекту, а саме підгалуззю ІІІ – машинного навчання та обробки природної мови.

Машинне навчання – це підгалузь штучного інтелекту, в якій системи мають здатність "навчатися" за допомогою даних, статистики та методу проб і помилок для оптимізації процесів та інновацій на більш швидких темпах. Машинне навчання надає комп'ютерам можливість розвивати здібності до навчання, подібні до людських, що дозволяє їм розв'язувати найскладніші проблеми світу, від досліджень раку до зміни клімату [14].

Обробка природної мови (Natural Language Processing) – це галузь штучного інтелекту, яка фокусується на взаємодії між комп'ютерами та людьми за допомогою природної мови. Метою є програмування комп'ютерів для обробки та аналізу великих обсягів даних природної мови. NLP включає в себе здатність машин розуміти, тлумачити та створювати людську мову таким чином, щоб це було якісним і значущим способом. [15]

Такого роду задачі можна розв'язати використовують такі алгоритми, методи та моделі ІІІ, як SVC, K-means, Random Forest, Naive Bayes, RNN.

SVC (класифікація методом опорних векторів) – опорні вектори вказують на об'єкти даних, які мають важливе значення для побудови моделі класифікації. Метод опорних векторів (SVM або SVC в випадку класифікації) є алгоритмом, який використовує опорні вектори для визначення оптимальної гіперплощини, яка розділяє дані на класи в просторі ознак. SVC входить в широкий клас методів опорних векторів і є одним з їх варіантів, спеціально призначеним для задач класифікації [16].

Кластеризація K-means є технікою навчання без вчителя для класифікації непозначених даних за групуванням їх за ознаками, а не за наперед визначеними категоріями. Змінна K представляє кількість створених груп або категорій. Мета полягає в розбитті даних на K різних кластерів та визначенні місця центра мас для кожного кластера. Потім новій точці даних може бути призначено кластер (категорія) на основі найближчого центру мас [17].

Алгоритм Random Forest (RF, випадковий ліс) є алгоритмом машинного навчання з керованим навчанням, який є дуже популярним та використовується для задач класифікації та регресії в машинному навчанні. Відомо, що ліс складається з численних дерев, і чим більше дерев, тим він стійкіший. Так само, чим більше дерев у випадковому лісі, тим вища його точність та здатність вирішувати завдання. Випадковий ліс є класифікатором, який містить кілька дерев рішень на різних підмножинах наданого набору даних та бере їх середнє значення для покращення прогнозованої точності цього набору даних. Він ґрунтується на концепції ансамблевого навчання, яке є процесом комбінування кількох класифікаторів для вирішення складної проблеми та покращення продуктивності моделі [18].

Naive Bayes (наївний Баєс) належить до сімейства дуже простих ймовірнісних класифікаторів, що базуються на Теоремі Баєса. Класифікатор отримав назву "Наївний Баєс" через спрощення обчислень для кожного класу, щоб вони були обчислюваними. Фактично, модель наївного Баєса є класифікацією за умовною ймовірністю з використанням Теоремі Баєса. Умовна ймовірність визначає ймовірність виникнення події за умови виникнення іншої події. Умовну ймовірність можна використовувати для розрахунку спільної ймовірності, яка є ймовірністю двох чи більше одночасних подій. З іншого боку, спільну ймовірність також можна використовувати для розрахунку умовної ймовірності, але часто важко розраховувати спільну ймовірність [19].

Recurrent Neural Network (RNN, рекурентна нейронна мережа) – це тип штучної нейронної мережі (ANN), яка використовується в Siri від Apple і голосовому пошуку Google. RNN запам'ятовує минулі введення завдяки

внутрішній пам'яті, яка корисна для прогнозування цін на акції, створення тексту, транскрипції та машинного перекладу. У традиційній нейронній мережі входи та виходи не залежать один від одного, тоді як вихід у RNN залежить від попередніх елементів у послідовності. Рекурентні мережі також спільно використовують параметри на кожному рівні мережі. У мережах прямого зв'язку для кожного вузла існують різні ваги. У той час як RNN має однакові ваги на кожному рівні мережі та під час градієнтного спуску, ваги та базис налаштовуються окремо, щоб зменшити втрати [20].

Отже, використовуючи різні алгоритми, методи та моделі машинного навчання та обробки природної мови, що призначені для роботи з текстовою інформацією та аналізу мови, можна реалізувати метод виявлення схильності до нетипової поведінки за текстовим дописом користувача. У рамках роботи буде використано нейромережевий підхід на основі рекурентних нейронних мереж, оскільки він дає вищу точність та можливість розуміння природної мови.

### **1.3 Аналіз існуючих програмних рішень та наукових публікацій**

Виявлення схильності до нетипової поведінки за текстовими дописами користувачів є актуальним завданням в різних сферах. У соціальних мережах і форумах такий метод може служити для попередження кібербулінгу та негативних явищ в онлайн-спільнотах. В сфері психічного здоров'я виявлення аномального способу вираження метод може допомагати попередженню кризових ситуацій та наданні підтримки через телемедицину. В освіті метод може служити для виявлення студентів, які потребують додаткової підтримки. У сфері корпоративного управління може використовуватися для виявлення стресу серед працівників. Крім того, в бізнесі виявлення нетипової поведінки може допомагати в розумінні потреб клієнтів та поліпшенні обслуговування. Цільове використання таких програм повинно дотримуватися етичних стандартів та забезпечувати конфіденційність особистих даних користувачів.

Одним із яскравих прикладів програмних застосунків є соціальна мережа Facebook, алгоритми якої дозволяють аналізувати контент соцмережі. Алгоритми Facebook спрямовані на аналіз та виявлення різноманітного контенту, наприклад, текстових та мультимедійних публікацій для виявлення порушень правил спільноти, неправдивої інформації, ксенофобії, ненависті та інших негативних аспектів. Facebook здійснює моніторинг та виявлення можливих загроз безпеці користувачів, таких як кіберзлочинці, шахраї та інші небезпечні дії, і також обмежує дії таких користувачів (рисунок 1.1).

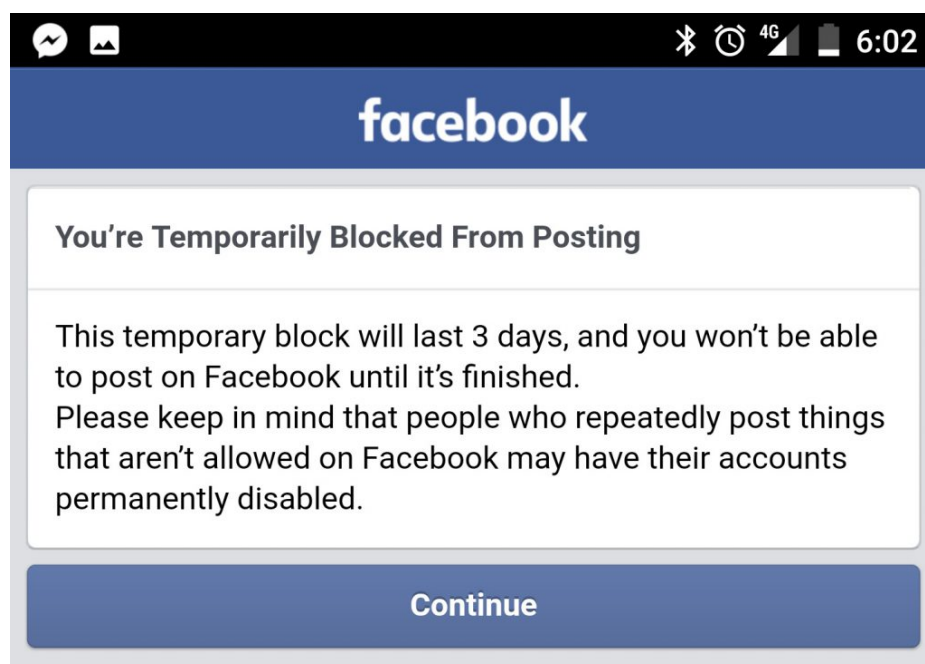


Рисунок 1.1 – Повідомлення про обмеження розміщення контенту користувачем [21]

Під час перегляду контенту, який може суперечити стандартам спільноти, контент, такий як публікації, коментарі та історії, виявляється та вилючається з Facebook. У випадку, якщо контент не порушує стандартів спільноти, але може бути вважаний неприпустимим, неуважним чи образливим, застосовуються різні заходи, такі як обмеження його розповсюдження або відзначення як чутливого. Стандарти спільноти визначають, що є припустимим чи неприпустимим на Facebook, й вони діють у всьому світі з метою підтримки вільного вираження і створення безпечного онлайн-середовища. З метою виявлення, рецензування та

прийняття заходів щодо контенту, який може порушувати стандарти спільноти, використовуються технології та людські рецензенти.

Розроблені моделі ШІ можуть визначити, чи потрібно вжити заходів щодо вмісту, таких як його видалення з платформи чи зменшення розповсюдження. Іноді контент потребує додаткового огляду, і наш ШІ відправляє його для розгляду людською робочою групою для уважнішого огляду. У таких випадках робочі групи роблять остаточне рішення, а розроблена технологія вчиться та покращується з кожним рішенням. З часом, вивчаючи тисячі людських рішень, технологія стає кращою [22].

Також активно проводяться наукові дослідження з виявлення схильності до депресивного стану, або ж агресією, які можуть бути причинами нетипової поведінки.

У роботі [23] проведено огляд використання методів обробки природної мови для виявлення психічних захворювань протягом останнього десятиріччя. Робота висвітлює важливість проблеми психічних захворювань у сучасному світі та їхній вплив на життя людей і суспільство в цілому. Використання методів NLP виявляє перспективи для покращення проактивної психічної допомоги та ранньої діагностики. Огляд включає 399 досліджень, проведених протягом десяти років. Він показує зростання інтересу до використання NLP для виявлення психічних захворювань, з особливою увагою до глибокого навчання, яке виявляється більш ефективним, ніж традиційні методи машинного навчання. Робота також висуває рекомендації для майбутніх досліджень, включаючи розробку нових методів виявлення, глибоких навчальних парадигм та моделей з високою ступенем інтерпретованості.

У роботі [24] досліджується проблема виявлення депресивних користувачів у віртуальних форумах. Дослідники аналізують поведінку користувачів у віртуальному форумі ReachOut.com, який надає підтримку молоді для обговорення повсякденних проблем, включаючи депресію. Запропоновано метод на основі навчання без вчителя з використанням рекурентних нейронних мереж та виявлення аномалій у тексті, що дає можливість виявлення

депресивних користувачів. Аналізується лінгвістичний стиль користувачів у поєднанні з нейромережевими характеристиками, які моделюють взаємодії користувачів у форумі. Результати показують, що як лінгвістичні особливості, так і нейромережеві характеристики є ефективними предикторами користувачів із депресією. Комбінування цих двох наборів характеристик дозволяє досягти значення F1-міри 0.64 та виявляється ефективнішим за базові методи.

У роботі [25] досліджується проблема емоційного переїдання як симптому розладів харчування, зокрема в онлайн-спільноті Reddit. Застосовуються методи машинного навчання та тематичного моделювання для аналізу поведінкових даних. Виділено чотири ключові теми емоційного переїдання та п'ять основних тем для зворотного зв'язку. Результати свідчать про можливість використання тематичних моделей та класифікаторів машинного навчання для розуміння поведінкових паттернів користувачів із симптомами емоційного переїдання. Це дослідження вказує на потенційне застосування в персоналізованій медицині для надання ефективної підтримки онлайн-користувачам із симптомами емоційного переїдання.

Сучасні дослідження, представлені в різних роботах, свідчать про те, що виявлення схильності до нетипової поведінки за текстовим дописом користувача є вкрай актуальною та важливою задачею. Це особливо стосується сфери психічного здоров'я, де визначення нетипових паттернів поведінки може слугувати підставою для ранньої діагностики та надання підтримки. Застосування нейромережевих засобів, таких як рекурентні нейронні мережі і методи глибокого навчання, виявляється дуже ефективним для аналізу текстових даних та виявлення нетипових сценаріїв поведінки. Це дозволяє автоматизувати процес виявлення та забезпечити швидше реагування на можливі ризики або схильності до певних поведінкових відхилень. У цілому, використання нейромережевих інструментів для виявлення нетипової поведінки в текстових дописах користувачів визнається перспективним напрямком для досліджень.

## 1.4 Мета, задачі та вимоги до реалізації інформаційної системи

Метою кваліфікаційної роботи бакалавра є покращення виявлення схильності до нетипової поведінки за текстовим дописом користувача за рахунок його аналізу нейромережевими засобами. Для покращення виявлення схильності до нетипової поведінки за текстовим дописом необхідно створити програмне забезпечення, а також слід вирішити такі завдання:

- виконати аналіз інформаційних моделей області виявлення схильності до нетипової поведінки;
- виконати огляд теоретичних підходів, а також обрати підхід для нейромережевого виявлення схильності до нетипової поведінки;
- виконати аналіз існуючих публікацій за напрямком дослідження;
- провести аналіз існуючого програмного забезпечення області виявлення схильності до нетипової поведінки за текстовим дописом користувача;
- створити метод виявлення схильності до нетипової поведінки за текстовим дописом користувача;
- описати інформаційну структуру системи для виявлення схильності до нетипової поведінки за текстовим дописом користувача;
- спроектувати базу даних;
- обрати набір даних для навчання нейромережі;
- створити програмну реалізацію на основі створеного методу;
- виконати тестування створеної інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами;
- виконати дослідження ефективності створеного методу виявлення схильності до нетипової поведінки з використанням розробленої інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами.

## Розділ 2 Розробка методу виявлення схильності до нетипової поведінки за текстовим дописом

### 2.1 Схеми методу виявлення схильності до нетипової поведінки за текстовим дописом нейромережевими засобами

Метод виявлення схильності до нетипової поведінки за текстовим дописом присвячений покращенню виявлення схильності до нетипової поведінки за текстовим дописом користувача за рахунок його аналізу нейромережевими засобами. Схеми методу наведена на рисунку 2.1.

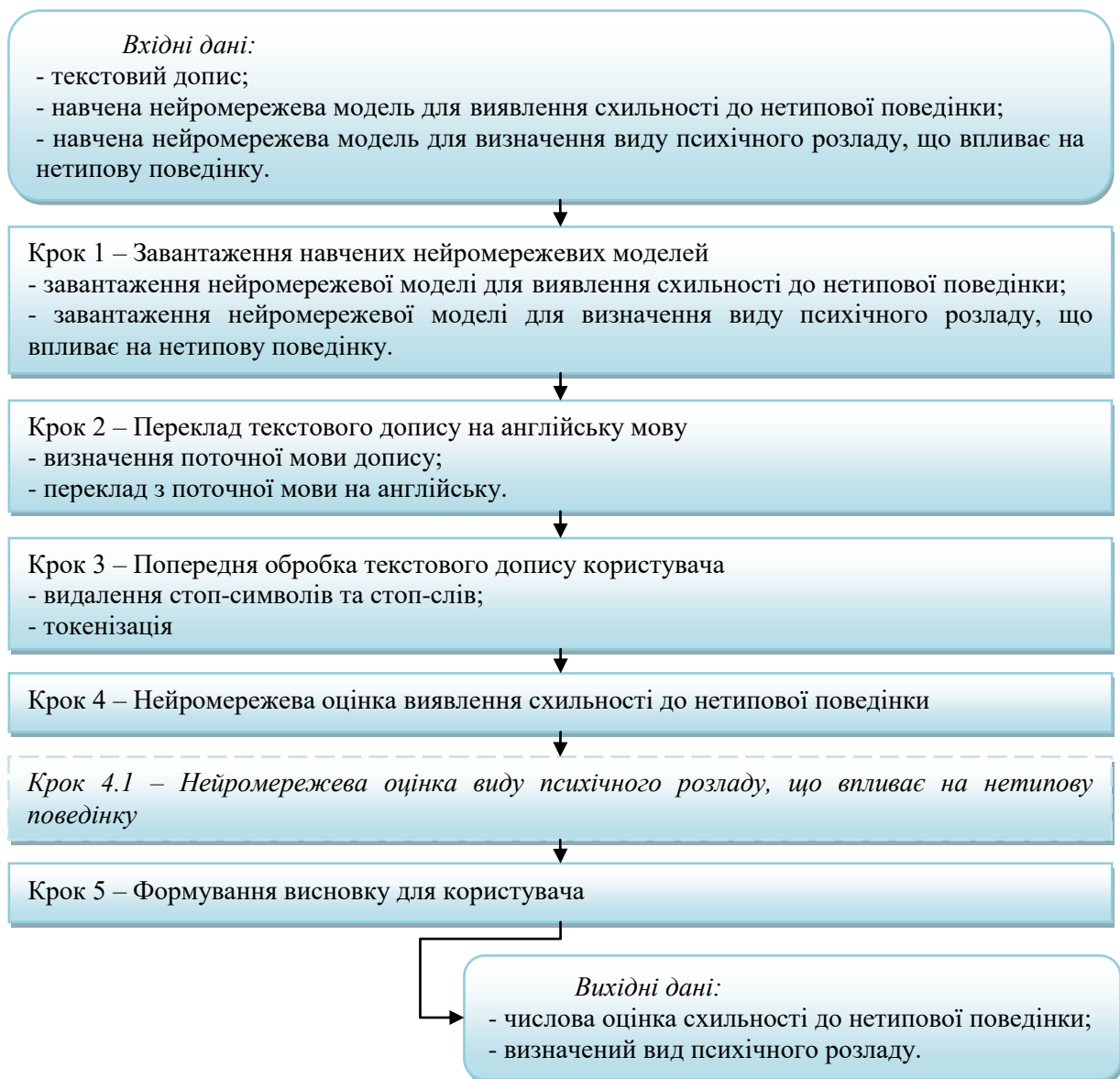


Рисунок 2.1 – Схеми методу виявлення схильності до нетипової поведінки

Вхідними даними методу виявлення схильності до нетипової поведінки за текстовим дописом є текстовий допис, навчена нейромережева модель для виявлення схильності до нетипової поведінки та навчена нейромережева модель для визначення виду психічного розладу, що впливає на нетипову поведінку.

Першим кроком є завантаження навчених нейромережевих моделей. Для роботи методу використовується дві натреновані нейромережеві моделі: для виявлення схильності до нетипової поведінки та для визначення виду психічного розладу, що впливає на нетипову поведінку.

Другим кроком є переклад текстового допису на англійську мову. Він відбувається у два етапи, першим з яких є визначення поточної мови допису, а потім і переклад з поточної визначеної мови на англійську.

Наступним кроком здійснюється попередня обробка текстового допису користувача, що включає видалення стоп-символів та стоп-слів, а також процес токенізації.

Четвертим кроком є нейромережева оцінка виявлення схильності до нетипової поведінки. Дана оцінка є бінарною. Якщо оцінка позитивна (є схильність до нетипової поведінки), відбувається крок 4.1, якщо оцінка негативна, крок 4.1 не відбувається.

На кроці 4.1 здійснюється нейромережева оцінка виду психічного розладу, що впливає на нетипову поведінку. В роботі представлено десять видів можливих розладів поведінки.

На п'ятому кроці відбувається формування висновку для користувача, що включає в себе оцінку виявлення схильності до нетипової поведінки, а також вид можливого психологічного розладу, що може призвести до нетипової поведінки.

