

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА
Факультет інформаційних технологій
Кафедра інтелектуальних технологій

ВИПУСКНА КВАЛІФІКАЦІЙНА РОБОТА
БАКАЛАВРА
НА ТЕМУ:

Програмний модуль визначення емоційного забарвлення
постів в соціальних мережах

Галузь знань 12 «Інформаційні технології»

Спеціальність 122 «Комп'ютерні науки»

Освітня програма «Комп'ютерні науки»

Освітній рівень: бакалавр

Виконав: студент 4 курсу, групи КН-41

Парнет А.О.

(прізвище та ініціали)

Керівник Федусенко О.В.

(прізвище та ініціали)

(науковий ступінь, звання)

Випускна кваліфікаційна робота бакалавра допущена до захисту
рішенням кафедри *інтелектуальних технологій*

Протокол № 13 від 05.06.2023 р.

зав. кафедри _____ доц. Іларіонов О.Є.

Київ – 2023

ЗМІСТ

Вступ	3
1. Аналіз процесу визначення емоційного забарвлення постів у соціальних мережах	4
1.1 Аналітичний огляд літератури за темою дослідження визначення емоційного забарвлення тексту	4
1.2. Аналіз існуючих інформаційних систем для визначення емоційного забарвлення тексту	7
1.3. Аналіз основних процесів предметного дослідження	10
1.4. Постановка задачі на розробку визначення емоційного забарвлення постів у соціальних мережах	16
1.5. Висновки	18
2. Розробка архітектури застосунку для визначення емоційного забарвлення постів у соціальних мережах	20
2.1 Розробка архітектури	20
2.1.1 Функціональний аналіз	20
2.1.2 IDEF0 діаграма процесу визначення емоційного забарвлення тексту	25
2.1.3 Архітектура інформаційної системи	29
2.2 Інформаційне забезпечення	30
2.3 Висновки	36
3. Програмне забезпечення для визначення емоційного забарвлення постів у соціальних мережах	38
3.1 Обґрунтування вибору програмних засобів	38
3.2 Структура програмного забезпечення	40
3.3 Керівництво користувача	44
3.4. Оцінка точності роботи	46
3.5 Огляд процесу тестування	48
3.5 Висновки	54
Висновки	56
Список використаних джерел	58

Вступ

Темою даної дипломної роботи є визначення емоційного забарвлення постів у соціальних мережах. Це є дуже цікавою і в той же час складною задачею, адже на даний час існує багато методів вирішення цієї задачі, але жоден з них не може гарантувати стовідсоткової точності знаходження оптимального розв'язку. Об'єктом дослідження даної роботи є соціальні мережі, так як саме там зберігається найбільше інформації, яку необхідно класифікувати та визначити характеристики (позитивний чи негативний зміст має даний текст). Предметом дослідження можна вважати пости в соціальних мережах. Метою дипломної роботи є розробка застосунку для визначення емоційного забарвлення постів у соціальних мережах.

Для досягнення поставленої мети необхідно вирішити наступні задачі: проаналізувати літературу та вже існуючі інформаційні системи для визначення емоційного забарвлення тексту, розробити архітектуру та сформувавши інформаційне забезпечення застосунку. Також в практичній частині необхідно визначитись з використовуваними програмними засобами, створити структуру програмного забезпечення, реалізувати програмний продукт та провести його тестування.

1. Аналіз процесу визначення емоційного забарвлення постів у соціальних мережах

1.1 Аналітичний огляд літератури за темою дослідження визначення емоційного забарвлення тексту

Питання визначення емоційного забарвлення тексту не є новим і ним вже займалися вчені з різних країн. Тож розглянемо, які варіанти існують на даний час.

Robert A. Stine в статті «Sentiment Analysis»[5] дуже детально описує як задача аналізу настроїв вирішується за допомогою нейронних мереж та глибокого навчання. Розглядається, що типова побудова нейронної мережі буде складатися з трьох етапів, що накладаються один одного. На першому етапі обирається спосіб представлення тексту у векторному просторі. Поширеним варіантом є представлення лексеми кожного типу слів як точки у багатовимірному векторному просторі. Цей спосіб ще називають *embedding words* (вбудовування слів). Коли в тексті існує десятки тисяч різних слів, зберігати для кожного з них власне число значно знизить ефективність моделі. Натомість *embedding layer* (шар вбудовування) дозволяє перетворити кожне слово у вектор фіксованої довжини визначеного розміру. Результуючий вектор є щільним і містить реальні значення. Розмірність такого вектору зазвичай становить 300 або більше.

Другий етап побудови нейронної мережі перетворює вбудований текст у прихований вектор стану. У НН зі зворотним зв'язком, яка ігнорує послідовність розташування тексту, прихований стан є відносно простою функцією. Більшість слів в ізоляції мають нейтральний настрій. Лише при об'єднанні слів у фрази більш чітко проявляється емоційне забарвлення. Для того, щоб НН відображала послідовну структуру тексту, були досліджені різноманітні представлення. Прихований стан тепер повинен враховувати вплив попередніх або оточуючих слів на поточне слово.

Третій і останній етап побудови є більш простим. Параметри прихованого стану адаптуються до поставленої задачі шляхом налаштування оцінок моделі з міченими навчальними даними. По суті, мережа другого етапу виробляє приховане

представлення тексту документа, що підлягає класифікації. Цей прихований стан потім подається в класифікатор настроїв, часто логістичну регресію. Вся модель навчається для максимізації відповідної логістичної ймовірності. Нейронна мережа відстежує прогрес навчання і демонструє загальний прогноз до даних, необхідних для глибокого навчання.

Також Bing Liu у своїх дослідженнях «Sentiment Analysis and Opinion Mining»[2] пропонує розглянути задачу класифікації емоційного забарвлення тексту зі сторони керованого та некерованого навчання (також на основі нейронної мережі). Так як перший метод є дуже схожим до того, який було розглянуто вище, тому перейдемо до алгоритму некерованого навчання. Він був наведений у роботі Turney[6] та складається з трьох кроків.

Крок 1. Два слова, що йдуть підряд, застосовуються для навчання, якщо їх POS (part-of-speech) теги відповідають одному з шаблонів, наведених у таблиці 1.1. Наприклад, шаблон 2 означає, що обираються два слова, що йдуть підряд якщо перше слово є прислівником, друге слово є прикметником, а третє слово (яке не береться) не є іменником. Як приклад, у реченні "Це піаніно видає прекрасні звуки", "прекрасні звуки" виділяється, оскільки воно задовольняє першому шаблону. Причиною використання цих шаблонів є те, що слова на JJ, RB, RBR та RBS часто виражають думку. Іменники або дієслова виступають в якості контекстів, оскільки в різних контекстах слова JJ, RB, RBR та RBS можуть виражати різні почуття. Наприклад, прикметник "непередбачуваний" може мати негативне значення в огляді автомобіля як "непередбачуване керування", але може мати позитивне значення в огляді фільму як "непередбачуваний сюжет".

Таблиця 1.1. – Шаблиони POS-тегів для вилучення двослівних фраз

	Перше слово	Друге слово	Третє слово (не витягується)
1	JJ	NN or NNS	anything
2	RB, RBR, or RBS	JJ	not NN nor NNS
3	JJ	JJ	not NN nor NNS

4	NN or NNS	JJ	not NN nor NNS
5	RB, RBR or RBS	VB, VBD, VBN, or VBG	anything

JJ – adjectives (прикметники)

RB – adverbs (прислівники)

RBR – comparative adverbs (порівняльні прислівники)

RBS – superlative adverbs (вищий ступінь порівняння прислівників)

NN – common nouns (загальні іменники)

NNS – common noun plural form (загальна іменники у множині)

VB – verb base forms (початкові форми дієслова)

VBD – verb past tense forms (дієслова минулого часу)

VBN – verb past participles (дієприкметники минулого часу)

VBG – verb gerund or present participle (герундій або дієприкметник теперішнього часу)

Крок 2. Оцінюється орієнтація настрою (SO) витягнутих фраз за допомогою міри точкової взаємної інформації (PMI). PMI вимірює ступінь статистичної залежності між двома термінами. Сентиментальна орієнтація (SO) фрази обчислюється на основі її асоціації з позитивним опорним словом "відмінно" та негативним опорним словом "погано". Ймовірності розраховуються шляхом видачі запитів до пошукової системи та збору кількості переглядів. На кожен пошуковий запит пошукова система зазвичай видає кількість релевантних документів до запиту, що і є кількістю збігів. Таким чином, здійснюючи пошук за двома термінами разом і окремо, можна оцінити ймовірності в рівнянні.

Крок 3. Отримавши текст, для якого необхідно визначити емоційне забарвлення, алгоритм обчислює середній SO всіх фраз у тексті і класифікує його як позитивний, якщо середній показник позитивний, і як негативний в іншому випадку.

Остаточна точність даного методу класифікації для текстів коливається від 84% до 66%.

1.2. Аналіз існуючих інформаційних систем для визначення емоційного забарвлення тексту

Звичайно, в наш час вже напевно не існує питання, яке ще не намагались вирішити. Тому було розроблено не один застосунок, який допомагає з аналізом емоційного забарвлення тексту. Розглянемо найбільш популярні серед них[11].

1. MonkeyLearn

Створена в першу чергу для бізнесу. Має такі можливості

- аналіз індексу підтримки споживача (аналіз коментарів промоутерів, їх настроїв та покращення того, що не підходить)
- аналіз відгуків (допомагає визначити сильні та слабкі сторони кожного окремого бізнес-підрозділу та функцій продукту)
- аналіз задоволення споживача (визначає негативний досвід роботи з продуктами та процесами, завдяки вирішенню яких можна підвищити рівень задоволеності клієнтів)
- аналіз підтримки (аналізує запити до служби підтримки, скорочує проблеми, які повторюються, розділяє їх по темах, що значно полегшує вирішення, відслідковування та оцінювання проблеми)
- аналіз зворотнього зв'язку (дає доступ до очищених, систематизованих та візуалізованих даних для будь-якого виду спілкування з клієнтами).

Загалом цей додаток спрощує очищення, маркування та візуалізацію відгуків клієнтів[16].

2. Lexalytics

Дає змогу налаштувати інструменти аналізу тональності в хмарі, призначене для фахівців з обробки та аналізу даних і архітекторів, яким потрібен повний доступ до технології або яким необхідне локальне розгортання з міркувань безпеки. Містить в собі доступ до потужної системи аналізу настроїв NLP платформи, яку можна налаштувати відповідно до конкретних потреб. Після того, як система виміряє почуття та емоції, що стоять за відгуками клієнтів, їх можна візуалізувати (і зберегти) результати в комплексному інструменті бізнес-аналітики. Хоча

Lexalytics забезпечує як аналіз тексту, так і візуалізацію даних, для його використання будуть потрібні технічні знання[14].

3. Brandwatch

Brandwatch був названий лідером у рейтингу Forrester Wave 2020 для платформ соціального прослуховування, а точність і достовірність аналізу настроїв платформи була визнана вищою, ніж у інших постачальників, у порівнянні з іншими платформами.

Компанії використовують Brandwatch, щоб відстежувати згадки в Інтернеті та розуміти голос клієнта, виявляти коливання настроїв і вимірювати впізнаваність бренду в режимі реального часу, 24 години на добу, 7 днів на тиждень. Платформа, яка називає себе "споживчим інтелектом", також надає користувацькі класифікатори машинного навчання, а також автоматичне розмежування об'єктів для виявлення тем і сюжетів у даних, які вже були проаналізовані на основі настроїв.

Як додатковий бонус, Brandwatch включає унікальну функцію "розуміння зображень", яка дозволяє вийти за межі текстових даних, щоб користувачі могли відстежувати, як їх логотипи поширюються в мільйонах онлайн-ресурсів і де вони з'являються[12].

4. Social Searcher

Social Searcher відстежує ключові слова, хештеги та імена користувачів на всіх платформах соціальних мереж. Вони отримують безліч соціальних аналітичних даних, включаючи інформацію про аудиторію, популярні хештеги та соціальних інфлюенсерів. Безкоштовна версія включає в себе інструмент аналізу настроїв, який надає загальний настрій даних соціальних мереж на кожній платформі і розбивку популярних постів, які були класифіковані як негативні і позитивні[19].

5. MeaningCloud

Виконує багатомовний аналіз настроїв за допомогою API. Цей онлайн-інструмент виконує аналіз настроїв на основі аспектів, щоб визначити, чи згадуються певні теми в позитивному, негативному або нейтральному ключі. Крім

того, користувач може власноруч визначити словник, щоб включити будь-яку специфічну лексику, яку може використовувати у власній галузі. Деякі з найкращих функцій - це виявлення глобальних настроїв (загальне уявлення про те, що клієнт висловив у певному тексті), ідентифікація думки та факту, а також виявлення настроїв у кожному реченні тексту[15].

Отже для того щоб порівняти та проаналізувати вже існуючі застосунки, було визначено такі необхідні характеристики:

- простота використання
- адаптивність
- оновлюваний словник з даними
- орієнтованість на середньостатистичного користувача
- можливість отримання інформації за заданими параметрами
- безкоштовне використання

Таблиця 1.2 – Порівняльний аналіз існуючих рішень

	Monkey Learn	Lexalytics	Brandwatch	Social Searcher	MeaningCloud
Простота використання	+	-	+	+	+
Адаптивність	+	-	+	+	-
Оновлюваний словник з даними	+	+	-	+	-
Орієнтованість на середньостатистичного користувача	-	-	-	-	-

Можливість отримання інформації за заданими параметрами	+	+	+	-	-
Безкоштовне використання	-	-	-	-	-

Дійсно, всі розглянуті вище застосунки допомагають з аналізом тексту у соціальних мережах, але більшість з них орієнтовані на компанії, які хочуть створити клієнтоорієнтований досвід та вдосконалити маркетингову стратегію свого бренду. Вони не розраховані для звичайних людей, які просто хочуть відфільтрувати інформацію, яку вони отримують. І це можна зрозуміти, так як бізнес, якому потрібен даний застосунок, готовий заплатити більше, ніж середньостатистичний користувач соціальних мереж.

Додаток, який планується розробити в даній дипломній роботі, буде розрахований на більшу частину людства, тобто саме звичайних користувачів. Адже його головною задачею є не отримання прибутку з даної розробки, а допомога у вирішенні проблеми фільтрації отриманої інформації з соціальних мереж.

1.3. Аналіз основних процесів предметного дослідження

Для аналізу процесу визначення емоційного забарвлення тексту використовуватиметься методологія IDEF0 та DFD.

Для моделювання будемо використовувати CASE засіб AllFusion Process Modeler.

Розглянемо основні межі обстеження:

1. Функціональні. Для цього виділимо основні функції процесу «Визначення емоційного забарвлення постів у соціальних мережах»:

- структуризація даних

- робота з фільтрами
- отримання статистики

Предметна область – аналіз процесу визначення емоційного забарвлення постів у соціальних мережах.

Розглянемо схему моделі «Визначення емоційного забарвлення постів у соціальних мережах» на рисунку 1.1.

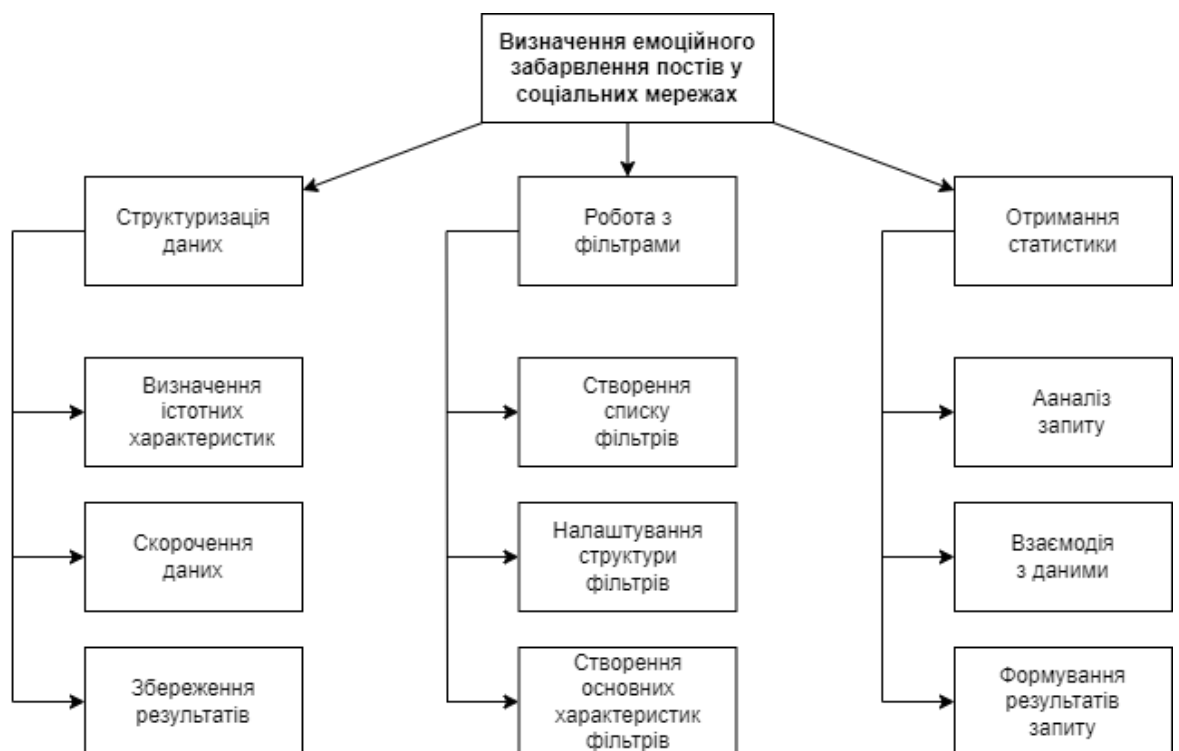


Рисунок 1.1 – Схема процесів " Визначення емоційного забарвлення постів у соціальних мережах"

Розглянемо контекстну діаграму процесу визначення емоційного забарвлення тексту ЯК-Є (рис. 1.2).

Основні зв'язки контекстної діаграми наведені в таблиці 1.3.

Таблиця 1.3 – Опис зв'язків контекстної діаграми ЯК-Є

Назва	Тип	Зовнішній об'єкт
Дані для аналізу	Вхідна інформація	Соціальні мережі
Вимоги з фільтрації	Вхідна інформація	Користувач
Статистика	Вихідна інформація	Користувач

Розглянемо декомпозицію контекстної діаграми (рис. 1.3). Даний бізнес-процес складається з таких бізнес-процесів:

1. Аналіз наявних даних

Вхідна інформація:

- Дані для аналізу

Вихідна інформація:

- Структуровані дані

2. Аналіз умов фільтрації

Вхідна інформація:

- Структуровані дані
- Фільтри

Вихідна інформація:

- Структурований список фільтрів

3. Отримання статистики

Вхідна інформація:

- Структуровані дані
- Структурований список фільтрів

Вихідна інформація:

- Статистика

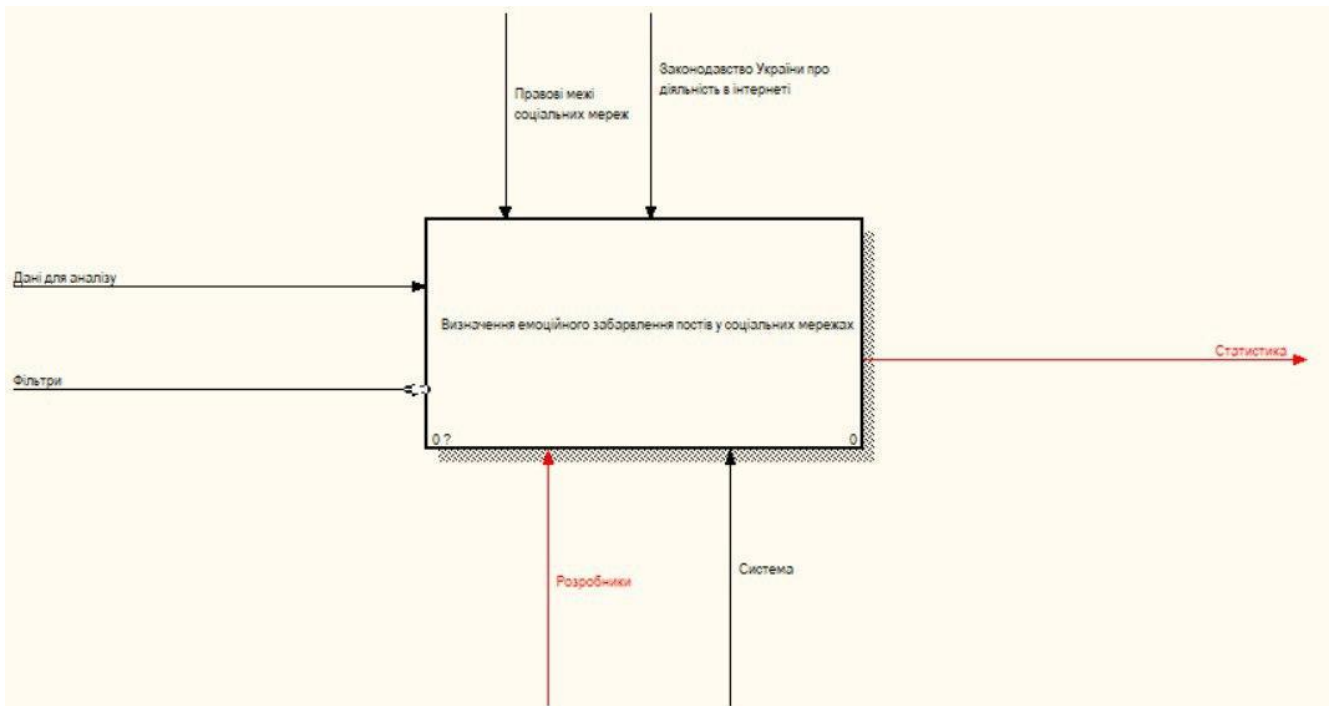


Рисунок 1.2 – Контекстна діаграма процесу "Визначення емоційного забарвлення постів у соціальних мережах"