Вихідними даними методу є числова оцінка схильності до нетипової поведінки та визначений вид психічного розладу.

Нетипова поведінка може включати в себе агресивні висловлювання, депресивні та суїцидальні думки, спроби провокації, розповсюдження неправдивої інформації чи образливі коментарі. Тому автоматизоване виявлення

нетипової поведінки важливе для вчасного надання підтримки особам, які можуть переживати емоційний стрес чи інші труднощі.

Отже, метод виявлення схильності до нетипової поведінки за текстовим дописом присвячений покращенню виявлення схильності до нетипової поведінки за текстовим дописом користувача за рахунок його аналізу нейромережевими засобами. Метод працює шляхом перетворення вхідних даних у вигляді текстового допису, навченої нейромережевої моделі для виявлення схильності до нетипової поведінки та навченої нейромережевої моделі для визначення виду психічного розладу, що впливає на нетипову поведінку у вихідні дані у вигляді числової оцінки схильності до нетипової поведінки та визначений вид психічного розладу.

## 2.2 Розробка архітектури нейронної мережі

Для роботи методу використовується дві рекурентні нейромережеві моделі: для виявлення схильності до нетипової поведінки та для визначення виду психічного розладу.

Модель для виявлення схильності до нетипової поведінки є типовою RNN-моделлю, з двома виходами. Архітектура нейромережевої моделі наведена на рисунку 2.2.

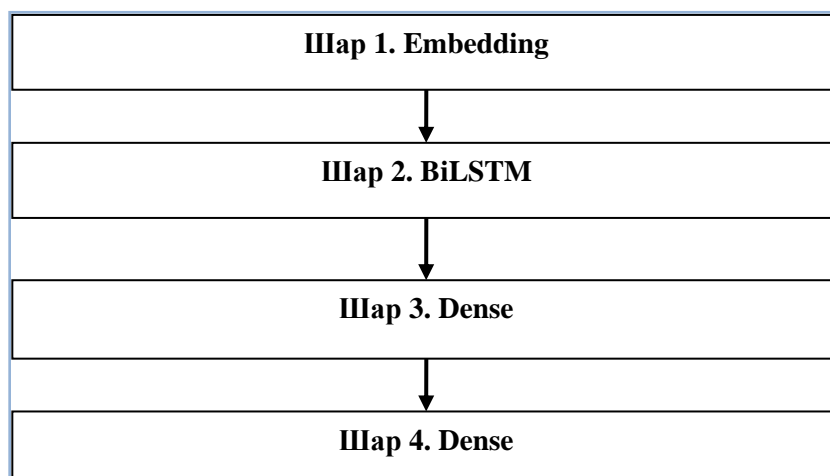


Рисунок 2.2 – Архітектура RNN-моделі для виявлення схильності до нетипової поведінки

Першим шаром є шар Embedding, що перетворює слова на вектори фіксованої довжини, що допомагає моделі зрозуміти семантичне значення слів та їхні взаємозв'язки в контексті.

Наступним шаром є BiLSTM, який представляє собою модифіковану версію звичайного LSTM-шару, що обробляє вхідні послідовності в обох напрямках за допомогою двох окремих LSTM шарів, кожен з яких містить по 128 нейронів.

У передостанньому шарі Dense кожен вхід з'єднується з кожним виходом, охоплюючи всі дані з попереднього шару. В рамках дослідження цей шар містить 128 нейронів. Це означає, що в шарі буде 128 нейронів з вагами, які будуть підлаштовуватись у процесі навчання.

Фінальний шар нейромережі, як і попередній, є шаром типу Dense. Оскільки він завершує модель, використовується функція активації softmax. Кількість нейронів у цьому шарі відповідає числу унікальних класів у навчальному наборі даних, що в даному дослідженні становить 2 (є схильність до нетипової поведінки / немає схильності до нетипової поведінки). Функція softmax генерує вектор ймовірностей для кожного класу, де кожне значення вказує на ймовірність того, що вхідний зразок належить до певного класу.

Модель для визначення виду психічного розладу, що може впливати на схильність до нетипової поведінки також є RNN-моделлю, однак зважаючи на роботу з десятьма класами, її архітектура є більш багатошаровою та складною. Архітектура нейромережевої моделі наведена на рисунку 2.3. Навідміну від попередньої моделі, тут окрім звичайних шарів присутні також дропаут шари.

Модель починається зі шару Embedding, який перетворює вхідні слова у вектори фіксованої довжини, що дозволяє моделі краще розуміти семантичне значення слів та їх контекстні взаємозв'язки. Вхідний розмір словника становить 5 000, вихідна розмірність векторів дорівнює 64, а довжина вхідної послідовності визначається параметром `max_length`, та може експериментально змінюватись.

Після цього йде шар LSTM з 64 нейронами, який обробляє послідовність даних і повертає повну послідовність результатів, що дозволяє наступним шарам мати доступ до всіх тимчасових етапів. Для запобігання перенавчанню використовується шар Dropout, який випадковим чином відкидає 50% нейронів.

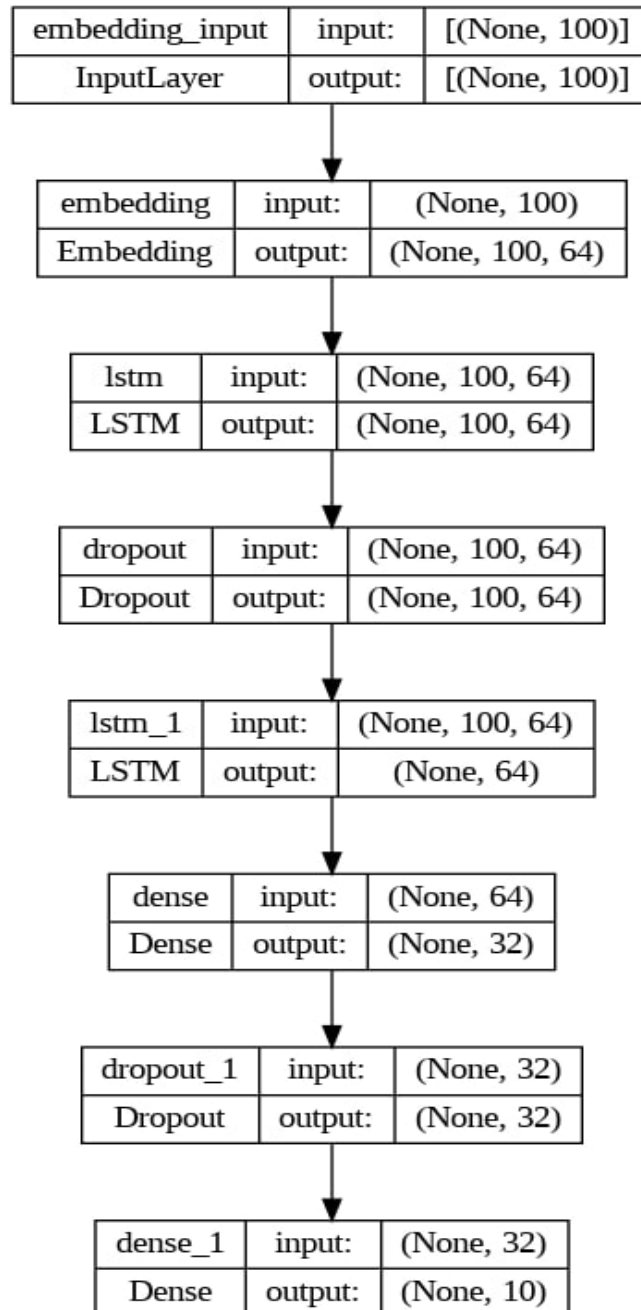


Рисунок 2.3 – Архітектура RNN-моделі для визначення виду психічного розладу

Наступний шар LSTM також має 64 нейрони, але на відміну від попереднього, він повертає лише останній вихідний стан, що представляє

загальне розуміння послідовності. Потім йде шар Dense з 32 нейронами і активацією ReLU, який забезпечує нелінійність і допомагає моделі вчитися складнішим взаємозв'язкам. Для подальшого зниження ймовірності перенавчання використовується ще один шар Dropout з такою ж ймовірністю 0.5.

Фінальний шар Dense має стільки нейронів, скільки є унікальних класів у навчальних даних. В даному випадку їх кількість визначається розміром множини міток класів `y_train`, що становить 10. Використання функції активації `softmax` у цьому шарі дозволяє отримати вектор ймовірностей для кожного класу, де кожне значення представляє ймовірність належності вхідного зразка до конкретного класу.

Отже, наведено архітектури для двох рекурентних нейромережових моделей, які є вхідними даними методу виявлення схильності до нетипової поведінки за текстовим дописом. Архітектура RNN-моделі для виявлення схильності до нетипової поведінки має 2 вихода, що відповідають за наявність схильності до нетипової поведінки або відсутність схильності до нетипової поведінки, а архітектура RNN-моделі для визначення виду психічного розладу передбачає 10 виходів, кожен з яких відповідає виду психічного розладу, що може впливати на схильність до нетипової поведінки.

### **2.3 Проектна архітектура інтелектуальної системи виявлення схильності до нетипової поведінки та взаємозв'язок компонентів**

При реалізації інтелектуальної системи на базі методу виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами необхідно сформувати структуру системи таким чином, щоб усі підсистеми та база даних коректно передавали між собою дані та результати.

На рисунку 2.4 наведено схему взаємодії підсистем інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом.

Підсистема для роботи з текстовими дописами призначена для взаємодії із контентом, аналіз якого буде проводити система. Підсистема передбачає зв'язок із базою даних, внесення нових записів, редагування та видалення існуючих, вибір записів для аналізу нетипової поведінки.

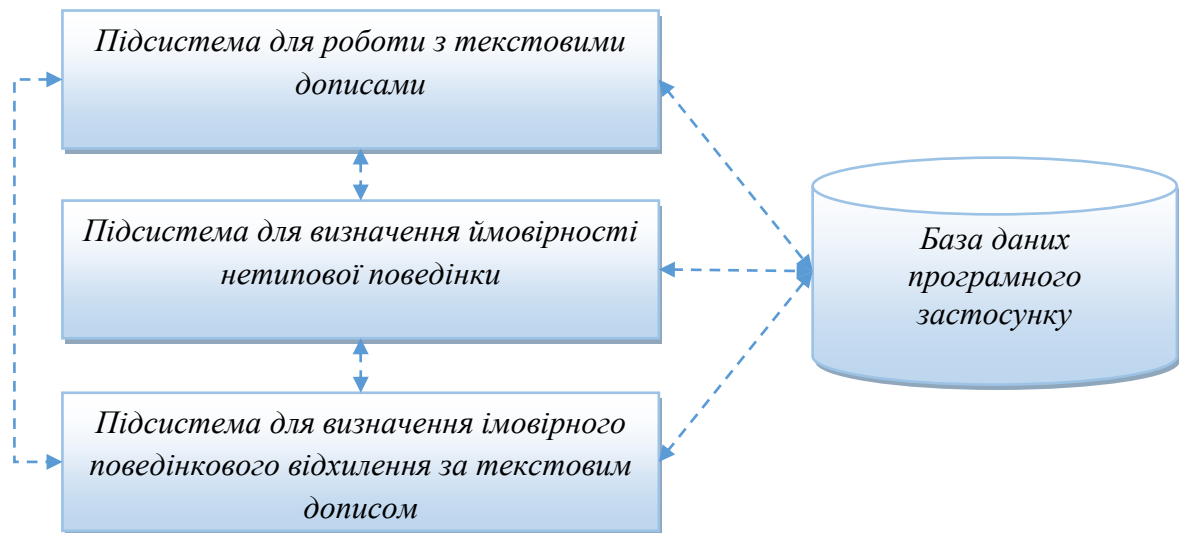


Рисунок 2.4 – Схема взаємодії підсистем інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом

Підсистема для роботи з текстовими дописами матиме декілька ключових функцій та компонентів, які дозволяють підготувати текст для подальшого аналізу для виявлення схильності до нетипової поведінки. Нижче наведено опис функцій підсистеми.

- Попередня обробка текстових даних, тобто підсистема очищатиме та нормалізуватиме текстові дані для подальшого аналізу. Це включає видалення стоп-слів, приведення до нижнього регістру та інші стандартні методи обробки тексту.

- Перегляд існуючого текстового допису для аналізу схильності до нетипової поведінки із дискового простору.

- Редагування існуючого текстового допису із дискового простору.

- Видалення існуючого текстового допису із дискового простору.

Підсистема для визначення ймовірності нетипової поведінки – це спеціалізований елемент програмного застосунку, що призначений для

виявлення імовірних випадків або тенденцій поведінки, що відхиляються від нормальних або очікуваних патернів. Підсистема містить наступні функції:

- вибір існуючого текстового допису для аналізу схильності до нетипової поведінки із дискового простору;
- визначення ймовірності нетипової поведінки в текстових дописах користувачів.

Також необхідно реалізувати окрему підсистему для визначення імовірного поведінкового відхилення за текстовим дописом, що виконує завдання аналізу текстового вмісту з метою виявлення потенційно аномальної або нетипової поведінки. Ця підсистема використовуватиме методи обробки природної мови, щоб зрозуміти зміст та контекст тексту, а також визначити, чи є поведінка, описана в дописі, відхиленням від норми. Підсистема містить наступні функції:

- вибір існуючого текстового допису для аналізу схильності до нетипової поведінки із дискового простору;
- визначення ймовірності нетипової поведінки в текстових дописах користувачів.

Отже, було створено схему взаємодії підсистем інтелектуальної системи на базі методу виявлення схильності до нетипової поведінки за текстовим дописом нейромережевими засобами.

#### **2.4 Проєктування бази даних програмної інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом**

Для розробки структури бази даних інтелектуальної системи на базі методу виявлення схильності до нетипової поведінки за текстовими дописами необхідно визначити сутності предметної області. На рисунку 2.5 зображено даталогічну модель бази даних інформаційної системи.

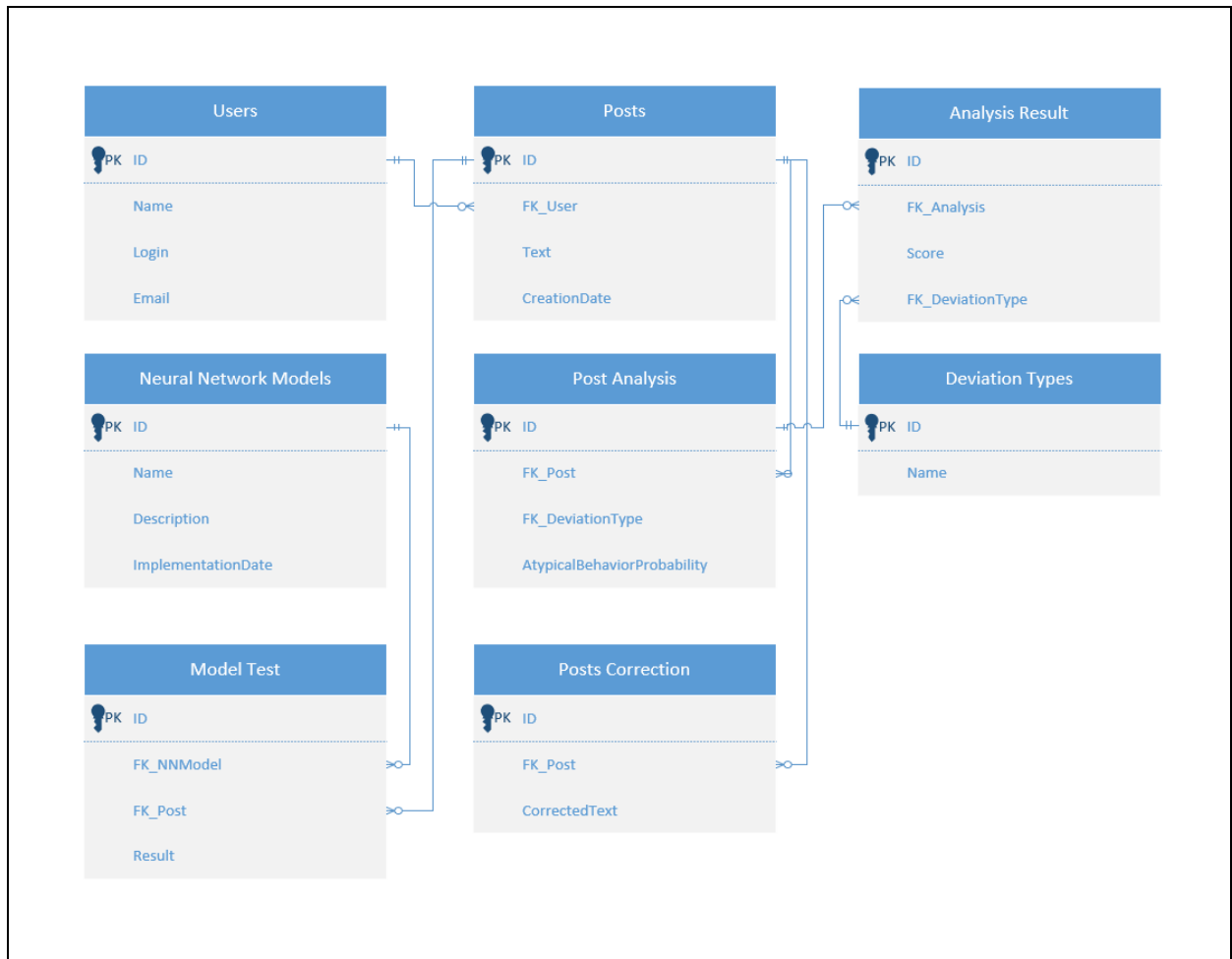


Рисунок 2.5 – Даталогічна модель інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача

Таблиця 2.1 – Атрибути таблиці «Users»

№ п/п	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ. Унікальний ідентифікатор користувача
2.	Name	varchar(255)	Ім'я користувача
3.	Login	varchar(255)	Логін користувача
4.	Email	varchar(255)	Електронна пошта користувача

В створеній БД є такі таблиці: Users, Posts, Analysis Result, NeuralNetworkModels, PostAnalysis, DeviationTypes, ModelTest та PostsCorrection. Кожна таблиця призначена для зберігання даних про окрему сутність.

Таблиця «Users» призначена для збереження даних про користувачів, які створюють пости. Атрибутами таблиці є: ID, Name, Login та Email (Таблиця 2.1).

Таблиця «Posts» містить дані про текстові дописи користувачів. Атрибутами таблиці є: ID, FK\_User, Text та CreationDate. (Таблиця 2.2).

Таблиця «AnalysisResult» зберігає інформацію про результати аналізу постів на предмет нетипової поведінки. Атрибутами таблиці є: ID, FK\_Analysis, Score та FK\_DeviationType . (Таблиця 2.3).

Таблиця 2.2 – Атрибути таблиці «Posts»

№ п/п	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ. Унікальний ідентифікатор допису
2.	FK_User	int	Вторинний ключ. Посилання на запис із таблиці “Users”. Використовується для визначення користувача що створив пост.
3.	Text	varchar(255)	Текстовий вміст допису
4.	CreationDate	datetime	Дата створення допису

Таблиця 2.3 – Атрибути таблиці «Analysis Result»

№ п/п	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ. Унікальний ідентифікатор запису результату
2.	FK_Analysis	int	Вторинний ключ. Посилання на запис із таблиці “Post Analysis”.
3.	Score	double	Числовий показник оцінки результату
4.	FK_DeviationType	int	Вторинний ключ. Посилання на запис із таблиці “Deviation Types”. Використовується для встановлення типу виявленого відхилення.