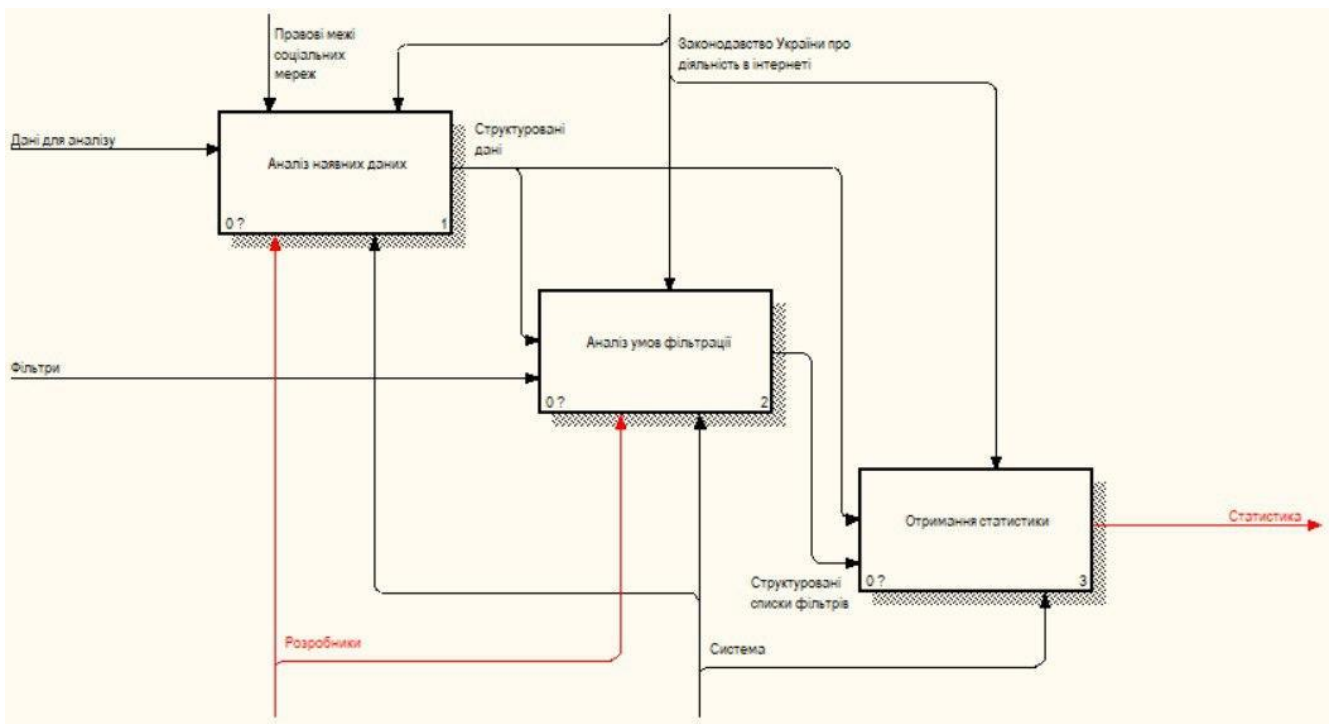


Рисунок 1.3 – Декомпозиція контекстної діаграми процесу

"Визначення емоційного забарвлення постів у соціальних мережах"

Розглянемо докладніше кожен процес у вигляді DFD діаграми (рисунок 1.4). Процес аналізу наявних даних здійснюється в такий спосіб. Спочатку система отримує дані у звичайному текстовому вигляді. Далі їй необхідно переформатувати

ці дані у читабельний для неї вигляд. Вже переформатовані дані необхідно скоротити (прибрати повтори, зайвий текст, який не несе у собі корисну інформацію). І далі скорочені дані треба структурувати та передати для подальшої роботи (без збереження для повторного використання).

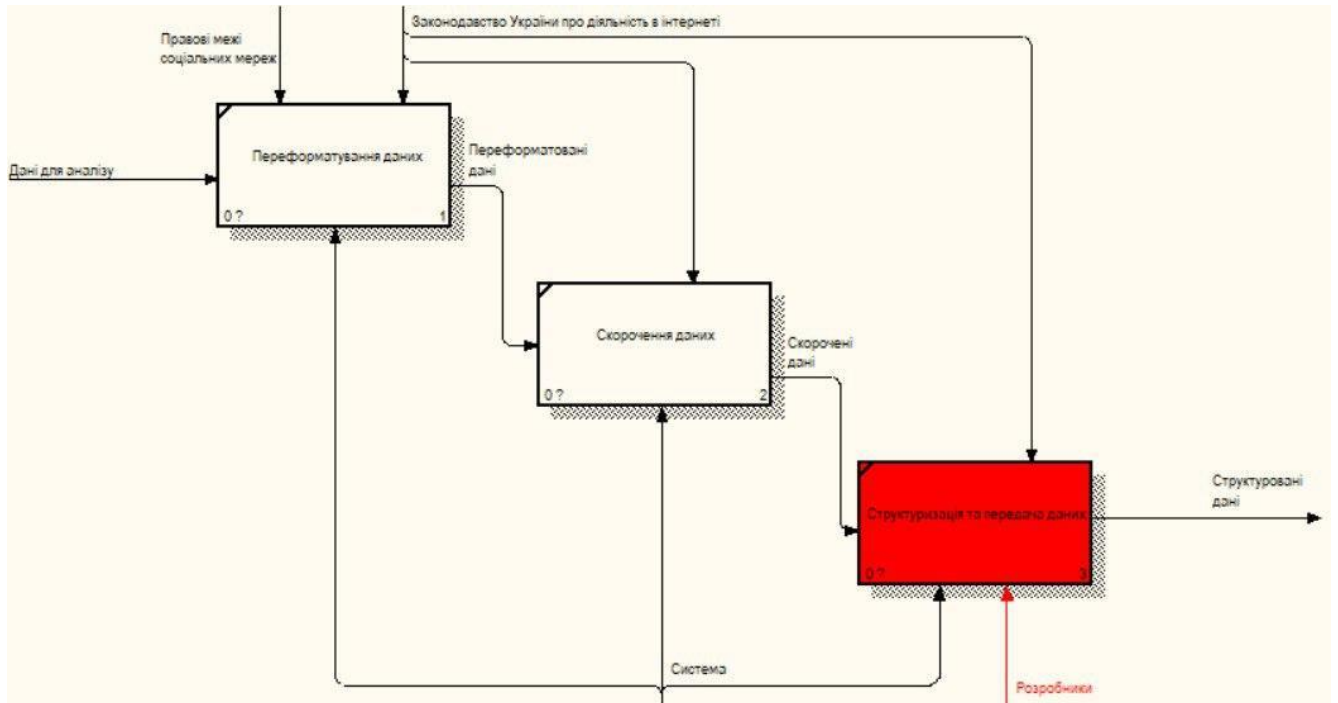


Рисунок 1.4 – Декомпозиція процесу "Аналіз наявних даних"

Процес аналізу умов фільтрації також необхідно розділити на декілька етапів. Це є формування списку наявних фільтрів зі структурованих даних. Далі система шукає зв'язок та подібності між наявним списком фільтрів та необхідними для роботи. Та вже дані зв'язки оброблює та повертає у вигляді структурованого списку фільтрів (рисунок 1.5).

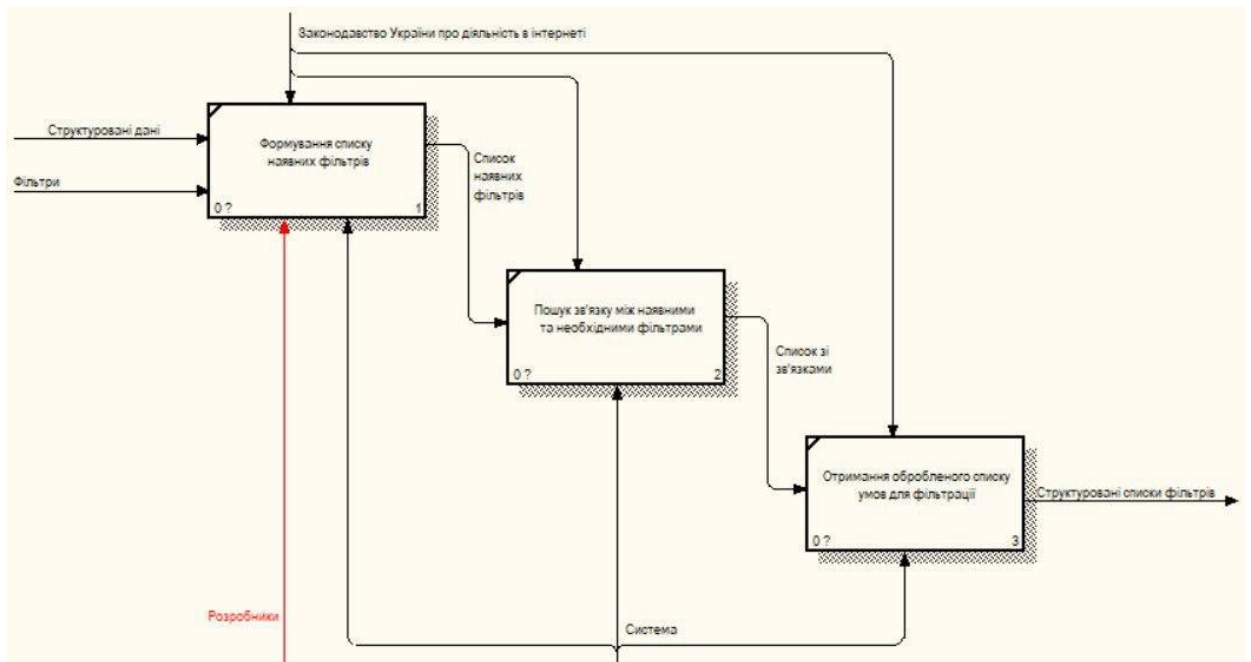


Рисунок 1.5 – Декомпозиція процесу "Аналіз умов фільтрації"

Третій, останній, процес отримання статистики здійснюється в такий спосіб. Так як вже є структуровані дані та структурований список фільтрів, то необхідно розділити дані за цими фільтрами. Серед даних груп виділити дані, які саме запитував користувач та сформувати на їх основі таблиці та графіки статистики.