Таблиця «NeuralNetworkModels» призначена для збереження інформації про моделі нейронних мереж. Атрибутами таблиці є: ID, Name, Description та ImplementationDate. (Таблиця 2.4).

Таблиця 2.4 – Атрибути таблиці «NeuralNetworkModels»

№ п/п	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ. Унікальний ідентифікатор нейронної мережі
2.	Name	varchar(255)	Назва нейромережевої моделі.
3.	Description	varchar(255)	Короткий опис нейронної мережі
4.	ImplementationDate	datetime	Дата початку використання нейромережевої моделі

Таблиця «DeviationTypes» містить список різних видів нетипової поведінки. Атрибутами таблиці є: ID та Name. (Таблиця 2.5).

Таблиця 2.5 – Атрибути таблиці «DeviationTypes»

№ п/п	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ. Унікальний ідентифікатор типу відхилення
2.	Name	varchar(255)	Назва типу відхилення, наприклад, "агресія", "депресія", "пропаганда".

Таблиця «Post Analysis» використовується для збереження даних про аналіз постів на предмет нетипової поведінки. Атрибутами таблиці є: ID, FK\_Post, FK\_DeviationType та AtypicalBehaviorProbability. (Таблиця 2.6).

Таблиця 2.6 – Атрибути таблиці «PostAnalysis»

№	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ. Унікальний ідентифікатор запису аналізу
2.	FK_Post	int	Вторинний ключ. Посилання на запис із таблиці “Posts”.
3.	FK_DeviationType	int	Вторинний ключ. Посилання на запис із таблиці “Deviation Types”. Використовується для встановлення типу виявленого відхилення.
4.	AtypicalBehavior-Probability	double	Числова ймовірність нетипової поведінки

Таблиця «Model Test» використовується для відстеження результатів тестування нейромережових моделей на конкретних постах. Атрибути таблиці є: ID, FK\_NNModel, FK\_Post та Result. (Таблиця 2.7).

Таблиця 2.7 – Атрибути таблиці «Model Test»

№	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ. Унікальний ідентифікатор тестування
2.	FK_NNModel	int	Вторинний ключ. Посилання на запис із таблиці “Neural Network Models”.
3.	FK_Post	int	Вторинний ключ. Посилання на запис із таблиці “Posts”. Використовується для визначення допису, який було проаналізовано.
4.	Result	varchar(255)	Опис результатів випробування, як у текстовому, так і в числовому форматі.

Таблиця «Posts Correction» зберігає дані про редагування дописів після аналізу. Атрибутами таблиці є: ID, FK\_Post та CorrectedText. (Таблиця 2.8).

Таблиця 2.8 – Атрибути таблиці «Posts Correction»

№	Назва	Тип даних	Опис
1.	ID	int	Первинний ключ. Унікальний ідентифікатор типу редагування
2.	FK_Post	int	Вторинний ключ. Посилання на запис із таблиці “Posts”. Використовується для визначення допису, який було відредаговано.
3.	CorrectionPosts	varchar(255)	Новий (або змінений) текст допису після редагування.

Таким чином, було створено базу даних для збереження інформації із застосунку на базі методу виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами та представлено відповідну даталогічну модель.

## 2.5 Підготовка робочих вхідних даних для системи

У якості робочих вхідних даних для виявлення схильності до нетипової поведінки за текстовим дописом буде використано набір, що містить інформацію, опубліковану в субредітах, пов'язаних з психічним здоров'ям. Набір даних «Human Stress Prediction» включає різні аспекти проблем психічного здоров'я, які люди висвітлюють у своїх повідомленнях [26].

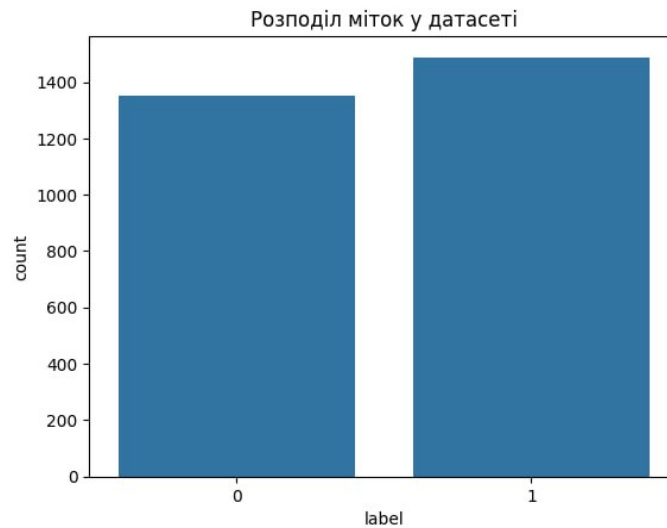


Рисунок 2.6 – Розподіл даних за категоріями

Даний набір містить маркери, де 0 означає відсутність стресу, а 1 вказує на наявність стресу, що є значним фактором схильності до нетипової поведінки. Розподіл даних за категоріями наведено на рисунку 2.6 Записів, що містять прояви схильності до нетипової поведінки налічується 1488, таких, що не містять – 1350.

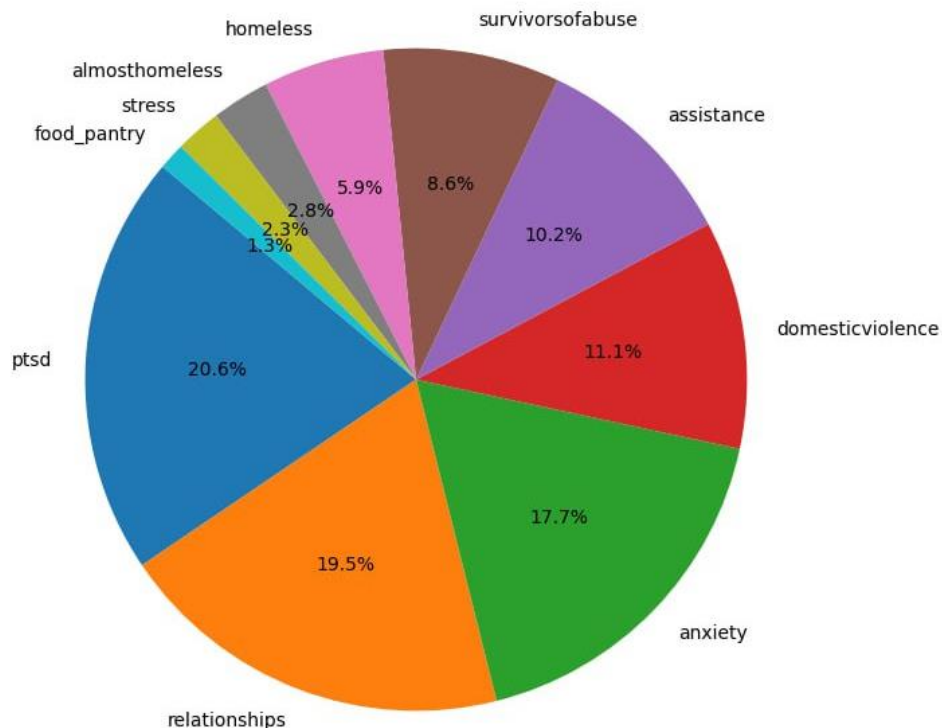


Рисунок 2.7 – Розподіл записів за видами психічних розладів

Окрім бінарного розподілу даний датасет містить додаткову розмітку по видах психічних порушень (10 класів). Розподіл даних за видами психічних порушень, що можуть впливати на прояви нетипової поведінки наведено на рисунку 2.7.

Кількісний розподіл за 10 категоріями психологічних розладів, що можуть впливати на виявлення схильності до нетипової поведінки наведено на рисунку 2.8.

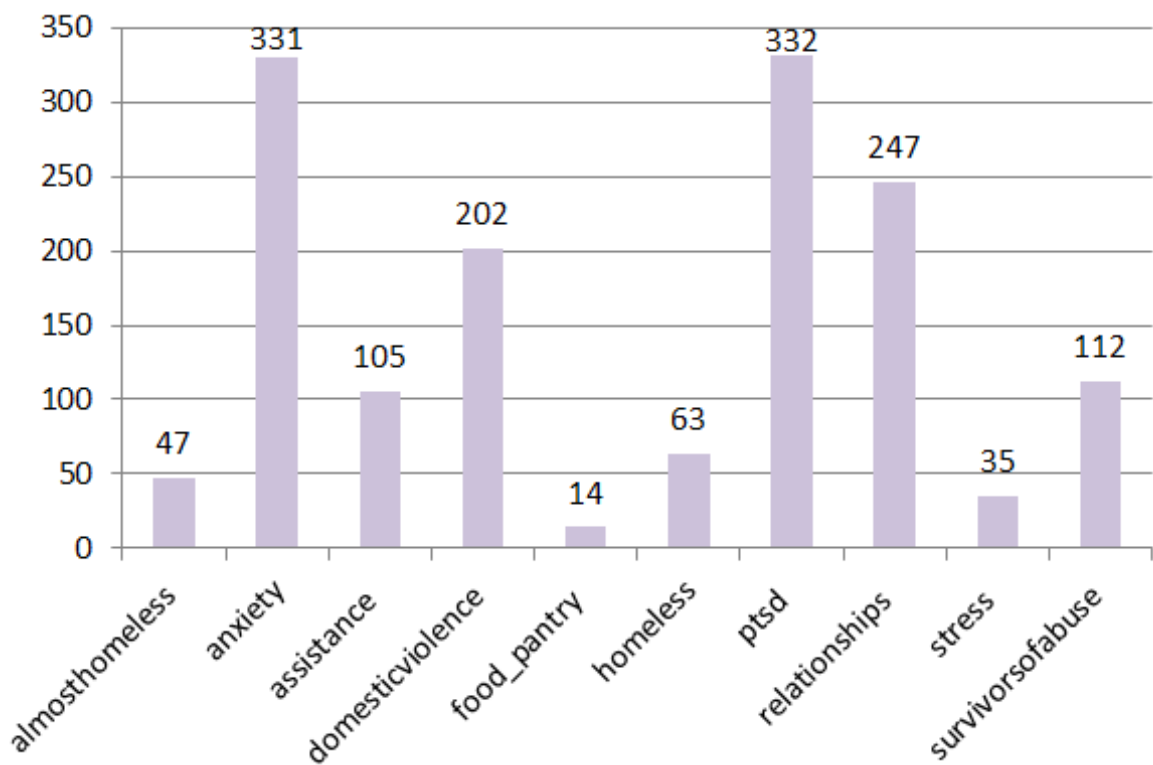


Рисунок 2.8 – Кількісний розподіл дописів, що містять прояви психологічних розладів

Як видно з рисунку 2.8, дані не збалансовані. Тому для класів, що містять кількість дописів менше 100, дані було доповнено за допомогою великої мовної моделі GPT 3.5.

Отже, для навчання нейромережі, яка буде здійснювати бінарну оцінку щодо проявів схильності до нетипової поведінки буде використано 1488 записів, що містять схильності до нетипової поведінки та 1350 таких, що не містять схильностей до нетипової поведінки. Для навчання нейромережі, яка буде

показувати на тип психічного розладу, що найімовірніше може призвести до проявів нетипової поведінки, буде використано набір даних з понад 1500 дописів, що представляють кожен вид психічного розладу.

## **2.6 Особливості використання спеціалізованих програмних компонентів для рекурентних неймереж**

Для подальшої програмної реалізації інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом доцільно використовувати спеціалізовані програмні компоненти, що допоможуть значно спростити розробку.

«TensorFlow» широко використовувана бібліотека для машинного навчання, розроблена компанією Google. Вона забезпечує гнучку екосистему інструментів, бібліотек і ресурсів спільноти, що дозволяє дослідникам і розробникам створювати та впроваджувати моделі машинного навчання. TensorFlow підтримує широкий спектр алгоритмів машинного навчання і глибокого навчання і здатний виконувати розрахунки на різних платформах, включаючи CPU, GPU і TPU, що забезпечує високу продуктивність і масштабованість [27]. Логотип наведено на рисунку 2.9.

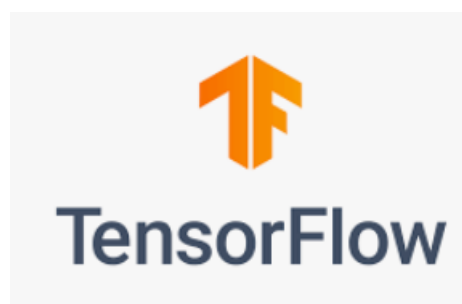


Рисунок 2.9 – Логотип бібліотеки «TensorFlow»

Архітектура «TensorFlow» дозволяє будувати обчислювальні графи, де вузли представляють математичні операції, а ребрами є багатовимірні масиви даних (тензори), що передаються між ними. Це забезпечує ефективне виконання

складних обчислень і дозволяє оптимізувати моделі. Бібліотека широко застосовується в академічних дослідженнях і промислових проектах, сприяючи розвитку таких галузей, як обробка природної мови, комп'ютерний зір і системи рекомендацій.

У кваліфікаційній роботі бакалавра дана бібліотека буде використана з метою побудови та навчання рекурентних нейронних мереж для виявлення схильності до нетипової поведінки.

Наступною бібліотекою, яку планується використати є «Imblearn», бібліотека Python, яка надає інструменти для роботи з незбалансованими наборами даних [28]. Одним із її основних компонентів є метод SMOTE, який використовується для збалансування наборів даних шляхом створення синтетичних зразків менш представленого класу. Приклад роботи бібліотеки наведено на рисунку 2.10. Дана бібліотека буде використана з метою доповнення навчальної вибірки для нейромережі для визначення виду психічного розладу.

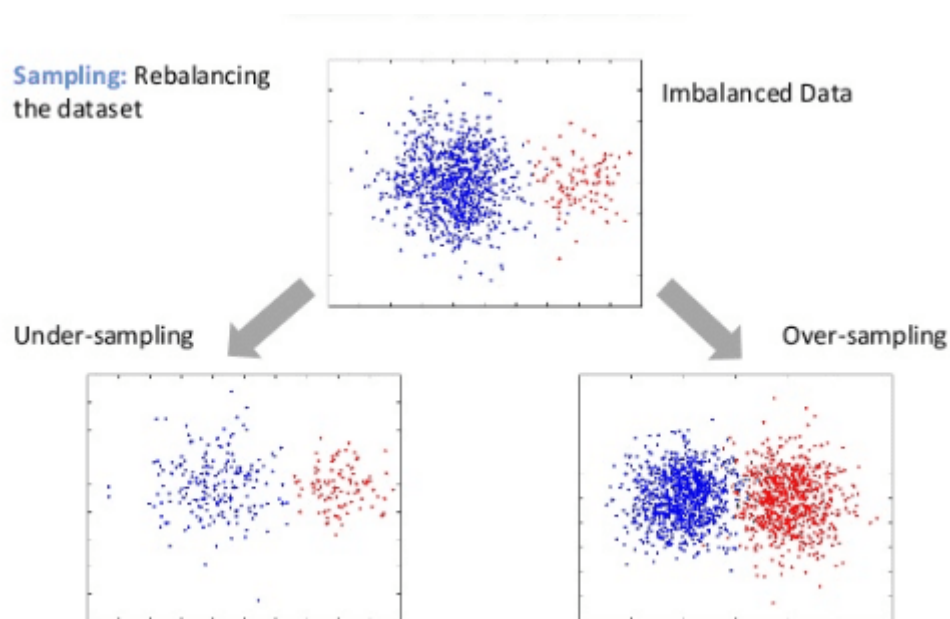


Рисунок 2.10 – Приклад доповнення датасету

Отже, для навчання та оцінки якості навчених нейронних мереж для виявлення схильності до нетипової поведінки буде використано бібліотеку «TensorFlow» з метою побудови та навчання рекурентних нейронних мереж, а

також «Imblearn», з метою доповнення навчальної вибірки для нейромережі для визначення виду психічного розладу.

## 2.7 Висновки до розділу 2

Створено метод виявлення схильності до нетипової поведінки за текстовим дописом, що присвячений покращенню виявлення схильності до нетипової поведінки за рахунок його аналізу нейромережевими засобами. Метод працює шляхом перетворення вхідних даних у вигляді текстового допису, навченої нейромережевої моделі для виявлення схильності до нетипової поведінки та навченої нейромережевої моделі для визначення виду психічного розладу, що впливає на нетипову поведінку у вихідні дані у вигляді числової оцінки схильності до нетипової поведінки та визначений вид психічного розладу.

Наведено архітектури для двох рекурентних нейромережевих моделей, що є вхідними даними методу виявлення схильності до нетипової поведінки за текстовим дописом. Архітектура RNN-моделі для виявлення схильності до нетипової поведінки має 2 виходи, що відповідають за наявність схильності до нетипової поведінки або відсутність схильності до нетипової поведінки, а архітектура RNN-моделі для визначення виду психічного розладу передбачає 10 виходів, кожен з яких відповідає виду психічного розладу, що може впливати на схильність до нетипової поведінки.

Наведено схему взаємодії підсистем інтелектуальної системи на базі методу виявлення схильності до нетипової поведінки за текстовим дописом нейромережевими засобами. Розроблена інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом складається із підсистеми для роботи з текстовими дописами, підсистеми для визначення ймовірності нетипової поведінки, підсистеми визначення імовірного поведінкового відхилення за текстовим дописом, а також бази даних.

Створено базу даних для збереження інформації із інтелектуальної системи на базі методу виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами та представлено відповідну даталогічну модель.

Виконано підготовку робочих вхідних даних, для навчання нейромережі, яка буде здійснювати бінарну оцінку щодо проявів схильності до нетипової поведінки буде використано 1488 записів, що містять схильності до нетипової поведінки та 1350 таких, що не містять схильностей до нетипової поведінки. Для навчання нейромережі, яка буде показувати на тип психічного розладу, що найімовірніше може призвести до проявів нетипової поведінки, буде використано набір даних з 1000 дописів, по 100 на кожен вид психічного розладу.

Було обрано спеціальні програмні компоненти, для навчання та оцінки якості навчених нейронних мереж в задачі виявлення схильності до нетипової поведінки буде використано бібліотеку «TensorFlow» з метою побудови та навчання рекурентних нейронних мереж, а також бібліотеку «Imblearn» з метою доповнення навчальної вибірки для нейромережі для визначення виду психічного розладу.

За створеним методом та спроектованою структурою інтелектуальної системи виявлення схильності до нетипової поведінки в подальшому необхідно розробити застосунок, з використанням якого провести дослідження ефективності методу виявлення схильності до нетипової поведінки. Розроблене програмне забезпечення необхідно протестувати на коректність виконання заявлених функцій.

### Розділ 3 Програмна реалізація інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом

#### 3.1 Визначення шляхів дослідження та засобів створення інтелектуальної системи виявлення схильності до нетипової поведінки

Метою кваліфікаційної роботи бакалавра є розробка методу виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами та відповідного програмного забезпечення – інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом.

Програмний застосунок для виявлення схильності до нетипової поведінки за текстовим дописом користувача має містити реалізацію наступних груп функцій (рисунок 3.1).



Рисунок 3.1 – Групи функцій користувача інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом

Програмний застосунок на базі методу виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами має виконувати наступні функції, а саме:

- надавати можливість користувачеві внесення нових записів, редагування та видалення існуючих;
- надавати можливість користувачеві обирати необхідний файл для аналізу для подальшого дослідження;
- повертати результати щодо імовірного поведінкового відхилення за текстовим дописом та імовірної нетипової поведінки.

Окрім того, необхідно провести ряд тестів інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом, дослідження ефективності отриманого методу та програмного застосунку на його базі.

### **3.2 Вибір засобів розробки інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача**

Для реалізації інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами було обрано мову програмування Python.

Python – це високорівнева мова програмування, розроблена Гвідо ван Россумом і вперше випущена у 1991 році. Ця мова відома своєю простотою та ефективністю, що робить її популярним вибором серед програмістів.

Python має велику кількість бібліотек для різних завдань, що робить його дуже популярним для наукових обчислень, обробки даних, штучного інтелекту, веб-розробки та багатьох інших сфер. У Python не потрібно явно вказувати типи змінних, оскільки вони визначаються автоматично при присвоєнні значення.

Python має велику кількість стандартних бібліотек, а також величезну кількість сторонніх бібліотек, що дозволяє вирішувати широкий спектр завдань. Зокрема, для створення застосувань, що використовують ШІ ця мова

програмування має ряд важливих переваг. У Python існує велика кількість бібліотек та фреймворків, призначених спеціально для розробки штучного інтелекту. Наприклад, TensorFlow, PyTorch, scikit-learn, NLTK тощо, що спрощує створення та розвиток різноманітних моделей ШІ.

Python має велику та активну спільноту розробників, яка постійно розвиває та підтримує різноманітні інструменти для ШІ. Це означає, що завдяки активному співробітництву і обміну знаннями ви можете знайти відповіді на багато питань та проблем, які виникають під час розробки. Python використовується у багатьох галузях ШІ, включаючи машинне навчання, обробку природної мови, комп'ютерне зору, обробку сигналів та багато інших. Це дає можливість створювати широкий спектр рішень у сфері штучного інтелекту за допомогою однієї мови програмування.

Python підтримує об'єктно-орієнтоване програмування, що дозволяє структурувати програми за допомогою класів та об'єктів, може працювати на різних операційних системах, включаючи Windows, macOS та різні дистрибутиви Linux.

Python використовується в таких областях, як наукові дослідження, аналіз даних, розробка веб-додатків, штучний інтелект, машинне навчання та багато інших. Його потужність та простота роблять його одним із найпопулярніших інструментів для програмування сучасності.

Для реалізації інтерфейсу користувача було обрано бібліотеку PySimpleGui, це проста у використанні бібліотека для створення графічних інтерфейсів користувача в програмах Python. Основною метою PySimpleGUI є забезпечення легкості створення інтерфейсів для користувачів без необхідності вивчення складних фреймворків або використання важких інструментів.

Особливістю PySimpleGUI є його простий та зрозумілий синтаксис, який дозволяє швидко створювати інтерфейси за допомогою мінімального коду. Вона підтримує різні типи віджетів, такі як кнопки, текстові поля, списки, таблиці та інші, що дозволяє створювати різноманітні та функціональні інтерфейси.

PySimpleGUI може бути використаний для різних цілей, включаючи створення додатків з графічним інтерфейсом для настільних програм, інструментів управління даними, інтерактивних веб-сайтів та багато іншого. Вона є ідеальним вибором для початківців, а також для досвідчених розробників, які шукають швидкий та простий спосіб створення GUI для своїх програм.

Застосунок на базі методу виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами використовуватиме відповідну базу даних, що була створена за допомогою SQLite. SQLite – це легка використання вбудована база даних, яка забезпечує можливість зберігати та керувати даними без необхідності встановлення окремого сервера баз даних. Одна з основних переваг SQLite полягає в його простоті та компактності, що робить його ідеальним вибором для невеликих проектів або додатків з обмеженими ресурсами.

Однією з особливостей SQLite є те, що вся база даних зберігається у єдиному файлі на диску, що спрощує розгортання та резервне копіювання даних. Крім того, SQLite підтримує стандартні SQL-запити, що дозволяє виконувати різноманітні операції з даними, такі як вставка, вибірка, оновлення та видалення записів.

Отже, було проведено вибір засобів розробки інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами. Було обрано наступний набір засобів: мова програмування Python, бібліотека для створення інтерфейсу користувача PySimpleGUI та СКБД SQLite.

### **3.3 Структура та функціональне призначення програмних складових інтелектуальної системи виявлення схильності до нетипової поведінки**

Реалізація методу виявлення схильності до нетипової поведінки за текстовим дописом користувача у вигляді програмного застосунку буде виконана за спроектованою інформаційною системою.

Для реалізації інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами необхідно сформулювати діаграму класів програми. На рисунку 3.2 представлено діаграму класів інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами.

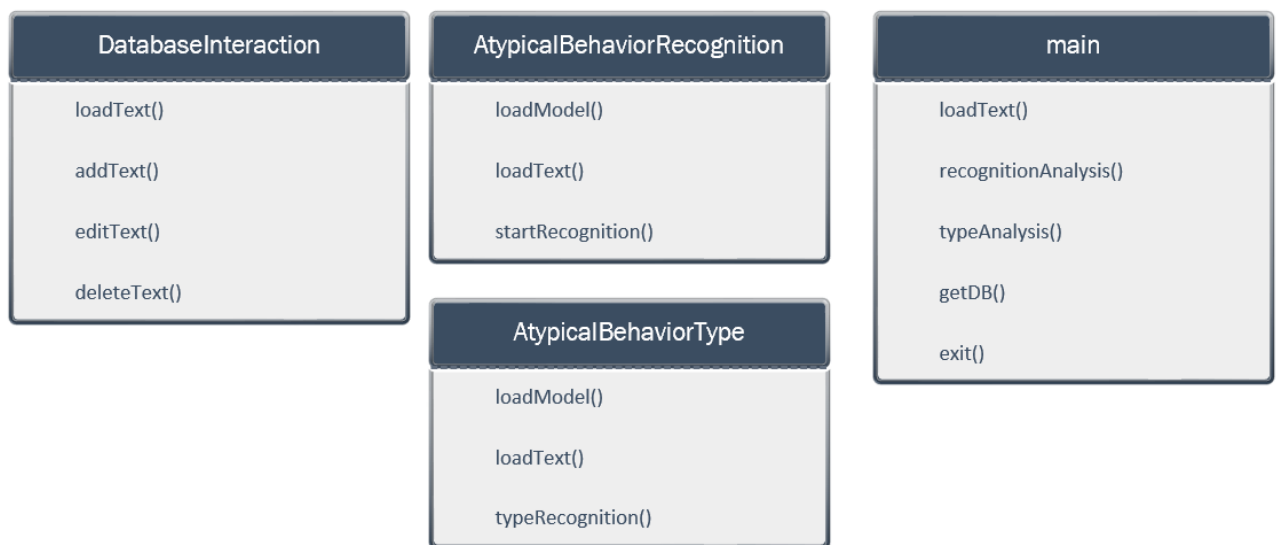


Рисунок 3.2 – Діаграма класів інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача

Клас `DatabaseInteraction` призначений для взаємодії з базою даних, і його методи дозволяють виконувати основні операції CRUD з текстовими даними. Метод `loadText()` завантажує текстові дані з бази даних, одне поле тексту або набір текстових записів.

Метод `addText()` додає новий текст до бази даних. Метод має параметри для визначення, який саме текст додавати та як він має бути структурований, рекомендовано завантажувати тексти від 50 до 150 слів. Метод `editText()` призначений для редагування існуючого тексту у базі даних. Метод `deleteText()` видаляє текст із бази даних.

Клас `AtypicalBehaviorRecognition` призначений для розпізнаванням патернів у тексті, що вказують на нетипову поведінку, таку як тривога, стрес, або

інші аномалії. Основне призначення класу – повернути бінарне значення класифікації, чи було в тексті визначено нетипову поведінку. Метод `loadModel()` завантажує попередньо навчену модель для розпізнавання нетипової поведінки. Метод `loadText()` завантажує текст, який буде аналізуватися на предмет нетипової поведінки. Метод `startRecognition()` запускає процес розпізнавання нетипової поведінки. Метод застосовує завантажену модель до текстових даних для виявлення аномалій або специфічних патернів.

Клас `AtypicalBehaviorType` відповідає за визначення конкретних типів нетипової поведінки та їх розпізнавання. Клас містить наступні методи для визначення ймовірного поведінкового відхилення. Метод `loadModel()` завантажує модель для розпізнавання типів нетипової поведінки. Метод `loadText()` призначений для завантаження тексту для аналізу на предмет певного виду нетипової поведінки. Метод `typeRecognition()` виконує розпізнавання конкретного типу нетипової поведінки за допомогою попередньо навченої нейромережевої моделі.

Клас `main` – основний клас програми, який поєднує в собі функціональність взаємодії з базою даних, визначення імовірності нетипової поведінки в тексті та його класифікації. Метод `loadText()` завантажує текст із бази даних, це початковий крок для подальшої обробки або аналізу. Метод `recognitionAnalysis()` перенаправляє до процесу аналізу розпізнавання. Він звертається до класу `AtypicalBehaviorRecognition`, щоб запустити розпізнавання та отримати результати. Метод `typeAnalysis()` відповідає за перенаправлення до форми або компонента, що виконує аналіз типу, що використовує клас `AtypicalBehaviorType` для визначення конкретних типів нетипової поведінки. Метод `getDB()` перенаправляє до механізму взаємодії з базою даних, що використовує клас `DatabaseInteraction`, щоб отримати доступ до бази даних для виконання CRUD-операцій. `Exit()` – метод, який перенаправляє на процедуру виходу із системи. Може завершувати поточні процеси, закривати відкриті ресурси і забезпечувати плавний вихід із програми.

Таким чином, в результаті опису структури та функціонального призначення програмних складових системи було створено діаграму класів інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами та описано групи функцій програмного застосунку.

### **3.4 Особливості реалізації програмних складових інтелектуальної системи виявлення схильності до нетипової поведінки**

Програмний застосунок на базі методу виявлення схильності до нетипової поведінки за текстовим дописом нейромережевими засобами міститиме наступні підсистеми:

- підсистема для роботи з текстовими дописами;
- підсистема для визначення ймовірності нетипової поведінки ;
- підсистема для визначення імовірного поведінкового відхилення за текстовим дописом.

Відповідно, необхідно реалізувати програмний функціонал для кожної з підсистем інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами. Основна задача – провести навчання нейромережевої моделі для визначення ймовірності нетипової поведінки та визначення імовірного поведінкового відхилення за текстовим дописом. Також потрібно реалізувати зручний інтерфейс користувача інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами для забезпечення функцій роботи із БД та нейромережевою моделлю.

Очистка датасету перед навчанням нейронної мережі є важливим етапом в процесі машинного навчання, оскільки якість вхідних даних безпосередньо впливає на продуктивність та точність моделі. Шум та зайва інформація в датасеті можуть негативно вплинути на тренування моделі, змушуючи її виявляти патерни, які не мають значення, або призвести до хибного навчання.

Для підготовки та очистки датасету було обрано алгоритм SMOTE, що означає Synthetic Minority Over-sampling Technique. Це техніка балансування наборів даних, яка використовується для вирівнювання розподілу класів у машинному навчанні. Вона особливо корисна, коли один або кілька класів у наборі даних значно менш представлені порівняно з іншими, що може призвести до проблем у навчанні моделей. SMOTE працює шляхом створення синтетичних зразків для меншості класів, щоб вирівняти кількість зразків між різними класами. На відміну від простого дублювання існуючих зразків, SMOTE створює нові зразки, що допомагає зменшити ризик перенавчання.

SMOTE вибирає випадковий зразок з менш представленого класу та визначає кілька «сусідів» для обраного зразка. Це робиться за допомогою алгоритму найближчих сусідів (наприклад, k-Nearest Neighbors).

На рисунку 3.3 наведено алгоритм, що був використаний під час написання програмного коду для балансування даних.

Для кожного обраного зразка SMOTE створює нові синтетичні зразки шляхом інтерполяції між обраним зразком та його сусідами. Цей процес передбачає обчислення різниці між вихідним зразком і його сусідом, а потім помноження цієї різниці на випадкове число (від 0 до 1), щоб отримати новий зразок.

Перший крок алгоритму – імпортування бібліотеки SMOTE з `imblearn.over_sampling`, який є основним методом для балансування. Після цього відбувається завантаження набору даних, а саме самі текстові дані з датасету та мітки, які ідентифікують клас кожного тексту.

Наступний крок – токенизація, тобто перетворення тексту в послідовності чисел. Це необхідно для обробки тексту нейронними мережами. Метод `pad_sequences()` заповнює (або обрізає) послідовності до однакової довжини. Це забезпечує сумісність у обробці даних. Далі відбувається ініціалізація `SMOTE(random_state=42)`, створюється об'єкт SMOTE, який буде використовуватися для балансування даних, а `random_state=42` встановлює фіксований випадковий стан, щоб результати були відтворюваними. Для

балансування даних із SMOTE використовується метод `smote.fit_resample(X_pad, labels)`, де `X_pad` – заповнені послідовності тексту, які є вхідними даними для SMOTE, `labels` – мітки, які відповідають кожному текстовому зразку.

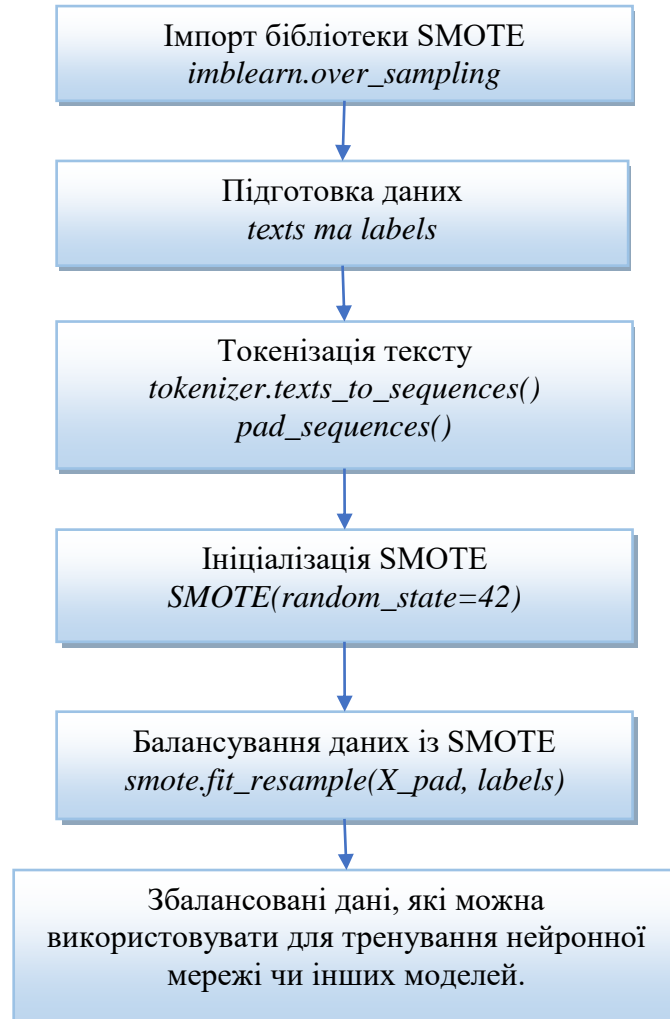


Рисунок 3.3 – Алгоритм SMOTE для балансування даних

Метод `fit_resample()` використовує алгоритм SMOTE для балансування даних, він знаходить зразки менш представленого класу та генерує нові зразки шляхом інтерполяції між існуючими зразками.

Вихідними результатами є `X_resampled` і `y_resampled`, які містять збалансований набір даних, де `X_resampled` – збалансовані послідовності тексту, `y_resampled` – збалансовані мітки, що відповідають новому набору тексту.

Вихідним результатом є збалансовані дані, які можна використовувати для тренування нейронної мережі чи інших моделей. В таблиці 3.1 представлені

результати, отримані після очищення та підготовки датасету перед навчанням нейромережевої моделі, в порівнянні з результатами навчання моделі на тому ж датасеті, але без попереднього балансування даних.

Перед балансуванням даних, продуктивність моделі була суттєво нижчою, що вказувало на потенційні проблеми з навчанням та узагальненням. Високий рівень втрат та низька точність на тренувальних даних, поряд із ще вищим рівнем валідаційних втрат і низькою валідаційною точністю, свідчили про те, що модель, ймовірно, не мала достатньої кількості релевантних даних для ефективного навчання. Ці проблеми часто є наслідком дисбалансу в наборі даних, коли один або кілька класів представлені значно менше, ніж інші.

Таблиця 3.1 – Порівняння отриманих результатів навчання моделі

<b>Параметр</b>	<b>Результат до балансування даних</b>	<b>Результат після балансування даних</b>
<b>loss</b>	1.6754	0.3104
<b>accuracy</b>	0.4523	0.9265
<b>val_loss</b>	1.9856	0.3221
<b>val_accuracy</b>	0.2905	0.9150

Після застосування балансування, результати значно покращилися. Загальні втрати на тренувальних даних зменшилися, тоді як точність суттєво підвищилася, наблизившись до 93%. Валідаційні втрати також зменшилися, вказуючи на поліпшення здатності моделі узагальнювати на невидимих даних. Валідаційна точність, яка раніше становила лише 29%, піднялася до 91,5%, що свідчить про значне поліпшення валідаційної продуктивності.

Ці зміни в результатах підкреслюють важливість балансування даних перед тренуванням моделей машинного навчання. Зокрема, використання методів, таких як SMOTE, сприяє вирівнюванню розподілу класів у наборі даних, зменшуючи ризик упередженості моделі. Загалом, результати демонструють, що балансування даних стало ефективною стратегією для

покращення продуктивності моделі машинного навчання для застосунку на базі методу виявлення схильності до нетипової поведінки за текстовим дописом нейромережевими засобами.

Також було реалізовано інтерфейс користувача інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами (рисунок 3.4).