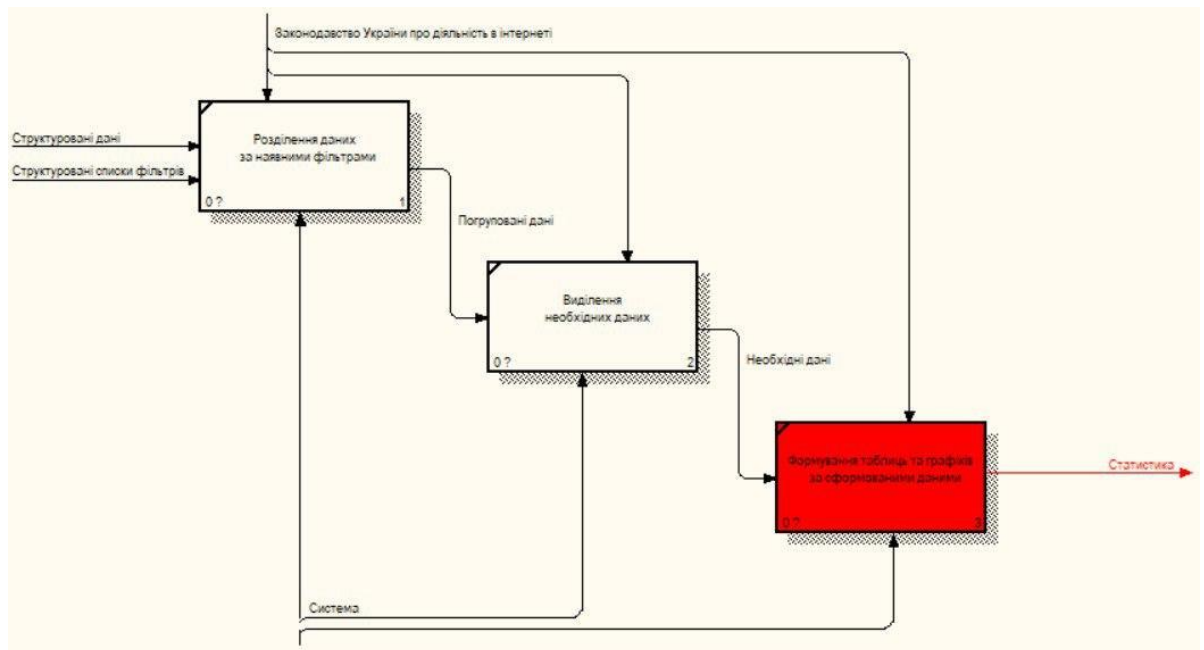


Рисунок 1.6 – Декомпозиція процесу «Отримання статистики»

На основі розробленої моделі можна виділити основні недоліки процесу «Визначення емоційного забарвлення постів у соціальних мережах». В першу чергу це є відсутність можливості отримати саме дані, а не лише статистику по них.

Також немає автоматичного оновлення даних. Тобто кожен раз, коли необхідно отримати нову інформацію, систему треба запускати наново та наново проходити кожен етап аналізу даних, пошуку фільтрів. Саме тому необхідно створити застосунок, який буде мати вже підготовлену для фільтрування систему, самостійно отримувати з мережі пости та повертати інформацію по заданому запиту користувача.

1.4. Постановка задачі на розробку визначення емоційного забарвлення постів у соціальних мережах

В першу чергу визначимо вимоги для вхідних даних. Вони повинні задовольняти декільком критеріям:

- репрезентативність — дані повинні ілюструвати дійсне положення речей у предметній області;
- несуперечність — суперечливі дані в навчальній вибірці призведуть до поганої якості навчання мережі;
- обсяг — як правило, число записів у вибірці повинне на кілька порядків перевершувати кількість зв'язків між нейронами в мережі. У протилежному випадку мережа просто «запам'ятає» усю навчальну вибірку і не зможе виконати узагальнення. Для вирішення задачі визначення емоційного забарвлення тексту було обрано взяти вибірку з 20 тисяч постів;
- різноманітність — система повинна навчатись всьому разом, знаходячи найкращі вагові коефіцієнти для загальної множини прикладів та дані повинні бути достатньо великі, щоб містити всю необхідну інформацію для виявлення важливих особливостей і зв'язків.

Також у нас є нефункціональні вимоги, які в першу чергу включають в себе:

- доступність на затребуваних пристроях (користувач повинен мати змогу відкрити це як на телефоні, так і на комп'ютері з різними операційними системами);
- унікальність продукту;

- зручний та зрозумілий інтерфейс (у користувача не має виникати питань що потрібно зробити, аби нарешті мати змогу використовувати застосунок за призначенням);
- легкість підтримки системи (код повинен бути зрозумілий не лише його розробникам, але і іншим співробітникам, які можуть в майбутньому приєднатись до цього проекту);
- надійність (необхідно перевірити систему за неочікуваних умов, великої кількості користувачів та під час перевантаження запитами);
- переносимість (здатність програмної системи працювати в іншому середовищі, якщо основний залежний фреймворк залишається незмінним);
- швидкість відповіді на запит (фільтрація постів повинна відбуватись швидше ніж за хвилину).

Функціональні вимоги будуть складатися з таких пунктів:

- початкова сторінка має містити в собі форму для реєстрації та форму входу;
- форма для реєстрації вимагає введення пошти, на яку має прийти лист з підтвердженням реєстрації та переадресацією на сторінку завершення реєстрації;
- під час завершення реєстрації користувач має двічі ввести пароль для власного облікового запису та натиснути на кнопку «Зберегти» для завершення реєстрації;
- форма для входу вимагає введення електронної пошти та паролю, та після натискання на кнопку «Увійти» переходить до облікового запису;
- користувач повинен мати змогу під'єднати свій аккаунт з соціальної мережі Twitter до системи;
- користувач має змогу обирати список користувачів, пости яких його цікавлять;
- користувач має широкий вибір настроїв (сумні, веселі, серйозні і т.д.) та може обрати один або декілька для застосування фільтрації за цими характеристиками (при цьому якщо користувач обрав, для прикладу, веселі,

йому блокується доступ до фільтру сумні, так як системі буде неможливо знайти пост, який буде задовольняти одночасно дві ці вимоги);

- система аналізує на настрої текст та зображення, щоб мати цілісну картину настрою даного посту;
- під часу процесу застосування фільтрів та завантаження відфільтрованих постів, користувач має бачити анімоване попередження про те що відбувається процес завантаження постів;
- якщо запит виконується занадто довго або виникли якісь проблеми під час обробки запиту, користувачу пропонується повторити запит або ж змінити умови цього запиту;
- вказувати часовий проміжок, за який цікавлять пости;
- система використовує додаткові сервіси для збору та аналізу використання застосунку користувачами (які помилки у них виникають, по якій причині, якими функціями додатку користуються найчастіше).

1.5. Висновки

Перший розділ був присвячений аналізу процесу визначення емоційного забарвлення постів у соціальних мережах. Розглянуто роботи Роберта Стіна та Бінг Люя. Перший у своїх роботах описав вирішення задачі за допомогою нейронних мереж. Бінг Лю запропонував варіант некерованого навчання, де слова, які застосовуються для навчання, обираються за спеціальним шаблоном (так як саме ці шаблони зазвичай виражають думку людини). Було наведено вже існуючі інформаційні системи для визначення емоційного забарвлення тексту, описано процес роботи одного з них з використанням методології IDEF0 та DFD. Основний мінус даних застосунків – вони розраховані для успішного ведення бізнесу, аналітики постів, але не для звичайного користувача, який хоче відфільтрувати інформацію, яка надходить з соціальних мереж. Також визначено основні вимоги до розроблюваного застосунку. В першу чергу це є вимоги до вхідних даних – репрезентативність, несуперечність, обсяг та різноманітність. Серед основних нефункціональних вимог є доступність на затребуваних пристроях, унікальність

продукту, зручний інтерфейс та швидка відповідь. У функціональних вимогах описано основні пункти для реалізації у застосунку, такі як авторизація, збереження даних та умови фільтрації постів.

2. Розробка архітектури застосунку для визначення емоційного забарвлення постів у соціальних мережах

2.1 Розробка архітектури

2.1.1 Функціональний аналіз

Основна мета функціонування системи: допомога користувачу у фільтруванні інформації, яку він отримує з соціальних мереж.

Таблиця 2.1 – Класифікаційні ознаки розроблюваної системи

№	Класифікаційна ознака	Клас за ознакою	Обґрунтування
1	Природа елементів	Абстрактні	Система не має фізичного розташування, вона працює лише у віртуальній мережі
2	Походження	Штучні	Її створюють розробники
3	Тип змінних	З якісними змінними	Оцінити роботу системи можна по тому наскільки добре вона справляється з поставленою задачею, тобто як добре вона фільтрує пости
4	Тип оператора системи	Чорний ящик	Користувачу невідомо яким чином система отримує відфільтровані дані
5	Термін існування	Тимчасові	Робота даної системи може бути припинена (так само як і підтримка її подальшої працездатності та оновлення)
6	Мінливість властивостей	Динамічні	Система має постійно оновлюватись, аби не втрачати актуальність та бути конкурентоспроможними

7	Ступінь складності	Складні	Багато незалежних підсистем
8	Відношення до зовнішнього середовища	Відкриті	Система співпрацює з соціальними мережами та отримує інформацію звідти
9	Реакція на вплив, що збурює	Пасивні	Зовнішні чинники не впливають на роботоспроможність системи
10	Характер управління	З комбінованим управлінням	Є розробники, які в процесі роботи вдосконалюють систему Але окрім цього вона сама отримує нові пости, навчається з ними працювати
11	Ступінь організованості	Добре організовані	Даний застосунок є самостійною системою, яка має свою ієрархію та кожна підсистема стоїть під контролем
12	Ступінь участі та впливу людини	Людино-машинні	Розробники займають створенням системи, яка в подальшому вже самостійно буде виконувати задачу її роботу

На рисунку 2.1 зображена карта процесів визначення емоційного забарвлення постів у соціальних мережах. Вона дозволяє візуально описати хід роботи та показати серію подій, які приведуть до кінцевого результату, в даному випадку до постів за визначеним настроєм.

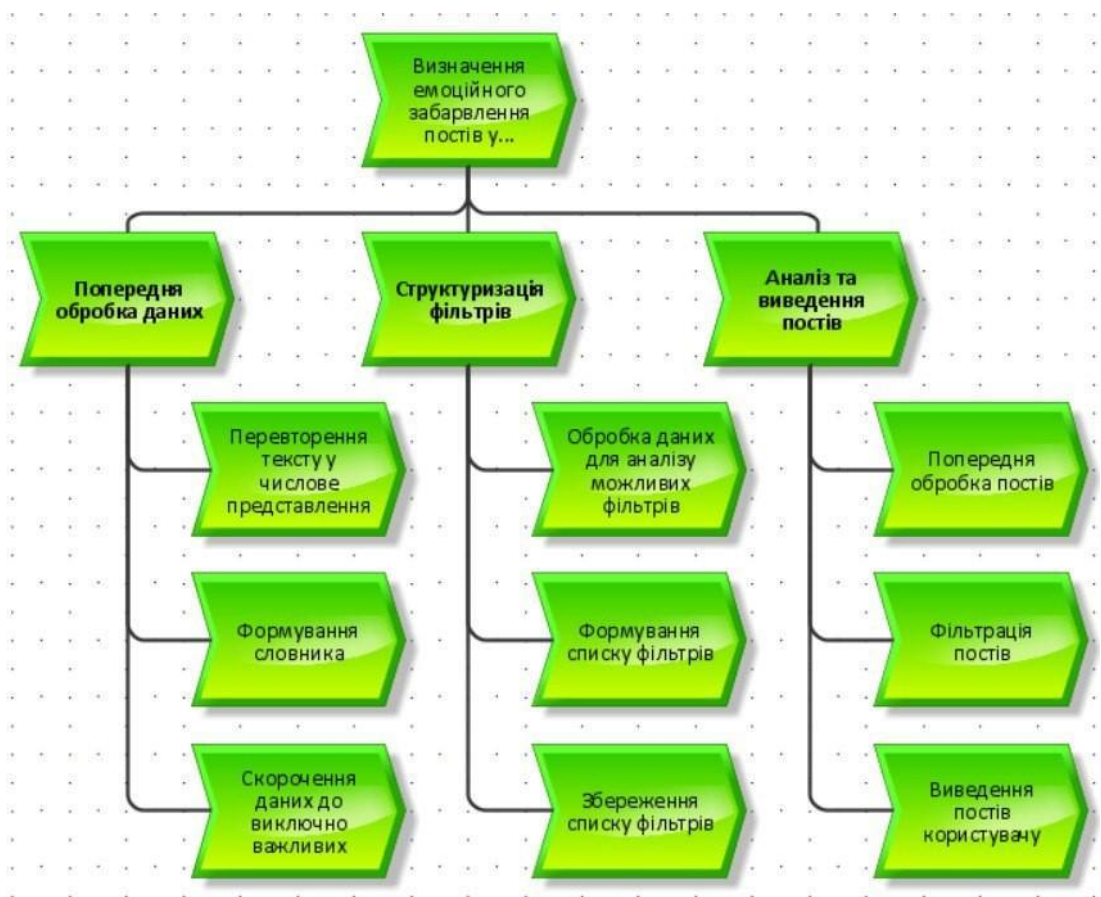


Рисунок 2.1 – Карта процесів

Для оформлення бізнес-процесу в нотації eEPC, в першу чергу необхідно визначити:

1.1. Елементи оточення бізнес-процесу:

Сутності – пости.

Документи – запит користувача на отримання певних постів.

Ролі – система.

1.2. Сховища даних – пости з Twitter.

Загалом для роботи над застосунком було сформовано 2 бізнес-процеси: навчання нейронної мережі та отримання даних за запитом користувача. На рисунку 2.2 схема першого запиту, що є підготовкою системи до роботи з заявками користувача.

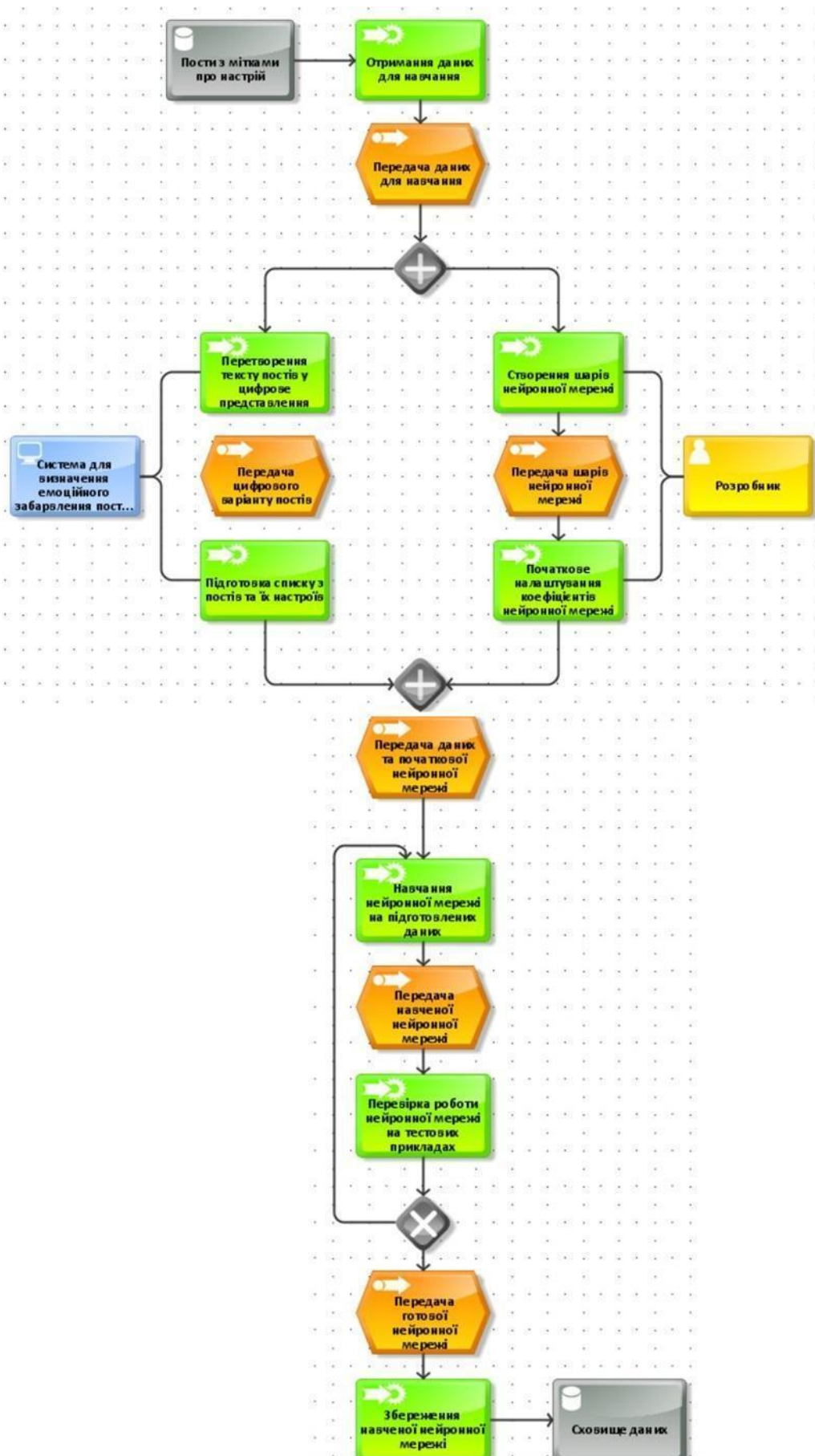


Рисунок 2.2 – Діаграма процесу отримання даних за запитом користувача

На рисунку 2.3 описано діаграму процесу отримання даних за запитом користувача, тобто які пункти проходить система від моменту отримання заявки на фільтрацію до моменту виведення користувачу постів, які його цікавлять.