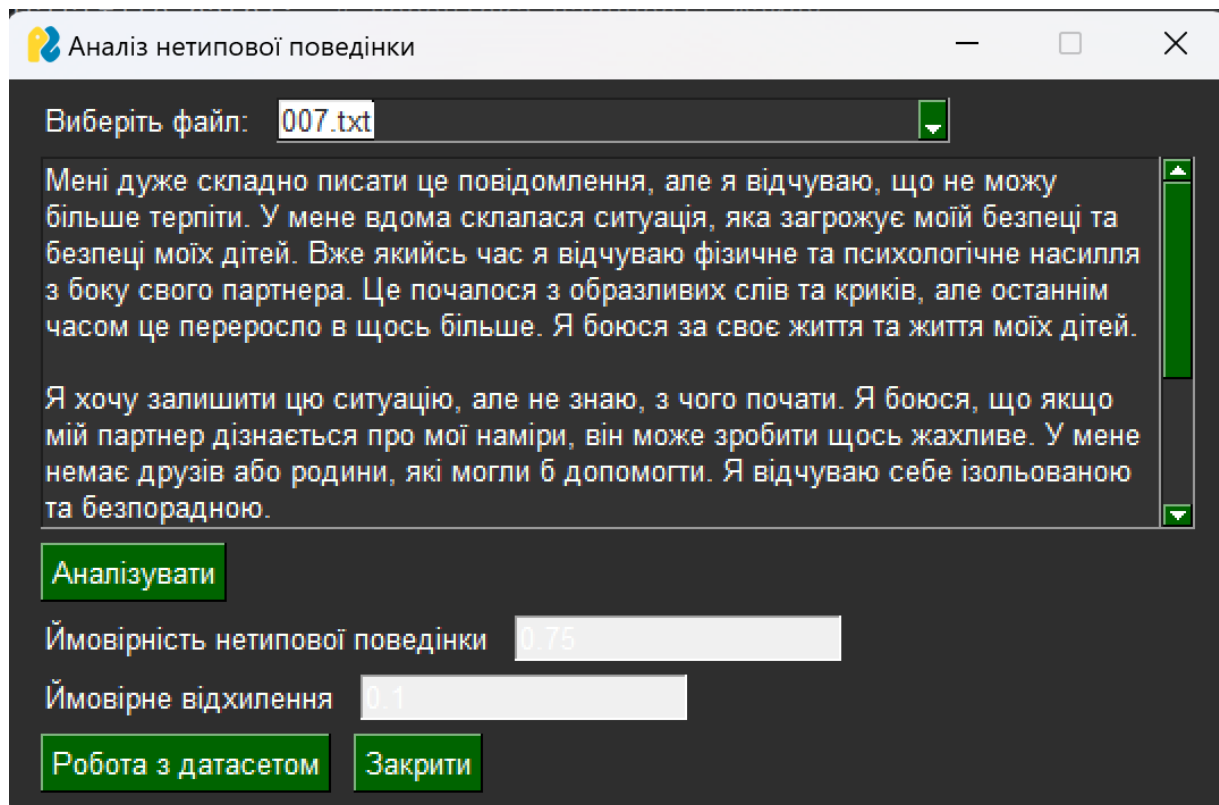


Рисунок 3.4 – Інтерфейс головного вікна інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами

Також необхідно реалізувати форму для взаємодії користувача інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами із базою даних. Форма має містити наступні поля: ID користувача, текст допису та ймовірне поведінкове відхилення, ідентифіковане в тексті. На рисунку 3.5 наведено результат реалізації програмного коду.

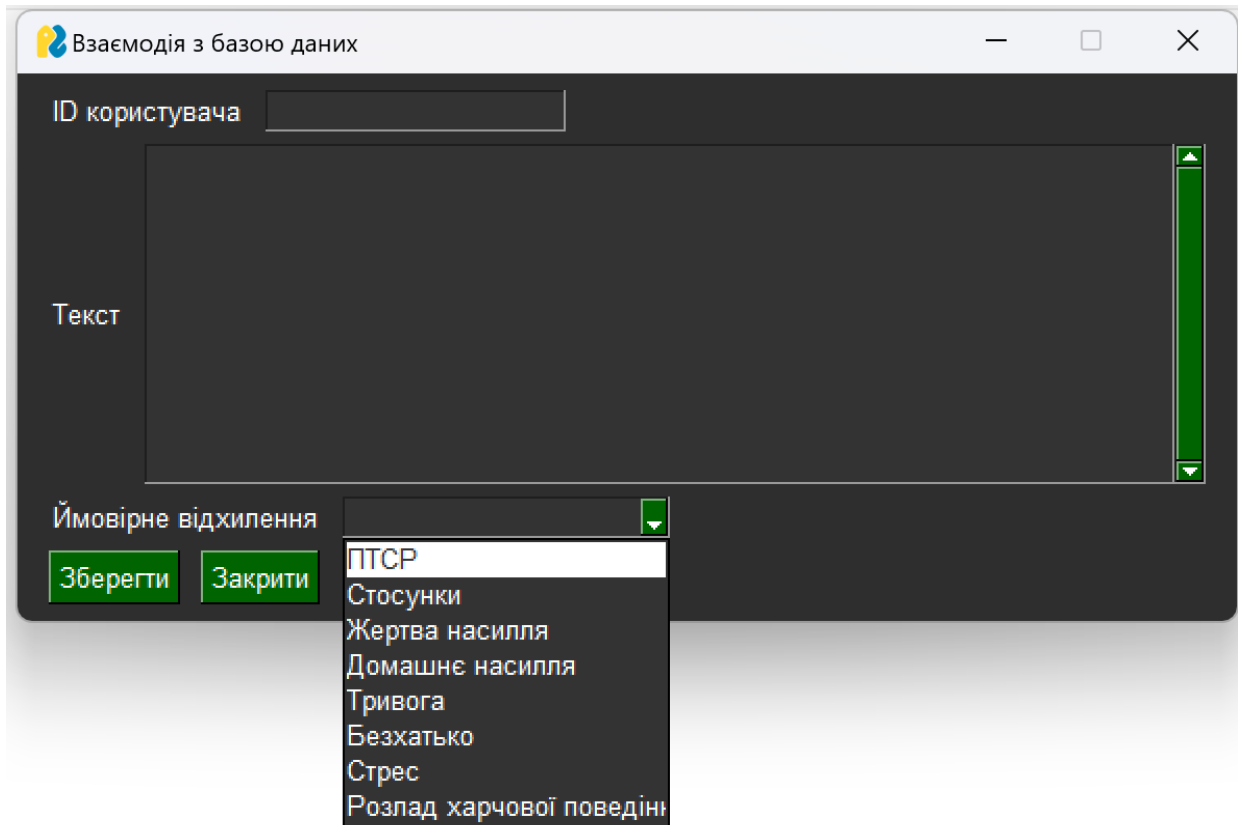


Рисунок 3.5 – Інтерфейс форми інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом для роботи із базою даних

На головній формі інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами користувач має змогу обрати запис, що вже розміщений в базі даних або ж ввести допис вручну. Користувач має змогу проаналізувати повідомлення одразу або перейти до роботи із датасетом.

Таким чином, було реалізовано програмний застосунок на базі методу виявлення схильності до нетипової поведінки за текстовим дописом нейромережевими засобами, що містить наступні підсистеми, а саме: підсистему для роботи з текстовими дописами, підсистему для визначення ймовірності нетипової поведінки, підсистему для визначення імовірного поведінкового відхилення за текстовим дописом.

### 3.5 Тестування інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом та вимоги до розгортання

Для забезпечення надійного виявлення схильності до нетипової поведінки за текстовими дописами необхідно провести ретельне тестування інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами. Цей застосунок використовує нейромережеві методи для аналізу текстів та визначення ймовірності нетипової поведінки. Він складається з кількох підсистем, кожна з яких відповідає за конкретний аспект процесу виявлення. Разом ці підсистеми створюють комплексний інструмент для аналізу текстових дописів з метою виявлення потенційно небезпечної або нетипової поведінки. Перший тестовий випадок наведено в таблиці 3.2

Перш за все необхідно перевірити стабільність підключення до бази даних, для цього було створено тест-кейс А-01 (таблиця 3.2).

Таблиця 3.2 – Тест-кейс А-01

Тест-кейс ID: А-01	Пріоритет: Major	Створено: 15.03.2024
Назва: Тест-кейс для перевірки підключення до бази даних		
Кроки		Очікуваний результат
1. Запустити програмний застосунок 2. Натиснути кнопку «Робота з датасетом» 3. На формі «Робота із базою даних» натиснути кнопку «Перевірити підключення» 4. Порівняти отриманий результат з очікуваним		Відображення повідомлення «Підключення до бази даних встановлене»
Результат виконання тест-кейсу: пройдено успішно		

На рисунку 3.6 наведено результат виконання тест-кейсу, отже перевірку підключення до бази даних успішно пройдено.

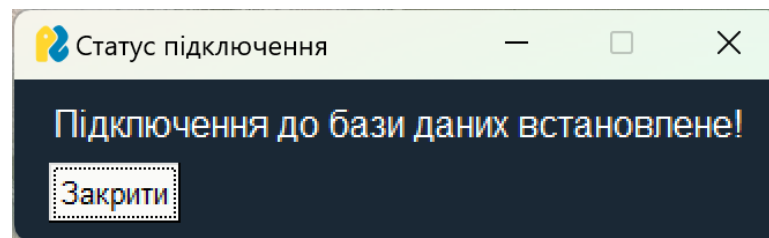


Рисунок 3.6 – Результат тест-кейсу А-01

Також необхідно перевірити функції для аналізу тексту, визначення ймовірності нетипової поведінки та типу ймовірного відхилення. Для цього було реалізовано тест-кейс А-02 (таблиця 3.3).

Таблиця 3.3 – Тест-кейс А-02

Тест-кейс ID: А-02	Пріоритет: Major	Створено: 15.03.2024
Назва: Тест-кейс для перевірки правильності визначення ймовірного поведінкового відхилення		
Кроки		Очікуваний результат
<ol style="list-style-type: none"> <li>Запустити програмний застосунок;</li> <li>Ввести запропонований текст із ознаками поведінкового відхилення «домашнє насилля» у відповідне поле;</li> <li>Натиснути кнопку «Аналізувати»;</li> <li>Порівняти фактичний результат з очікуваним</li> </ol>		<p>Відображення наступних результатів:</p> <p>Ймовірність нетипової поведінки &gt;0.5</p> <p>Ймовірне поведінкове відхилення «домашнє насилля»</p>
Результат виконання тест-кейсу: пройдено успішно		

Для виконання тест-кейсу необхідно підготувати текст із ймовірним поведінковим відхиленням «домашнє насилля»: *«Мені дуже складно писати це повідомлення, але я відчуваю, що не можу більше терпіти. У мене вдома*

*склалася ситуація, яка загрожує моїй безпеці та безпеці моїх дітей. Вже якийсь час я відчуваю фізичне та психологічне насилля з боку свого партнера. Це почалося з образливих слів та криків, але останнім часом це переросло в щось більше. Я боюся за своє життя та життя моїх дітей».*

Результат виконання тест-кейсу A-02 наведено на рисунку 3.7.

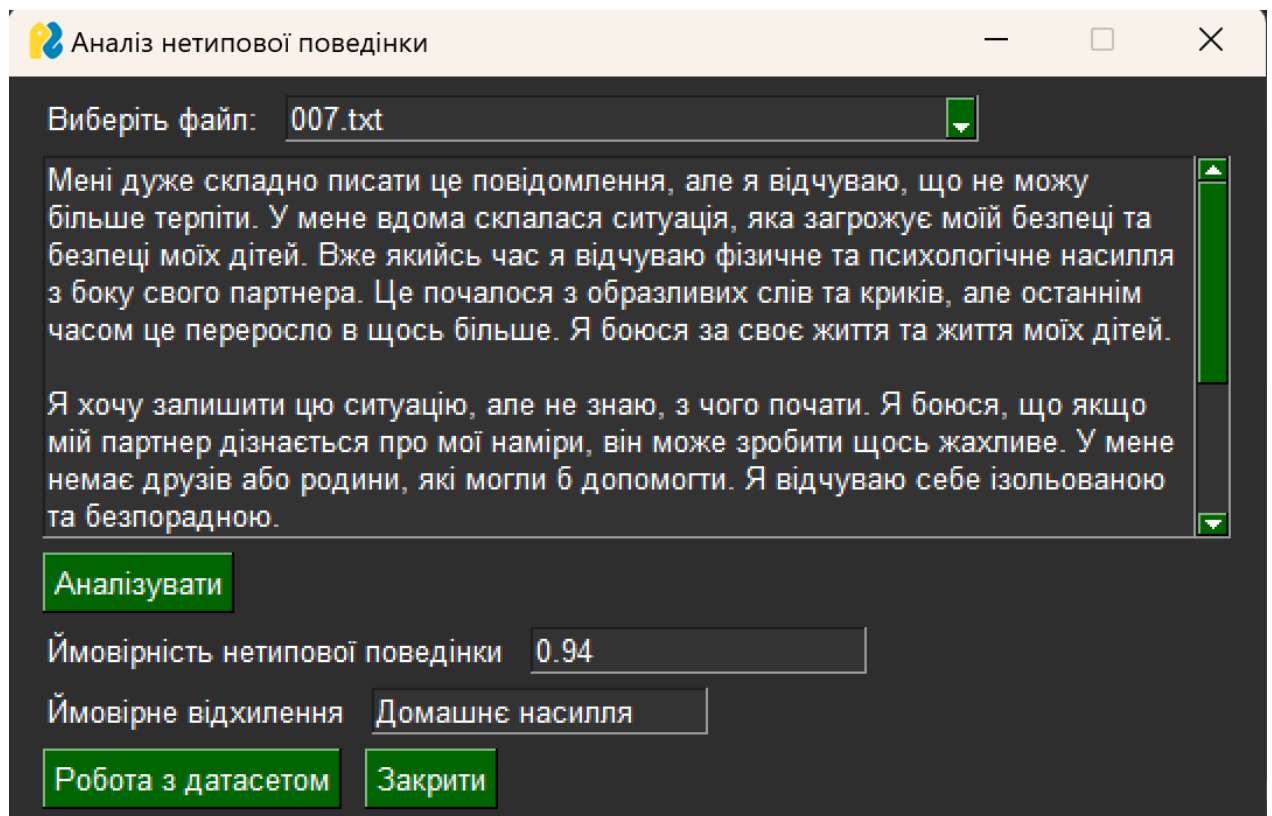


Рисунок 3.7 – Результат виконання тест-кейсу A-02

Також необхідно провести unit-тестування програмного застосунку. За допомогою бібліотеки `pytest` було створено тести для основних функцій програмного продукту. На рисунку 3.8 наведено результати виконання unit-тестування. Звіт показує, що всі тест-кейси пройшли успішно, без помилок. Значення [25%], [50%], [75%], [100%] вказують на прогрес у проходженні тестів, а «PASSED» – на успішне проходження тестування.

Таким чином, було проведено комплексне тестування програмного продукту, що використовує метод виявлення схильності до нетипової поведінки за текстовими дописами. Тестування охопило різні аспекти продукту, від



Таблиця 3.4 – Результати проведення дослідження

<i>До проведення SMOTE-балансування</i>			
Кількість епох	Точність	Функція втрат	Час навчання, секунд
10	0,27	2,19	80
25	0,35	1,93	170
50	0,51	1,76	350
<i>Після проведення SMOTE-балансування</i>			
10	0,34	0,81	80
25	0,62	0,55	154
50	0,915	0,3221	280

Після проведення SMOTE-балансування модель показує значне покращення у всіх аспектах: точність значно зросла, функція втрат знизилася, а час навчання зменшився. Це свідчить про те, що балансування даних допомогло в ефективному навчанні, що призвело до кращої продуктивності моделі. Результати підтверджують ефективність SMOTE для покращення навчання моделі, особливо в контексті дисбалансованих даних. На рисунку 3.9 наведено графіки, що відображають зміни точності при різній кількості епох.

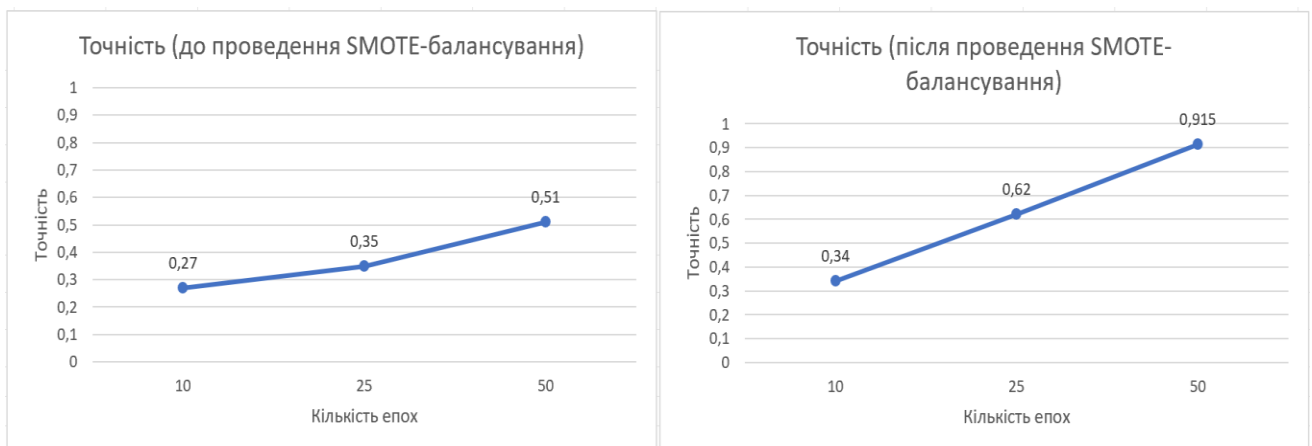


Рисунок 3.9 – Графіки змін значень точності моделі

Також на рисунку 3.10 представлено матрицю сплутувань до здійснення SMOTE-балансування датасету. Оскільки модель мала низьку точність, на рисунку відображено відповідні результати.

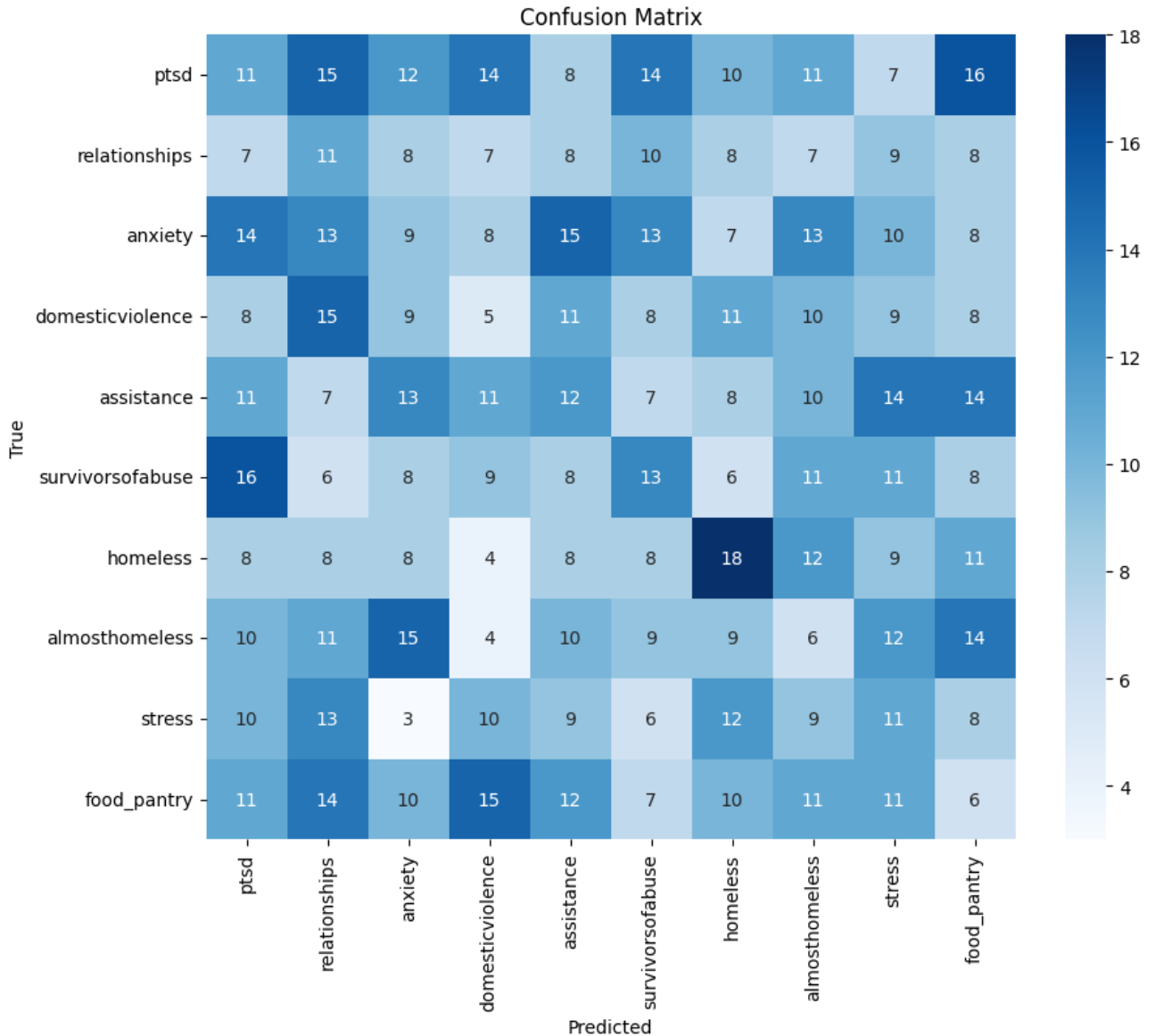


Рисунок 3.10 – Матриця сплутувань до SMOTE-балансування датасету

Оскільки до SMOTE-балансуванням модель навчається на даних з нерівномірним розподілом класів, вона може зсунути в бік класів з більшою кількістю прикладів, що призводить до низької точності в передбаченні міноритарних класів. SMOTE допомагає виправити це шляхом створення синтетичних прикладів для міноритарного класу, збільшуючи його кількість.