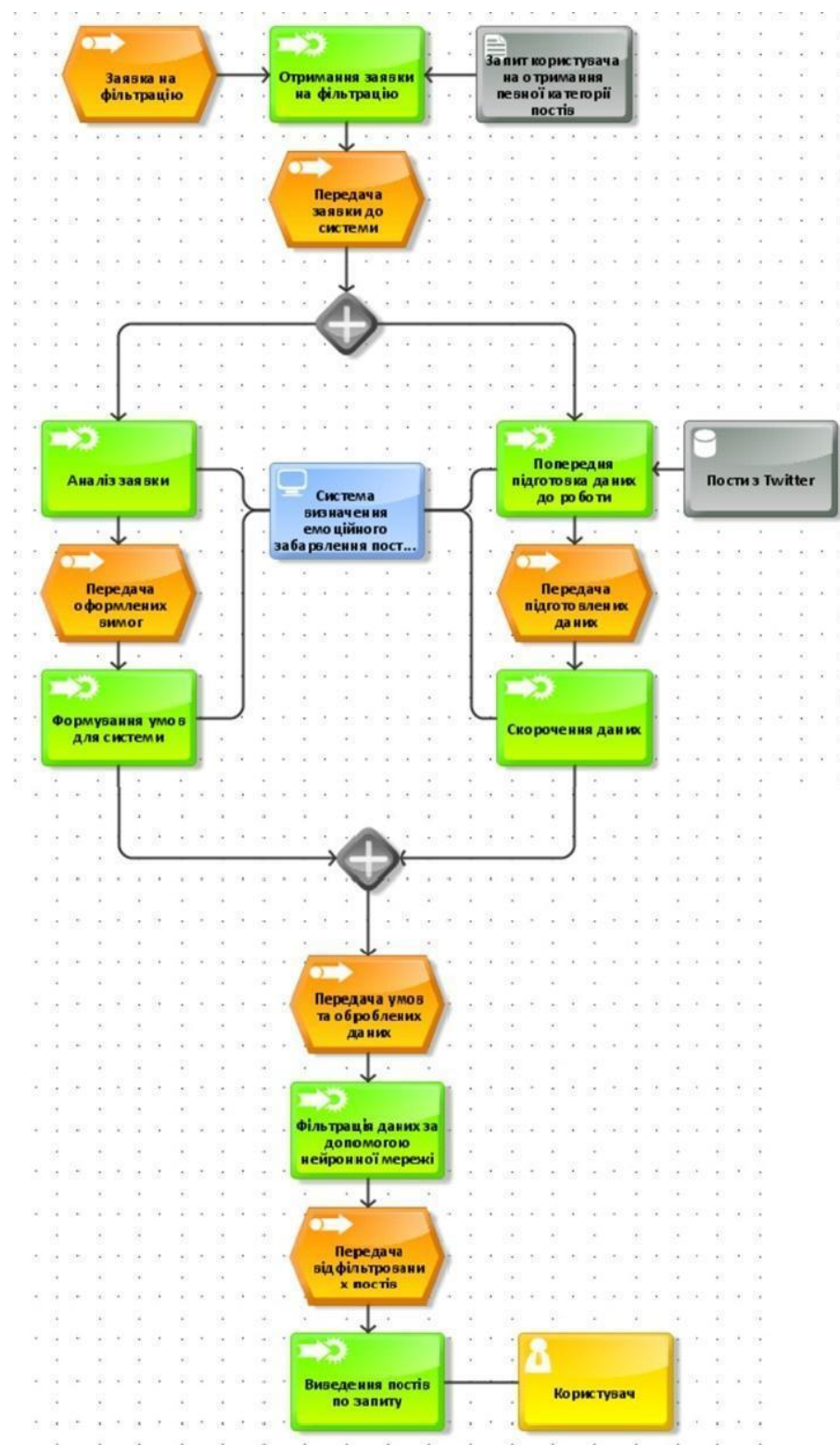


Рисунок 2.3 – Діаграма процесу навчання нейронної мережі

2.1.2 IDEF0 діаграма процесу визначення емоційного забарвлення тексту

Розглянемо контекстну діаграму процесу визначення емоційного забарвлення тексту ЯК-БУДЕ (рис. 2.4).

Основні зв'язки контекстної діаграми наведені в таблиці 2.2.

Таблиця 2.2 – Опис зв'язків контекстної діаграми ЯК-Є

Назва	Тип	Зовнішній об'єкт
Дані для аналізу	Вхідна інформація	Соціальні мережі
Вимоги з фільтрації	Вхідна інформація	Користувач
Відфільтровані пости	Вихідна інформація	Користувач

Розглянемо декомпозицію контекстної діаграми (рис. 2.5). Даний бізнес-процес складається з таких бізнес-процесів:

1. Аналіз наявних даних

Вхідна інформація:

- Дані для аналізу

Вихідна інформація:

- Структуровані дані

2. Аналіз умов фільтрації

Вхідна інформація:

- Структуровані дані
- Фільтри

Вихідна інформація:

- Структурований список фільтрів

3. Отримання постів

Вхідна інформація:

- Структуровані дані
- Структурований список фільтрів

Вихідна інформація:

- Відфільтровані пости

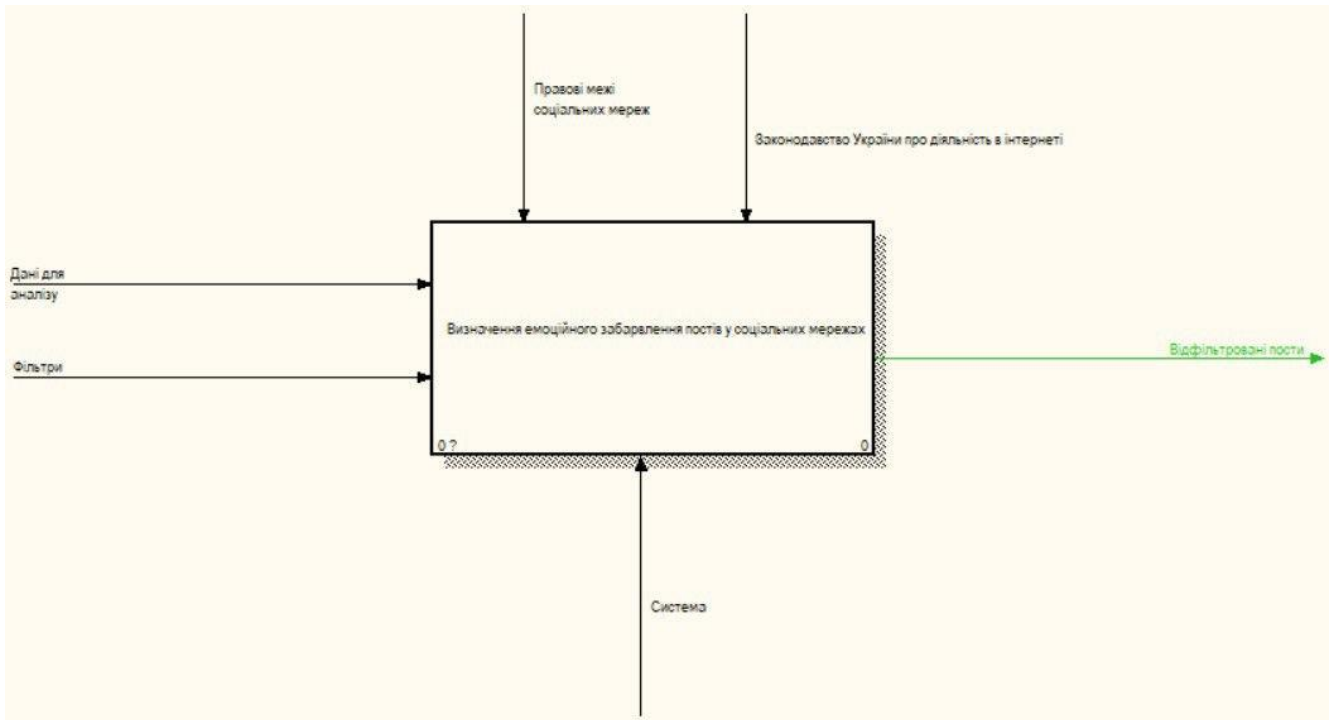


Рисунок 2.4 – Контекстна діаграма процесу "Визначення емоційного забарвлення постів у соціальних мережах"

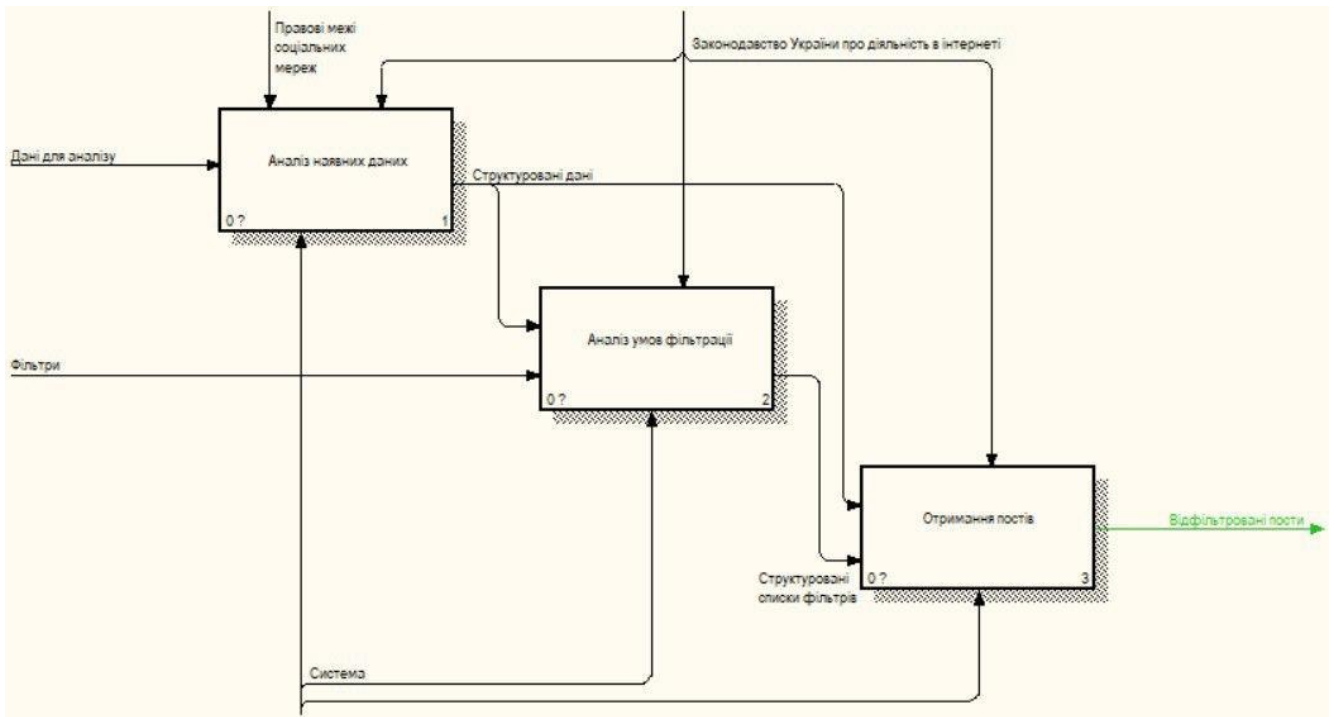


Рисунок 2.5 – Декомпозиція контекстної діаграми процесу "Визначення емоційного забарвлення постів у соціальних мережах"

Розглянемо докладніше кожен процес у вигляді DFD діаграми (рисунок 2.6).