Міноритарні класи можуть бути представлені недостатньо, що робить їх передбачення складним для моделі. Застосування SMOTE допомогло вирівняти кількість прикладів у кожному класі, надаючи моделі більше даних для навчання та покращуючи її продуктивність.

Для візуалізації результатів було повторно створено матрицю сплутувань (рисунок 3.11). Із отриманого зображення можна встановити, що за умови достатньої кількості даних для всіх класів, матриця сплутувань показує більше правильних передбачень і менше помилок (менше значень поза діагоналлю).

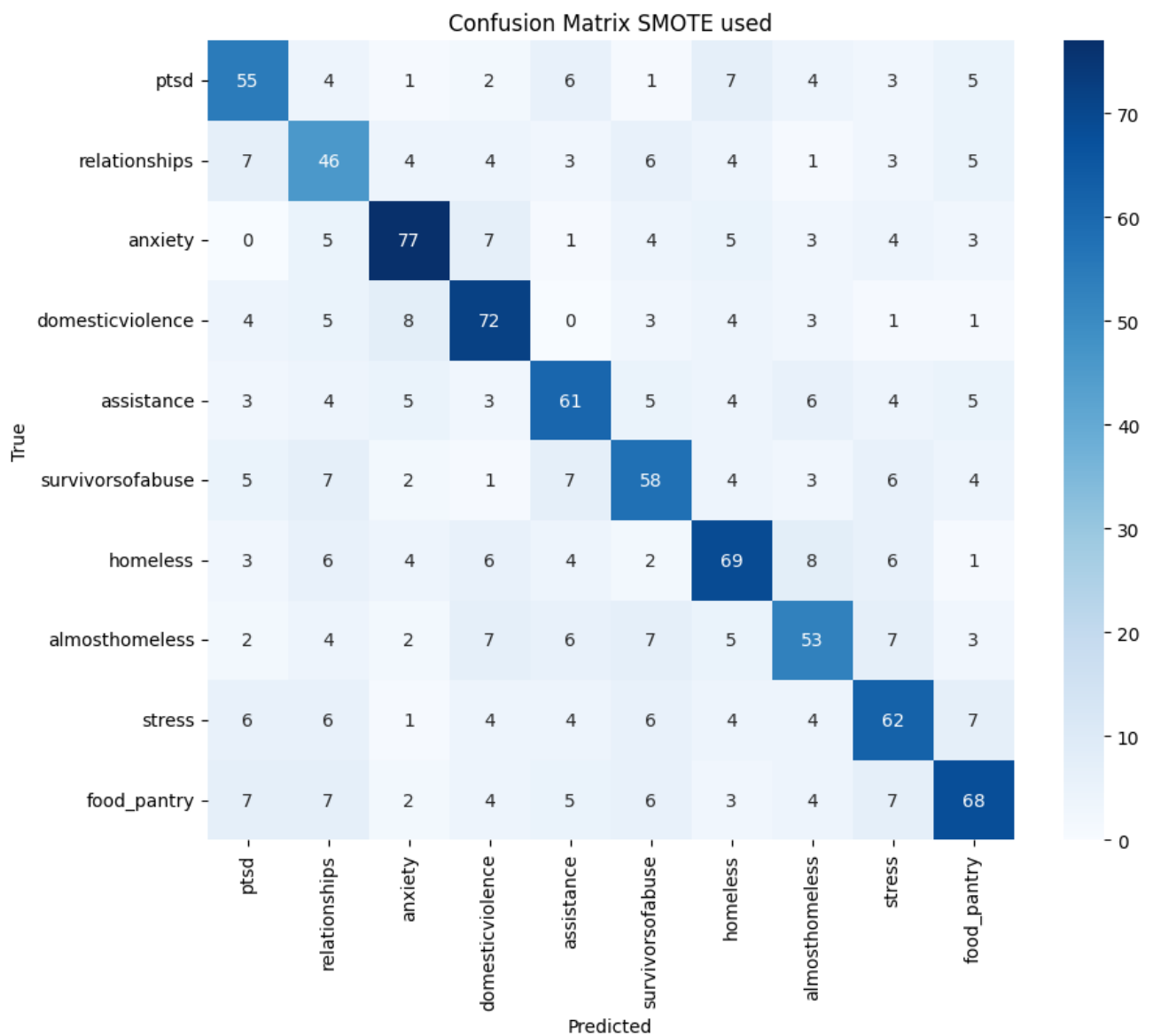


Рисунок 3.11 – Матриця сплутувань після SMOTE-балансування датасету

Таким чином, у ході реалізації інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами було досягнуто значних результатів у виявленні ймовірної аномальної поведінки у користувачів соціальних мереж та інших онлайн-платформ.

Здійснивши SMOTE-балансування датасету перед навчанням нейромережевої моделі, було досягнуто значного покращення показників продуктивності. Результати, які продемонстрували точність 0,915 та значення функції втрат 0,3221, вказують на те, що модель змогла успішно навчитися розпізнавати патерни в даних, при цьому мінімізуючи помилки та підвищуючи точність передбачень.

### **3.7 Висновки до розділу 3**

В результаті виконання розділу було створено програмну реалізацію системи на базі методу виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами. Програмний застосунок на базі методу виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами виконує наступні функції, а саме:

- надає можливість користувачеві внесення нових записів, редагування та видалення існуючих;
- надає можливість користувачеві обирати необхідний файл для аналізу для подальшого дослідження;
- повертає результати щодо імовірного поведінкового відхилення за текстовим дописом та імовірної нетипової поведінки.

Також було описано структуру та функціональне призначення програмних складових інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами, було створено схему взаємодії підсистем програмного застосунку на

базі методу виявлення схильності до нетипової поведінки за текстовим дописом нейромережевими засобами, діаграму класів програмного застосунку на базі методу виявлення схильності до нетипової поведінки за текстовим дописом нейромережевими засобами та описано групи функцій інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами.

Було проведено вибір засобів розробки інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами. Було обрано наступний набір засобів: мова програмування Python, бібліотека для створення інтерфейсу користувача PySimpleGUI та СКБД SQLite.

Результати тестування показали, що програмний продукт здатний аналізувати текстові дописи та визначати ймовірність нетипової поведінки. Усі критичні функції інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами пройшли тестування без помилок, а обробка помилок та валідація даних забезпечують стабільну роботу продукту у складних сценаріях.

Здійснивши SMOTE-балансування датасету перед навчанням нейромережевої моделі, було досягнуто значного покращення показників продуктивності. Результати, які продемонстрували точність 0,915 та значення функції втрат 0,3221, вказують на те, що модель змогла успішно навчитися розпізнавати патерни в даних, при цьому мінімізуючи помилки та підвищуючи точність передбачень.

Отриманий застосунок надає змогу аналізувати текстові дописи, виявляючи тривожні сигнали, які можуть свідчити про емоційні чи психологічні проблеми. Програмний продукт дозволяє вчасно ідентифікувати такі ризики, що дає змогу вжити превентивних заходів для підтримки благополуччя користувачів.

Таким чином, метод виявлення схильності до нетипової поведінки за текстовими дописами, реалізований у вигляді інтелектуальної системи

виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами, не лише допомагає у ранньому виявленні потенційних проблем, але також сприяє розвитку нових підходів до психосоціальної підтримки та вивченню поведінкових патернів, що може стати цінним ресурсом у сучасному світі.

## Висновки

Метою кваліфікаційної роботи бакалавра було покращення виявлення схильності до нетипової поведінки за текстовим дописом користувача за рахунок його аналізу нейромережевими засобами. Для досягнення мети було розроблено метод виявлення схильності до нетипової поведінки за текстовим дописом користувача й відповідне програмне забезпечення, а також були поставлені та вирішені такі завдання:

1. Виконано аналіз інформаційних моделей області виявлення схильності до нетипової поведінки. Встановлено, що автоматизоване знаходження ознак нетипової поведінки в текстових дописах є важливим завданням інформаційних технологій, що спрямоване на забезпечення безпеки та етичного спілкування в цифровому середовищі.

2. Виконано огляд теоретичних підходів, а також обрано підхід для нейромережевого виявлення схильності до нетипової поведінки на основі рекурентних нейронних мереж, оскільки він дає вищу точність та можливість розуміння природної мови.

3. Виконано аналіз існуючих публікацій за напрямком дослідження, а також проведено аналіз існуючого програмного забезпечення області виявлення схильності до нетипової поведінки за текстовим дописом користувача.

4. Створено метод виявлення схильності до нетипової поведінки за текстовим дописом користувача. Метод працює шляхом перетворення вхідних даних у вигляді текстового допису, навченої нейромережевої моделі для виявлення схильності до нетипової поведінки та навченої нейромережевої моделі для визначення виду психічного розладу, що впливає на нетипову поведінку у вихідні дані у вигляді числової оцінки схильності до нетипової поведінки та визначений вид психічного розладу.

5. Описано архітектуру системи та взаємозв'язок її компонентів для виявлення схильності до нетипової поведінки за текстовим дописом користувача.

6. Спроектовано базу даних інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами.

7. Обрано набір даних для навчання нейромережі. Для навчання нейромережі, що буде здійснювати бінарну оцінку проявів схильності до нетипової поведінки буде використано 1488 записів, що містять схильності до нетипової поведінки та 1350 таких, що не містять схильностей до нетипової поведінки. У якості навчального набору нейромережі, яка буде показувати на тип психічного розладу, що найімовірніше може призвести до проявів нетипової поведінки, буде використано набір даних з понад 1500 дописів, що представляють кожен вид психічного розладу.

8. Визначено шляхи дослідження та засоби створення інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами.

9. Створено програмну реалізацію на основі розробленого методу.

10. Виконано тестування створеного інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами, яке показало коректність описаних функцій.

11. Проведено дослідження ефективності створеного методу виявлення схильності до нетипової поведінки з використанням розробленої інтелектуальної системи виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами. Результати, які продемонстрували точність 0,915 та значення функції втрат 0,3221, вказують на те, що модель змогла успішно навчитися розпізнавати патерни в даних, при цьому мінімізуючи помилки та підвищуючи точність передбачень.

Кваліфікаційна робота бакалавра виконана у повному обсязі, отриманий результат відповідає поставленим завданням. Подальші дослідження будуть спрямовані на розширення психічних розладів, що можуть призводити до нетипової поведінки, пошуки відповідних наборів даних та перенавчання

нейромереж. Також планується порівняння використаної архітектури RNN з моделями-трансформерами.

Результат виконання кваліфікаційної роботи бакалавра повністю відповідає поставленим завданням та меті. Основні наукові та практичні результати доповідалися в доповіді «Recurrent Neural Network Model Architecture for Detecting a Tendency to Atypical Behavior Of Individuals by Text Posts» на XXVI Міжнародній науково-практичній конференції «Theoretical and Practical Aspects of Modern Research» 5 червня 2024 року (Ottawa, Canada, International Scientific Unity), за темою кваліфікаційної роботи бакалавра автором виконано наукову публікацію [29].

## Перелік посилань

1. Mayo Clinic. Stress symptoms: Effects on your body and behavior. URL: <https://www.mayoclinic.org/healthy-lifestyle/stress-management/in-depth/stress-symptoms/art-20050987>
2. Brainly. What is social dimension of stress? URL: <https://brainly.ph/question/23262165>
3. Lake behavioral hospital. What are the effects of stress on your mental health? URL: <https://www.lakebehavioralhospital.com/what-are-the-effects-of-stress-on-your-mental-health/>
4. Omh. Common Stress Reactions A Self-Assessment. URL: [https://omh.ny.gov/omhweb/disaster\\_resources/pandemic\\_influenza/doctors\\_nurses/common\\_stress\\_reactions.html](https://omh.ny.gov/omhweb/disaster_resources/pandemic_influenza/doctors_nurses/common_stress_reactions.html)
5. Study. Atypical Human Development: Definition & Examples. URL: <https://study.com/academy/lesson/abnormal-human-development-definition-examples.html>
6. Psychology writing. Human Behavior and Factors of Its Formation. URL: <https://psychologywriting.com/human-behavior-and-factors-of-its-formation/>
7. Lumen learning. What are Psychological Disorders? URL: <https://courses.lumenlearning.com/waymaker-psychology/chapter/what-are-psychological-disorders/>
8. WHO. Mental disorders. URL: <https://www.who.int/news-room/fact-sheets/detail/mental-disorders>
9. Mind. Trauma. URL: <https://www.mind.org.uk/information-support/types-of-mental-health-problems/trauma/effects-of-trauma/>
10. Living by example. How Is Social Media Changing Human Behavior? URL: <https://www.livingbyexample.org/how-is-social-media-changing-human-behavior/>
11. Sirivisa creative. Sharing Secrets on Social Media. URL: <https://www.sirivisacreative.com/blogs/sharing-secrets-on-social-media>

12. Study. Abnormal Behavior | Definition, Criteria & Examples. URL: <https://study.com/learn/lesson/abnormal-behavior-examples-criteria.html>
13. M. M. Rock. Social Media and Behavioral Health: Normal and Abnormal Behaviors. Behavioral Pediatric Healthcare for Nurse Practitioners. 2018. URL: <https://study.com/learn/lesson/abnormal-behavior-examples-criteria.html>
14. BuiltIn. What Is Machine Learning? URL: <https://builtin.com/machine-learning>
15. Datacamp. What is Natural Language Processing (NLP)? A Comprehensive Guide for Beginners. URL: <https://www.datacamp.com/blog/what-is-natural-language-processing>
16. Spiceworks. What Is a Support Vector Machine? Working, Types, and Examples. URL: <https://www.spiceworks.com/tech/big-data/articles/what-is-support-vector-machine/>
17. Deepai. What is K-Means Clustering? URL: <https://deepai.org/machine-learning-glossary-and-terms/k-means>
18. Simplilearn. Random Forest Algorithm. URL: <https://www.simplilearn.com/tutorials/machine-learning-tutorial/random-forest-algorithm>
19. Datatron. The Naive Bayes Classifier. URL: <https://datatron.com/the-naive-bayes-classifier/>
20. Datacamp. What are Recurrent Neural Networks (RNN). URL: <https://www.datacamp.com/tutorial/tutorial-for-recurrent-neural-network>
21. Facebook. URL: <https://www.facebook.com>
22. Facebook. How does Facebook use artificial intelligence to moderate content? URL: <https://www.facebook.com/help/1584908458516247>
23. Zhang T., Schoene A.M., Ji S. Natural language processing applied to mental illness detection: a narrative review. npj Digit. Med. 5, 46. 2022. URL: <https://doi.org/10.1038/s41746-022-00589-7>
24. Shrestha A., Serra E., Spezzano F. Multi-modal social and psycholinguistic embedding via recurrent neural networks to identify depressed users in

online forums. Netw Model Anal Health Inform Bioinforma 9, 22. 2020. URL: <https://doi.org/10.1007/s13721-020-0226-0>

25. Hwang Y., Kim H., Choi H., Lee J. Exploring Abnormal Behavior Patterns of Online Users With Emotional Eating Behavior: Topic Modeling Study. J Med Internet Res 2020;22(3):e15700. 2022. URL: <https://www.jmir.org/2020/3/e15700>

26. Kaggle. Human Stress Prediction. URL: <https://www.kaggle.com/datasets/kreeshrajani/human-stress-prediction>

27. TensorFlow і машинне навчання. URL: <https://foxminded.ua/tensorflow-shcho-tse/>

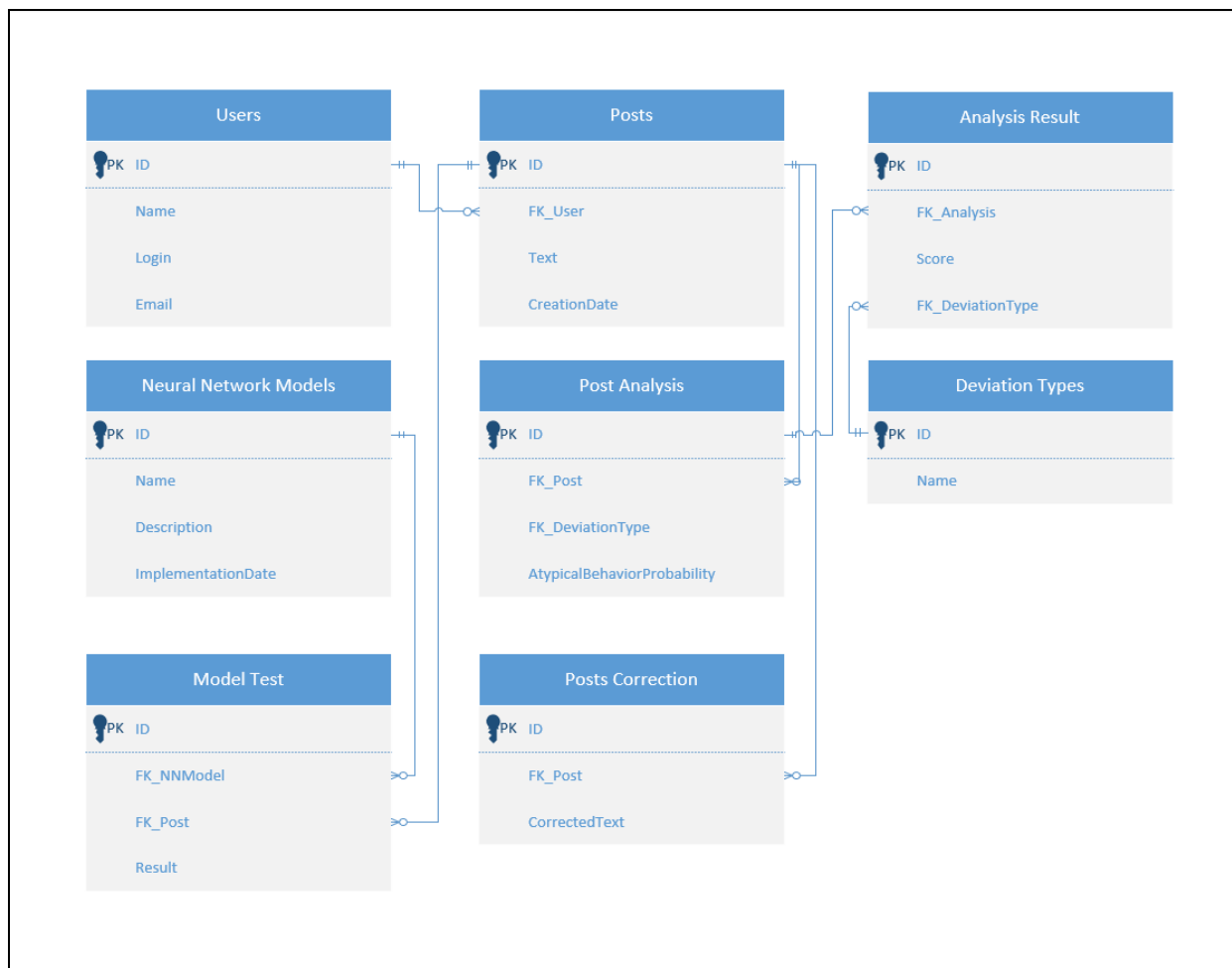
28. Handling Imbalanced Datasets With imblearn Library. URL: <https://medium.com/thecyphy/handling-imbalanced-datasets-with-imblearn-library-df5e58b968f4>

29. Sobko O., Mazurets O., Didur V., Chervonchuk I. Recurrent Neural Network Model Architecture for Detecting a Tendency to Atypical Behavior Of Individuals by Text Posts. Theoretical and Practical Aspects of Modern Research. Proceedings of XXVI International scientific and practical conference. June 5-7, 2024. International Scientific Unity. Ottawa, Canada. 2024. Pp. 113-117.