Процес аналізу наявних даних здійснюється в такий спосіб. Спочатку система отримує дані у звичайному текстовому вигляді. Далі їй необхідно перевести до чисельного вигляду, який буде прийнятний для системи. З цих даних необхідно створити словник, так як таким чином буде ще легше зберігати дані, просто давши кожному слову свій порядковий номер. Вже дані зі словника проходять процес скорочення, структурування та збереження у такому вигляді для подальшої роботи (так як ці дані можуть використовуватись не один раз і не хочеться кожен раз заново запускати весь процес аналізу).

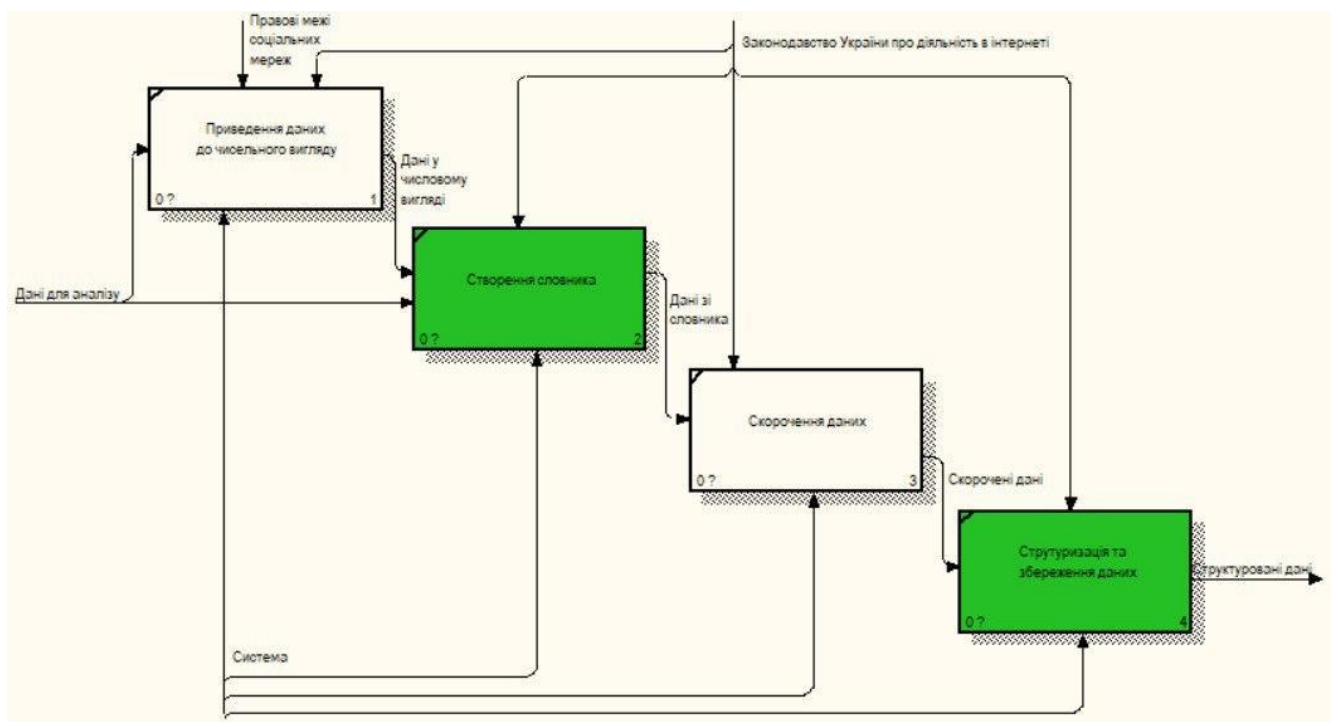


Рисунок 2.6 – Декомпозиція процесу "Аналіз наявних даних"

Процес аналізу умов фільтрації також необхідно розділити на декілька етапів. Це є формування списку наявних фільтрів зі структурованих даних. Далі система виділяє необхідні фільтри серед наявних. Та вже необхідні фільтри структуруються та повертаються у вигляді структурованого списку фільтрів (рисунок 2.7).

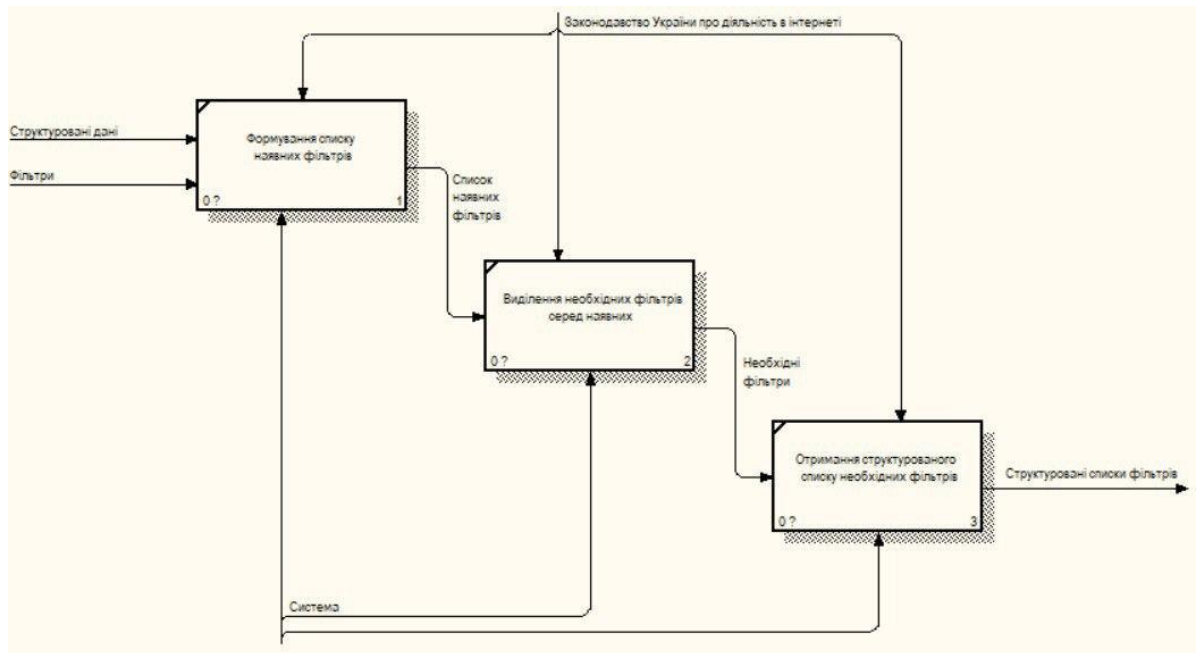


Рисунок 2.7 – Декомпозиція процесу "Аналіз умов фільтрації"

Третій, останній, процес отримання постів здійснюється в такий спосіб. Так як вже є структуровані дані та структурований список фільтрів, то необхідно розділити дані за цими фільтрами. Серед даних груп виділити дані, які саме запитував користувач та вивести їх користувачу для перегляду (рисунок 2.8).

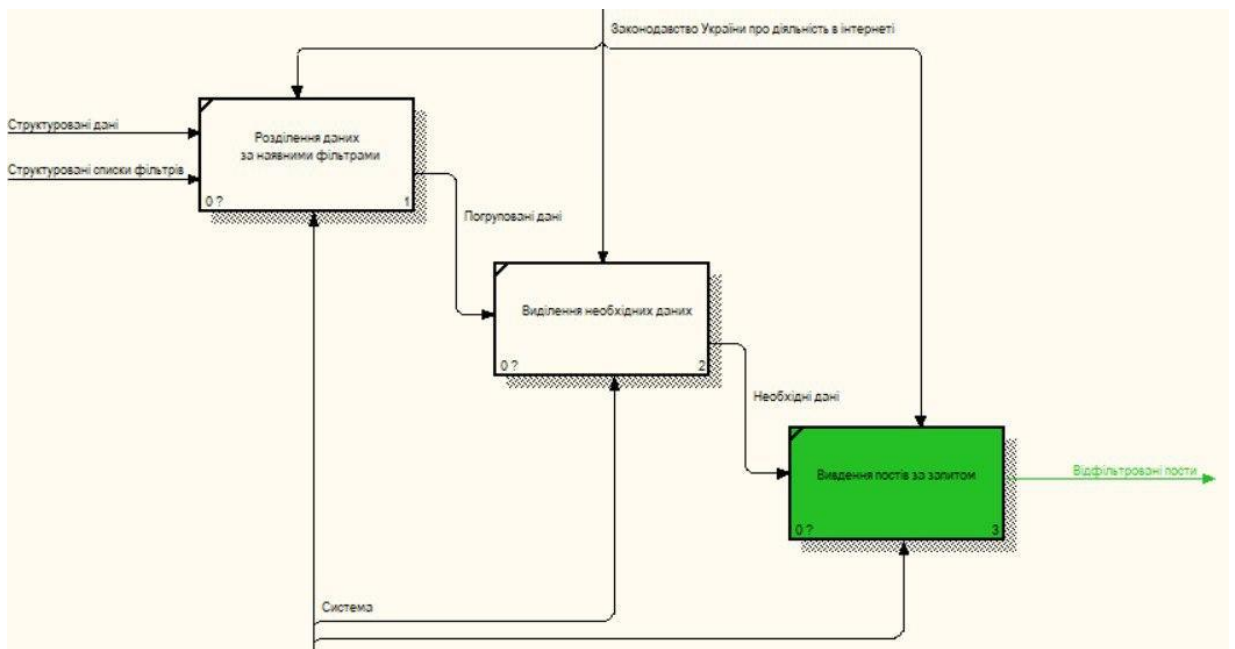


Рисунок 2.8 – Декомпозиція процесу «Отримання постів»

2.1.3 Архітектура інформаційної системи

Інформаційна система визначення емоційного забарвлення постів у соціальних складається з трьох основних підсистем: підсистема роботи з даними, роботи з фільтрами та роботи з користувачем. У свою чергу вони поділяються на такі модулі.