# ДОДАТКИ

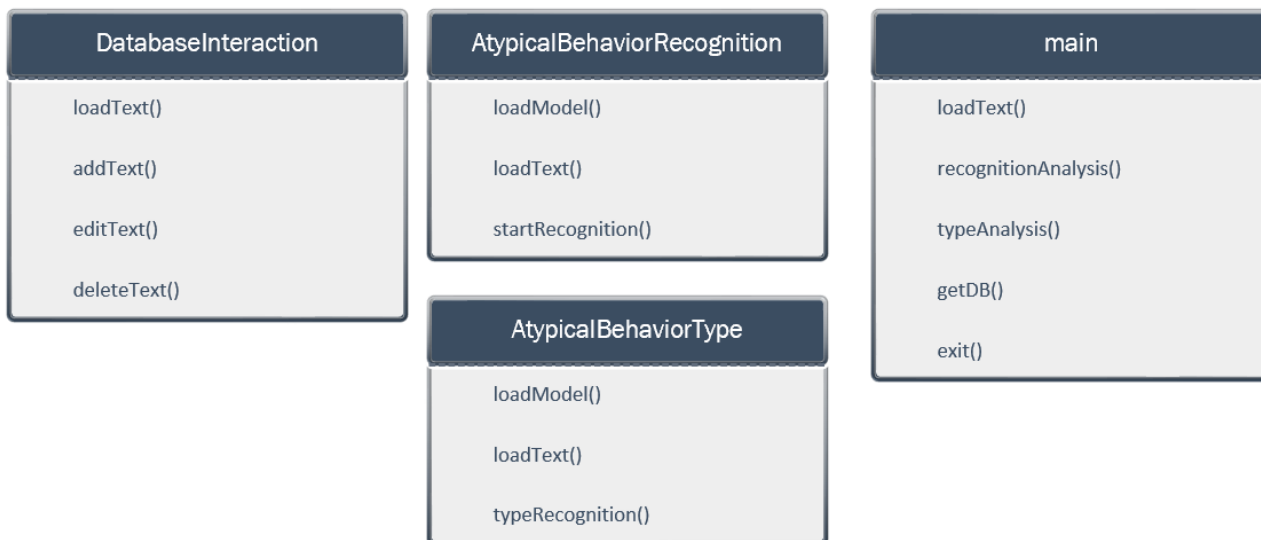
## Додаток А

### Структура бази даних інформаційної системи виявлення нетипової поведінки



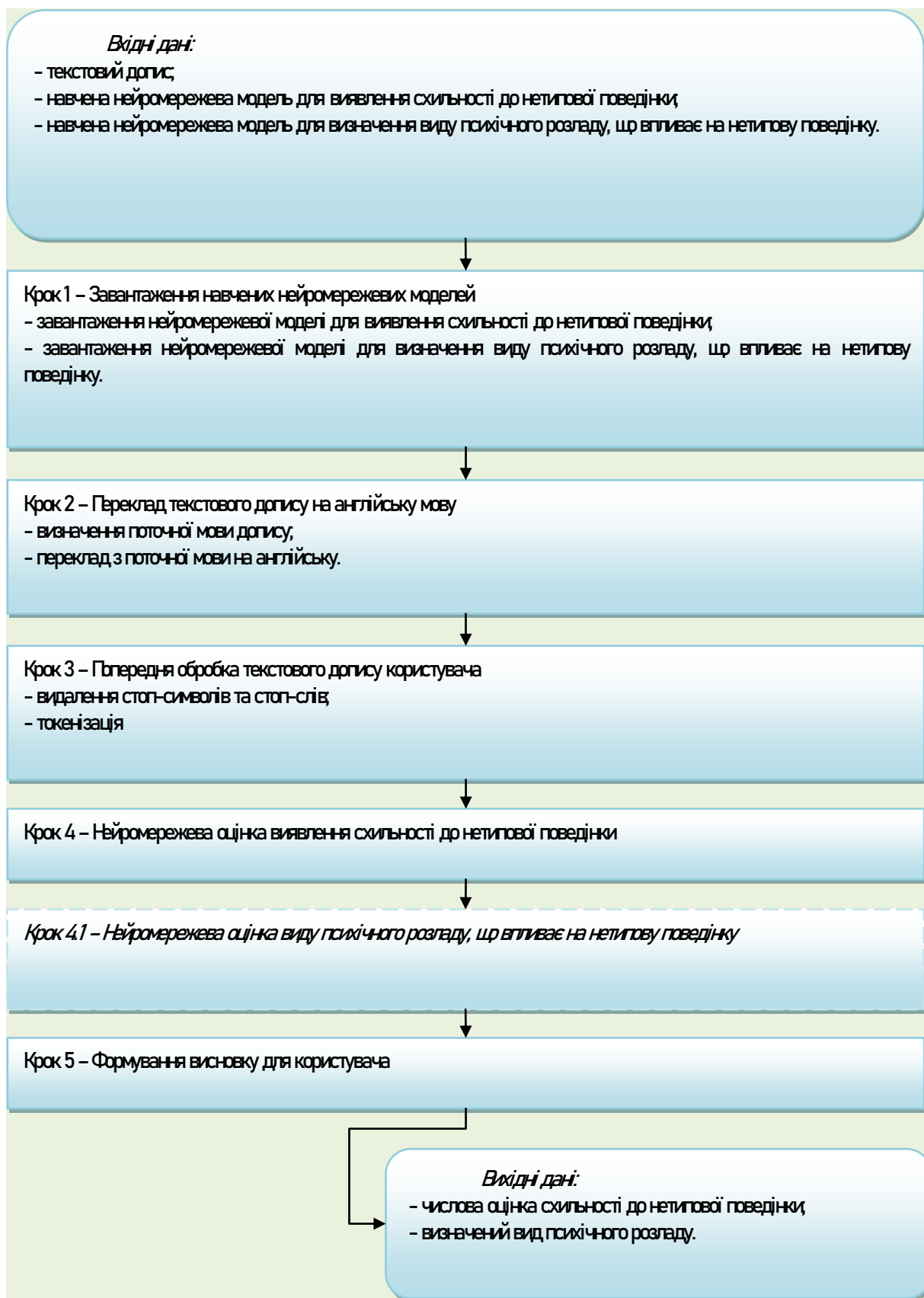
## Додаток Б

### Розгорнута структура класів інформаційної системи автоматизованого виявлення схильності до нетипової поведінки за текстовим дописом



## Додаток В

### Схема методу виявлення схильності до нетипової поведінки



## Додаток Г

### Презентаційний матеріал

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

# МЕТОД ВИЯВЛЕННЯ СХИЛЬНОСТІ ДО НЕТИПОВОЇ ПОВЕДІНКИ ЗА ТЕКСТОВИМ ДОПИСОМ КОРИСТУВАЧА НЕЙРОМЕРЕЖЕВИМИ ЗАСОБАМИ



**Виконав:**  
*студент групи КНС-21-1*  
**Ілля ЧЕРВОНЧУК**



**Керівник:**  
*викладач кафедри КН*  
**Олена СОБКО**

## Актуальність

З ростом популярності соціальних мереж люди віддають більше особистої інформації в онлайн-середовищі. Це створює унікальну можливість для застосування нейромережових методів для аналізу текстового контенту і виявлення потенційно нетипової поведінки.

На ряду з цим зростає інтерес до психологічних аспектів індивідуальної поведінки та її впливу на суспільство. Виявлення нетипової поведінки може допомогти в ранньому виявленні можливих проблем, таких як депресія, самогубство, насильницька поведінка тощо.

Компанії, що надають онлайн-сервіси, все більше зацікавлені в застосуванні таких методів для забезпечення безпеки своїх користувачів, виявлення шахраїв, боротьби з кіберзлочинністю та забезпечення загальної благополучності у віртуальному просторі.

Нейромережові методи стають все більш потужними і ефективними в аналізі тексту та виявленні складних закономірностей. Це відкриває нові можливості для створення точних моделей для виявлення нетипової поведінки на основі аналізу текстового контенту.

## Мета і задачі роботи

**Метою кваліфікаційної роботи бакалавра покращення виявлення схильності до нетипової поведінки за текстовим дописом користувача за рахунок його аналізу нейронмеревевими засобами.**

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

- виконати аналіз інформаційних моделей області виявлення схильності до нетипової поведінки;
- виконати огляд теоретичних підходів, а також обрати підхід для нейронмережевого виявлення схильності до нетипової поведінки;
- виконати аналіз існуючих публікацій за напрямком дослідження;
- провести аналіз існуючого програмного забезпечення області виявлення схильності до нетипової поведінки за текстовим дописом користувача;
- створити метод виявлення схильності до нетипової поведінки за текстовим дописом користувача;
- описати інформаційну структуру системи для виявлення схильності до нетипової поведінки за текстовим дописом користувача;
- спроектувати базу даних;
- обрати набір даних для навчання нейронмережі;
- створити відповідну програмну реалізацію на основі створеного методу;
- виконати тестування створеного програмного забезпечення;
- виконати дослідження ефективності створеного методу виявлення схильності до нетипової поведінки з використанням розробленого ПЗ

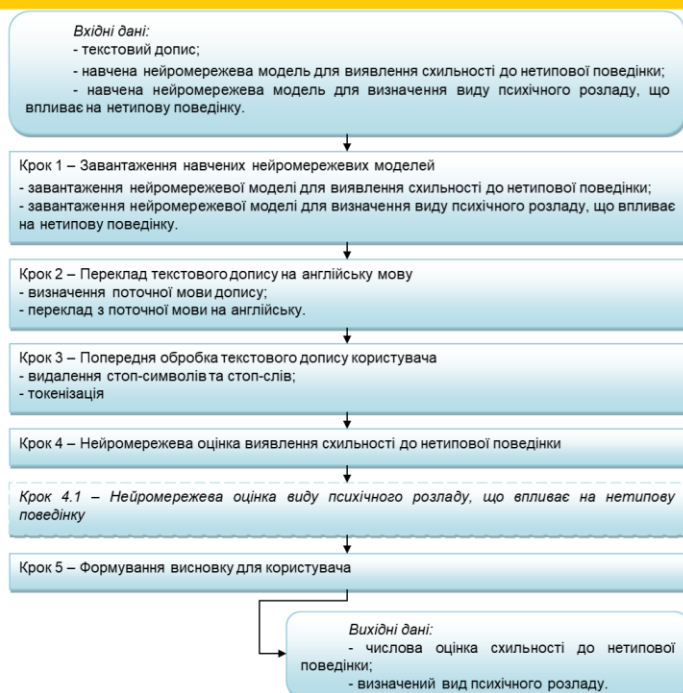
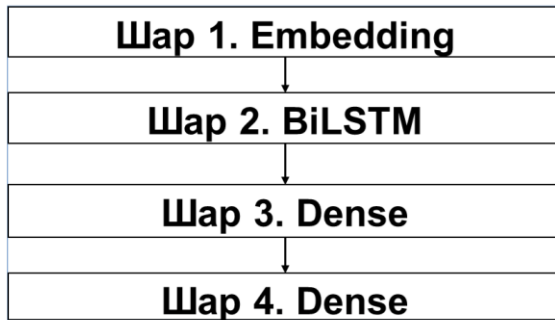
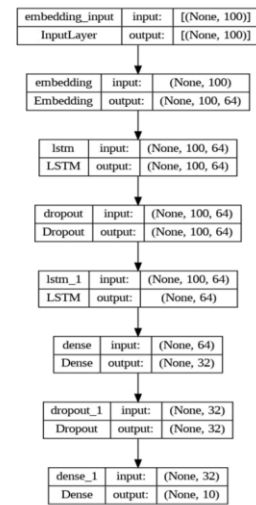


Схема методу  
виявлення  
схильності до  
нетипової  
поведінки за  
текстовим  
дописом

## Розробка архітектури нейронної мережі

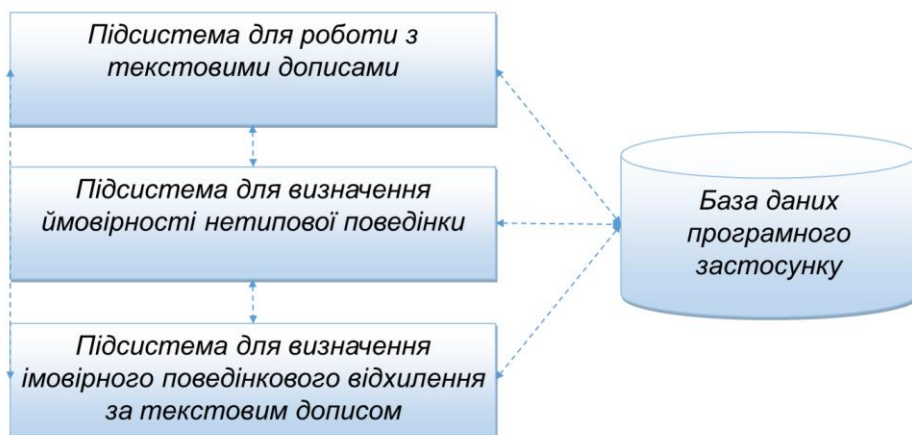


Архітектура RNN-моделі для виявлення схильності до нетипової поведінки

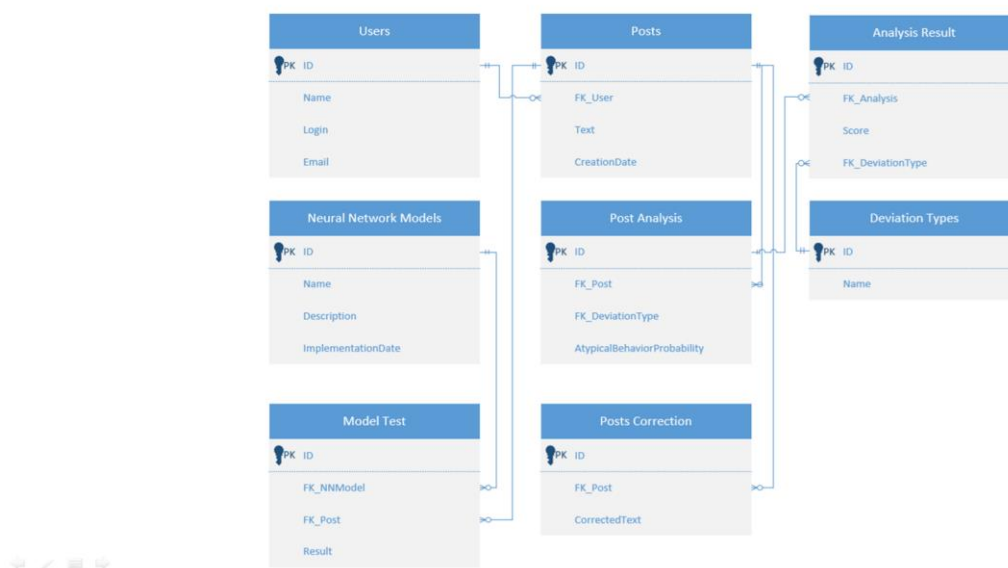


Архітектура RNN-моделі для визначення виду психічного розладу

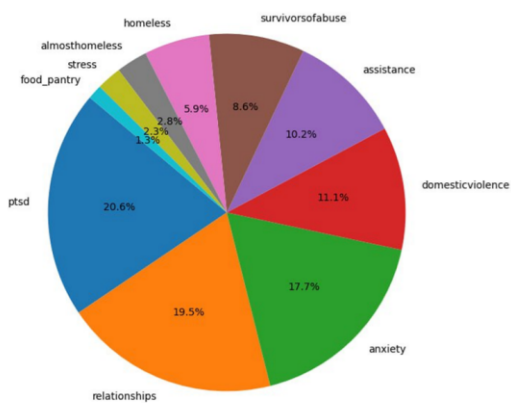
## Схема взаємодії підсистем інтелектуальної системи



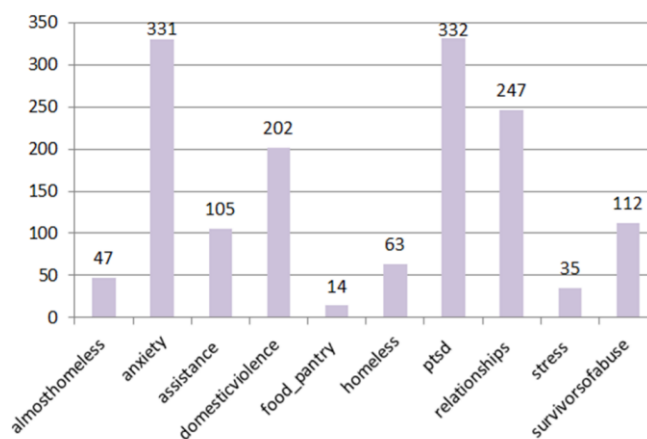
## Даталогічна модель бази даних інтелектуальної системи



## Набір даних дослідження

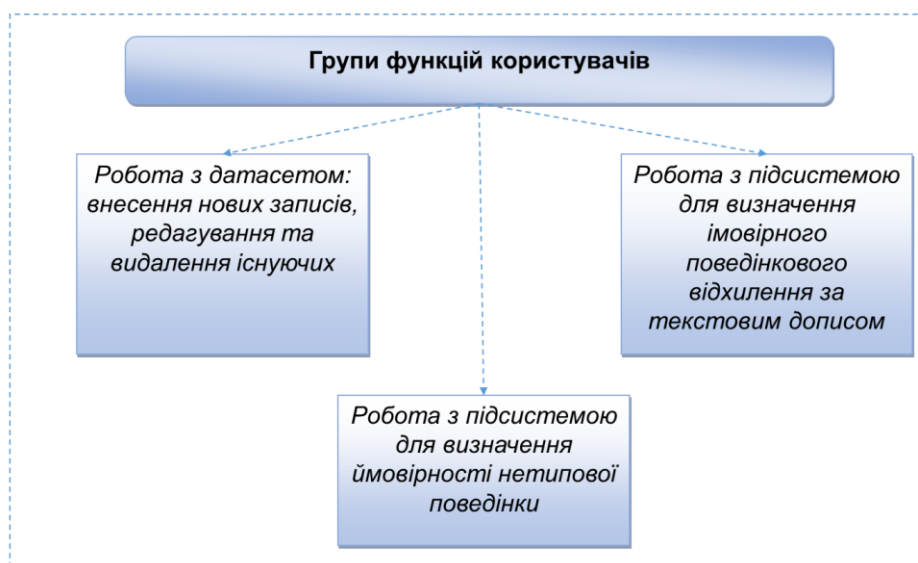


Розподіл записів за видами психічних розладів

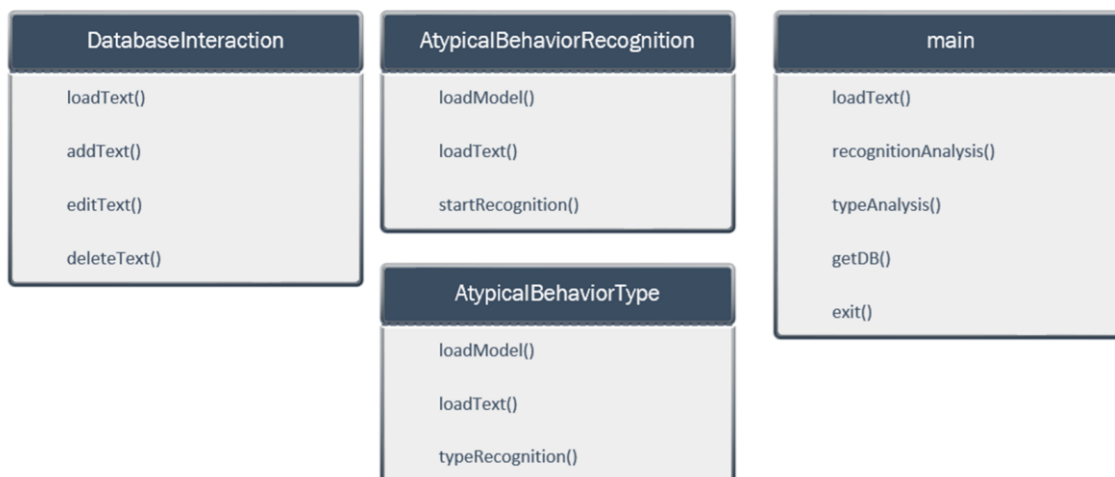


Кількісний розподіл дописів, що містять прояви психологічних розладів

## Групи функцій користувача застосунку



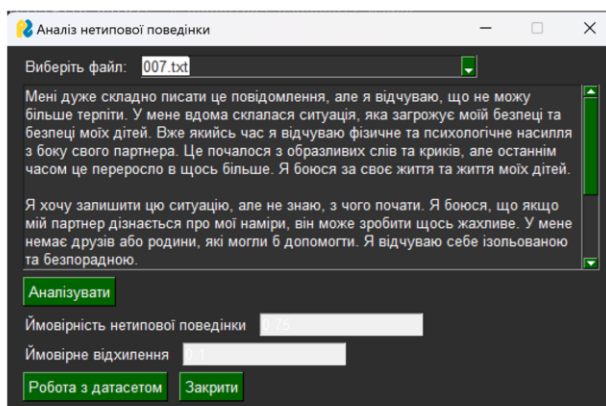
## Діаграма класів програмного застосунку



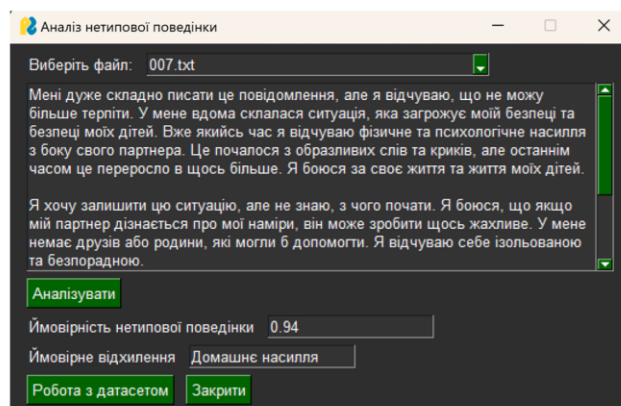


## Алгоритм SMOTE для балансування даних

## Інтерфейс інформаційної системи



Інтерфейс головного вікна застосунку



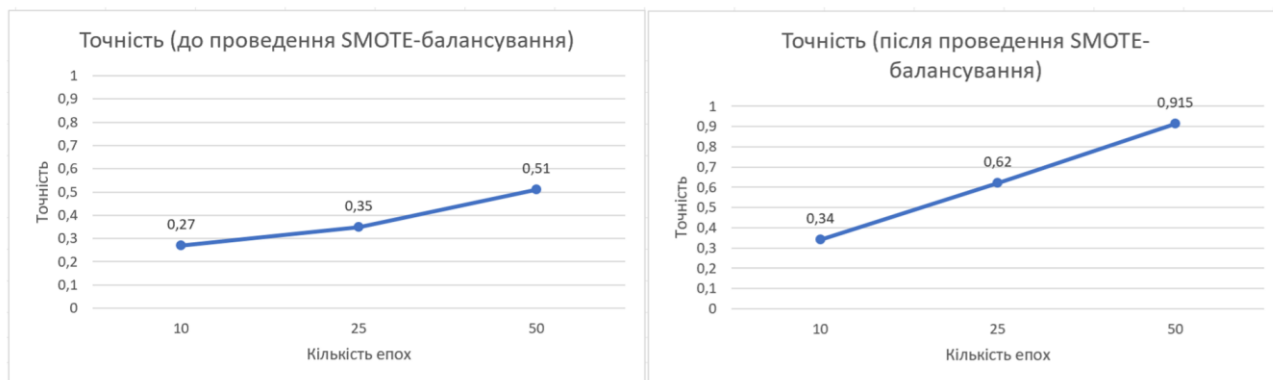
Визначення ймовірності нетипової поведінки та типу ймовірного відхилення

## Результати досліджень

Для валідації програмного продукту на базі методу виявлення схильності до нетипової поведінки за текстовими дописами необхідно провести детальне дослідження нейромережевої моделі та її показників за умови налаштування різних параметрів.

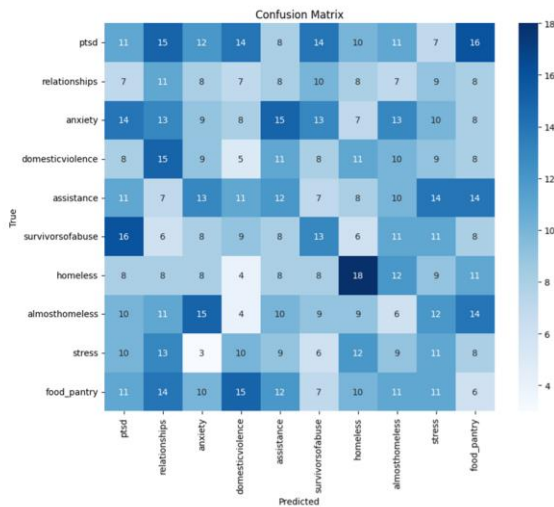
До проведення SMOTE-балансування			
Кількість епох	Точність	Функція втрат	Час навчання, секунд
10	0,27	2,19	80
25	0,35	1,93	170
50	0,51	1,76	350
Після проведення SMOTE-балансування			
10	0,34	0,81	80
25	0,62	0,55	154
50	0,915	0,3221	280

## Результати досліджень

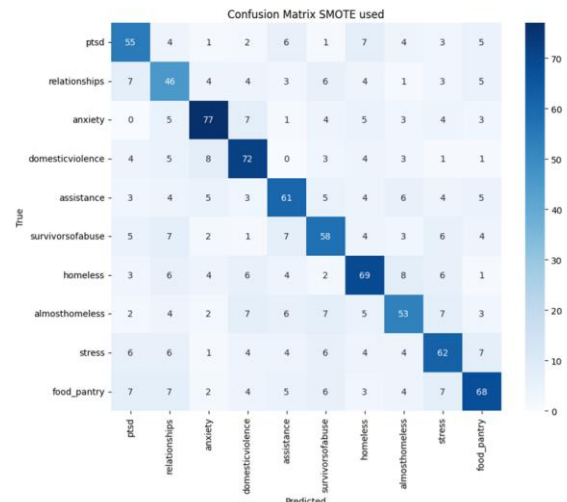


Графіки змін значень точності моделі

## Результати досліджень



Матриця сплутувань до SMOTE-балансування датасету



Матриця сплутувань після SMOTE-балансування датасету

## Висновки

Метою кваліфікаційної роботи бакалавра було покращення виявлення схильності до нетипової поведінки за текстовим дописом користувача за рахунок його аналізу нейромережевими засобами. Для покращення виявлення схильності до нетипової поведінки за текстовим дописом було створено програмне забезпечення, а також були поставлені та вирішені такі завдання:

1. Виконано аналіз інформаційних моделей області виявлення схильності до нетипової поведінки.
2. Виконано огляд теоретичних підходів, а також обрано підхід для нейромережевого виявлення схильності до нетипової поведінки на основі рекурентних нейронних мереж, оскільки він дає вищу точність та можливість розуміння природної мови.
3. Виконано аналіз існуючих публікацій за напрямком дослідження, а також проведено аналіз існуючого програмного забезпечення області виявлення схильності до нетипової поведінки за текстовим дописом користувача.
4. Створено метод виявлення схильності до нетипової поведінки за текстовим дописом користувача.
5. Описано архітектуру системи та взаємозв'язок її компонентів для виявлення схильності до нетипової поведінки за текстовим дописом користувача.
6. Спроектовано базу даних.
7. Обрано набір даних для навчання нейромережі.
8. Визначено шляхи дослідження та засоби створення програмного застосування.
9. Створено відповідну програмну реалізацію на основі розробленого методу.
10. Виконано тестування створеного програмного забезпечення, яке показало коректність описаних функцій.
11. Проведено дослідження ефективності створеного методу виявлення схильності до нетипової поведінки з використанням розробленого ПЗ.

Ім'я користувача:  
Кафедра КН

Дата перевірки:  
21.06.2024 07:43:13 EEST

Дата звіту:  
21.06.2024 13:25:32 EEST

ID перевірки:  
1016379891

Тип перевірки:  
Doc vs Internet + Library

ID користувача:  
100005671

Назва документа: КНС-21-1 Червончук\_ЗАПИСКА

Кількість сторінок: 69 Кількість слів: 12406 Кількість символів: 100485 Розмір файлу: 1.58 MB ID файлу: 1016189016

Виявлено модифікації тексту (можуть впливати на відсоток схожості)

## 10.3% Схожість

Найбільша схожість: 4.26% з джерелом з Бібліотеки (ID файлу: 1016172604)

5.88% Джерела з Інтернету

572

Сторінка 71

7.5% Джерела з Бібліотеки

116

Сторінка 75

## 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

11  
сторінок

# Anti-Plagiarism v-15.257

Максимальне співпадіння з одним документом 6.0%

Словники перевірки: en\_US, ru\_RU, ua\_UA. Помилки в документах: 13%

ID: 132033 Назва: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами Додано в БД: 2024-06-21 Автора: Ілля ЧЕРВОНЧУК Керівники: Олена СОБКО Консультанти: Опоненти:	Документ		Сумарний збіг по Базі Даних	
	Символи	Лексеми	Символи	Лексеми
	120045	1016	9757 (8%)	37 (4%)

## 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 схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами

Автор: студент групи КНс-21-1 Ілля Сергійович ЧЕРВОНЧУК

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.	<b>відповідає</b>
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:

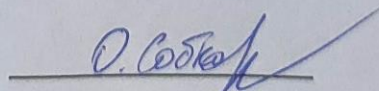
Запозичення, виявлені в роботі Червончука Іллі, не є плагіатом, оскільки: запозичення розміщені в розділі огляду існуючих підходів, не описують безпосередньо авторську роботу і не стосуються її результатів; усі запозичення фрагментарні; до запозичень входять фрагменти, що не мають авторства і містять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

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

- за системою Anti-Plagiarism: 6%;

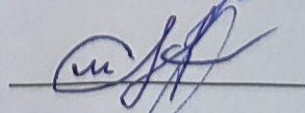
- за системою Unichек: 10.3 %

Керівник роботи



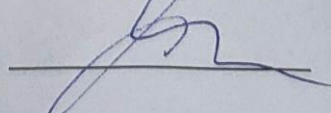
Олена СОБКО

Гарант ОП



Олександр МАЗУРЕЦЬ

Завідувач кафедри КН



Олександр БАРМАК



**ВІДГУК НАУКОВОГО КЕРІВНИКА  
на кваліфікаційну роботу бакалавра**

студента гр. КНс-21-1 Червончука Іллі Сергійовича

за темою Метод виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами

**1. Актуальність теми**

Нейромережеві методи у текстовому аналізі стають все потужнішими і ефективними, що відкриває нові можливості для розробки точних моделей виявлення нетипової поведінки через аналіз текстового контенту. Зростає інтерес до психологічних аспектів індивідуальної поведінки та її впливу на суспільство, де виявлення нетипової поведінки може допомогти ранньому виявленню можливих проблем, таких як депресія, самогубство, насильницька поведінка та інші. Тому, тема виявлення виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами набуває все більшої актуальністю.

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

За стандартом, а саме описом предметної області, об'єктами вивчення та діяльності є процес виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами. Метою кваліфікаційної роботи бакалавра є покращення виявлення схильності до нетипової поведінки за текстовим дописом користувача за рахунок його аналізу нейромережевими засобами. Мета була досягнута шляхом розробки методу виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами. При вирішенні поставленої задачі використано математичні моделі, методи та алгоритми розв'язання теоретичних і прикладних задач, що виникають при розробці інформаційних технологій. Тому результати виконання кваліфікаційної роботи бакалавра відповідають стандарту бакалавра спеціальності 122 – Комп'ютерні науки.

**3. Професійні та особистісні якості бакалавра**

Під час написання своєї кваліфікаційної роботи бакалавра Червончук Ілля Сергійович продемонстрував високу кваліфікацію та відповідальність як студент. Він здійснив систематичний і якісний підхід до всіх етапів дослідження, виконуючи їх вчасно.

#### **4. Ступінь самостійності під час виконання кваліфікаційної роботи**

Одержані в роботі результати є наслідком особистої діяльності студента, який самостійно виконував всі поставлені задачі.

#### **5. Ступінь оволодіння методами дослідження**

При реалізації кваліфікаційної роботи показав достатній рівень компетентностей та володіння необхідними інструментами та обладнанням, методами, методиками та технологіями предметної області комп'ютерних наук.

#### **6. Повнота та якість розкриття теми роботи**

Тема дослідження була обґрунтована і детально розкрита, що свідчить про глибоке розуміння автором контексту та значущості обраної проблеми. В роботі було проведено аналіз актуальності теми на основі вивчення наявних досліджень і підходів, що підтверджує важливість проведення такого дослідження в сучасних умовах. Поставлені завдання були чітко сформульовані і досягнуті в процесі проведення дослідження.

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

Структура роботи та послідовність викладення логічні та відповідають поставленій меті. Викладення матеріалу послідовне, аргументоване, літературно грамотне.

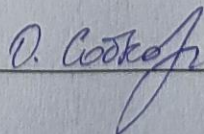
#### **8. Можливість практичного застосування кваліфікаційної роботи бакалавра, окремих її частин**

Метод виявлення схильності до нетипової поведінки за текстовими дописами користувачів за допомогою нейромережеских засобів може бути застосований у соціальних мережах, онлайн-форумах, чат-платформах та інших веб-сервісах для автоматичного виявлення ознак нетипової поведінки або інших потенційно небезпечних витоків, що дозволяє оперативно реагувати та надавати підтримку користувачам у вразливому стані.

#### **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 «добре».

Керівник \_\_\_\_\_



викладач каф. КН Олена СОБКО



## РЕЦЕНЗІЯ

### на кваліфікаційну роботу бакалавра

студента гр. КНс-21-1 Червончука Іллі Сергійовича

за темою: Метод виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами

#### 1. Актуальність обраної теми

Нейромережеві методи у текстовому аналізі зростають у потужності і ефективності, що відкриває нові можливості для розробки точних моделей виявлення нетипової поведінки через аналіз текстового контенту. Інтерес до психологічних аспектів індивідуальної поведінки та її впливу на суспільство зростає, оскільки виявлення нетипової поведінки може сприяти ранньому виявленню потенційних проблем, таких як депресія, самогубство, насильницька поведінка та інші. Тому, тема виявлення схильності до нетипової поведінки за текстовими дописами користувачів за допомогою нейромережевих методів стає все більш актуальною.

#### 2. Повнота розкриття мети та завдань роботи

В даній кваліфікаційній роботі бакалавра мета та завдання були повністю розкриті. Автор докладно описав процес розробки та програмну реалізацію методу виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами. Робота містить достатній рівень деталізації та використовує науковий підхід, що дозволяє чітко зрозуміти її суть та важливість.

#### 3. Зміст кожного розділу роботи

Кожен розділ бакалаврської роботи містить інформацію, що стосується теми, від теоретичних аспектів виявлення схильності до нетипової поведінки за текстовим дописом користувача нейромережевими засобами до практичної реалізації. Чітка організація та логічний зв'язок між розділами сприяють зрозумінню послідовності проведення дослідження та втілення проєкту.

4. Оцінка розробленої інформаційної системи, її практична цінність

Розроблена інформаційна система для виявлення схильності до нетипової поведінки за текстовими дописами користувачів з використанням нейромережесих методів показує високі результати та має значні перспективи застосування і практичну цінність. Реалізоване програмне забезпечення продемонструвало достатньо високу точність та швидкість у виявленні схильності до нетипової поведінки за текстовими дописами користувачів.

5. Якість оформлення кваліфікаційної роботи бакалавра

Оформлення бакалаврської роботи виконане належно і включає всі потрібні елементи, такі як розділи, таблиці, графіки, а також посилання на джерела. Робота має чітку структуру, дотримується наукового стилю та правильну організацію тексту, що забезпечує легке сприйняття і точну оцінку.

6. Недоліки кваліфікаційної роботи бакалавра

Кваліфікаційна робота бакалавра демонструє достатній потенціал та практичне досягнення, однак можна відзначити деякі недоліки. Для покращення якості роботи було б доцільно провести дослідження продуктивності нейромережесивої моделі при налаштуванні різних параметрів. Також є ряд орфографічних помилок та неоптимальна компоновка пояснювальної записки для друку, зокрема незаповнені області внизу аркушів.

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 «добре».

Рецензент

К.Г.М., доц. кафедри КііС Міжнародн Д.О