1. Підсистема роботи з даними

- модуль приведення тексту до системного вигляду (пости приходять у вигляді звичайних символів, але система їх не розпізнає, тому необхідно їх перевести як мінімум до вигляду чисел)
- модуль роботи зі словником (необхідно кожному слову дати певний номер, аби потім системі було легше знаходити закономірності, правила для класифікації)
- модуль збереження даних (так як для одних і тих самих даних немає потреби ще раз проходити етап попередньої обробки і для повторного використання краще зберегти їх вже в такому вигляді)

2. Підсистема роботи з фільтрами

- модуль формування списку фільтрів (відбувається пошук даних, за якими можна класифікувати наявні пости)
- модуль пошуку зв'язку між списком та необхідними фільтрами (запит користувача порівнюється з тим, за якими класами є можливість розділити дані)
- модуль отримання обробленого списку (проаналізувавши список фільтрів, повертається структурований список фільтрів, які користувачу необхідні)

3. Підсистема роботи з користувачем

- модуль розділення даних за фільтрами (дані діляться на групи за їх подібністю до характеристик певного фільтру)

- модуль отримання даних за запитом (виділення групи, яка саме цікавить користувача)
- модуль виведення даних за запитом (користувачу необхідно побачити результат його запиту)

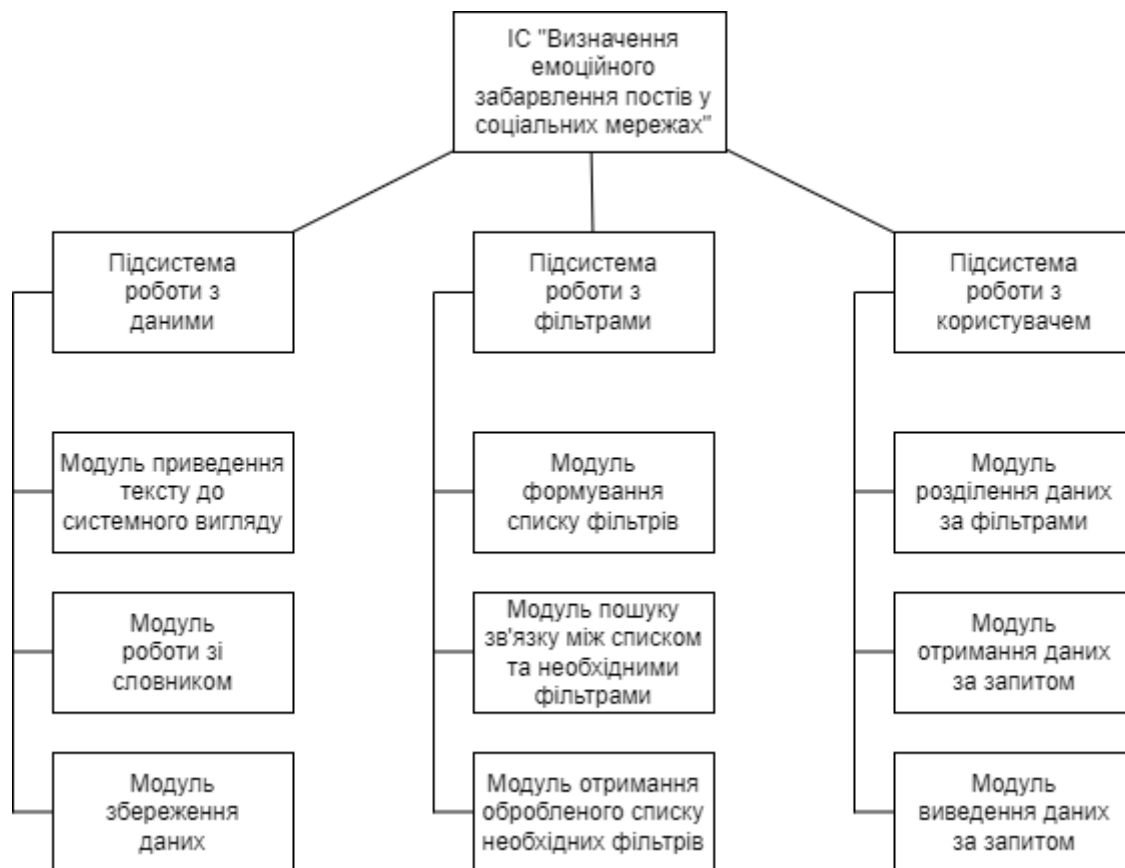


Рисунок 2.9 – Архітектура ІС

2.2 Інформаційне забезпечення

Для розроблюваного застосунку необхідно згідно всіх вимог створити БД, яка буде описувати всі сутності та роботу з ними.

Таблиця 2.3 – Словник даних

№	Назва елемента даних	Ідентифікатор	Тип даних	Документ-джерело	Призначення елемента даних
User					

1	Id користувача	id	ID	API-запит до Twitter	Ідентифікація користувача
2	Ім'я користувача	name	String	API-запит до Twitter	Ідентифікація користувача
3	Ім'я в профілі	screen_name	String	API-запит до Twitter	
4	Розташування	location	String	API-запит до Twitter	
5	Опис користувача	description	String	API-запит до Twitter	
6	Кількість підписників	followers_count	Integer	API-запит до Twitter	
7	Кількість друзів	friends_count	Integer	API-запит до Twitter	
8	Кількість слухачів	listed_count	Integer	API-запит до Twitter	
9	Дата створення аккаунту	created_at	Date	API-запит до Twitter	
10	Кількість вподобань	favourites_count	Integer	API-запит до Twitter	
11	Посилання на зображення профілю	profile_image_url	String	API-запит до Twitter	
Tweet					
12	Id посту	id	ID	API-запит до Twitter	Ідентифікація посту
13	Дата публікації	created_at	Date	API-запит до Twitter	

14	Текст посту	text	String	API-запит до Twitter	
15	Джерело	source	String	API-запит до Twitter	З якого пристрою зробили публікацію (PC, Android, iOS)
16	Координати	coordinates	String	API-запит до Twitter	
17	Id користувача	user_id	ID	API-запит до Twitter	Хто створив пост
18	Id посту, ретвіт якого було зроблено	retweeted_id	ID	API-запит до Twitter	
19	Кількість ретвітів	retweet_count	Integer	API-запит до Twitter	
20	Кількість лайків	favorite_count	Integer	API-запит до Twitter	
21	Мова	lang	String	API-запит до Twitter	
Retweeted post					
22	Id посту	id	ID	API-запит до Twitter	Ідентифікація посту
23	Дата публікації	created_at	Date	API-запит до Twitter	
24	Текст посту	text	String	API-запит до Twitter	

25	Джерело	source	String	API-запит до Twitter	
26	Id користувача	user_id	ID	API-запит до Twitter	Хто створив пост
27	Координати	coordinates	String	API-запит до Twitter	
28	Кількість ретвітів	retweet_count	Integer	API-запит до Twitter	
29	Кількість лайків	favorite_count	Integer	API-запит до Twitter	
30	Мова	lang	String	API-запит до Twitter	
Tweet for show					
31	Id посту	id	ID	Нейронна мережа	Ідентифікація посту
32	Текст посту	text	String	Нейронна мережа	Інформація для користувача
33	Ім'я користувача	user_name	String	Нейронна мережа	Інформація для користувача
34	Дата публікації посту	created_at	Date	Нейронна мережа	Інформація для користувача
35	Id настрою посту	sentiment_id	ID	Нейронна мережа	Інформація для користувача

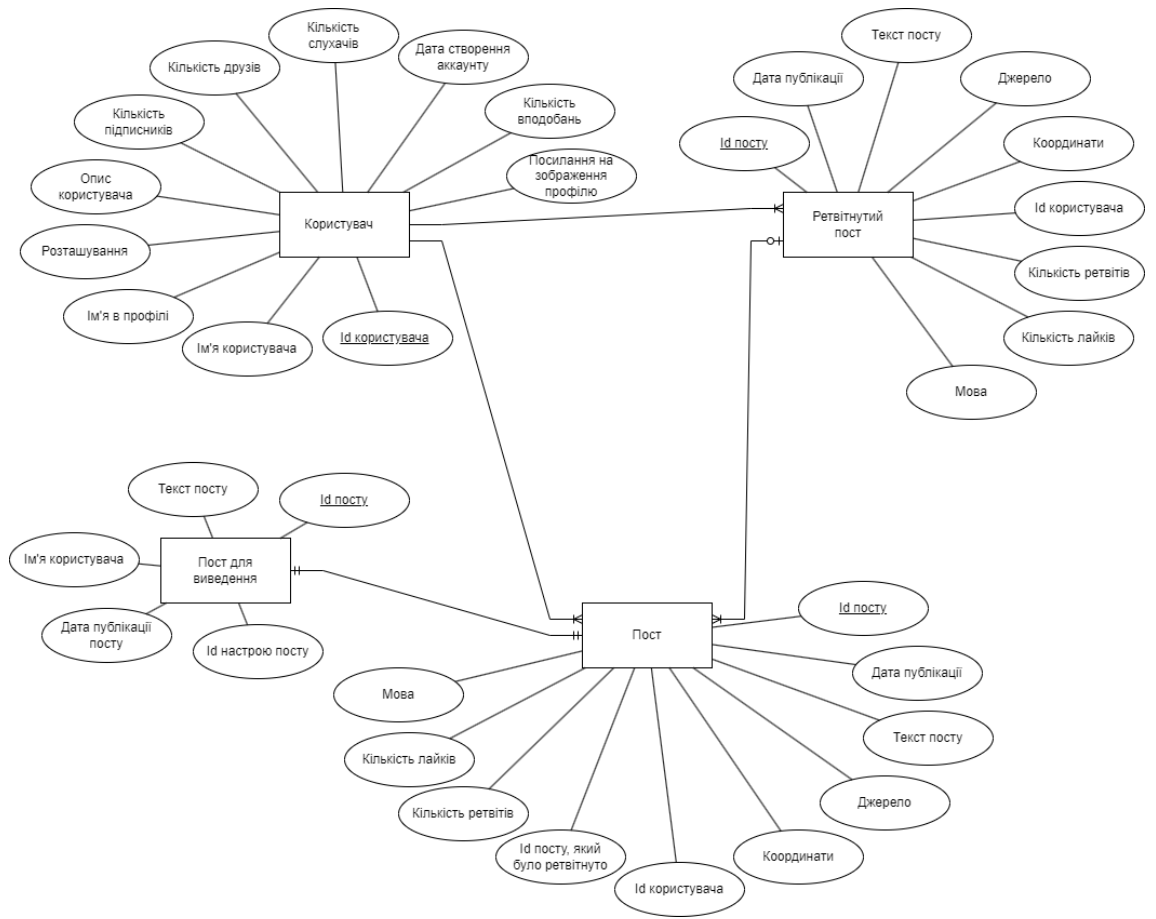


Рисунок 2.10 – Концептуальна модель даних

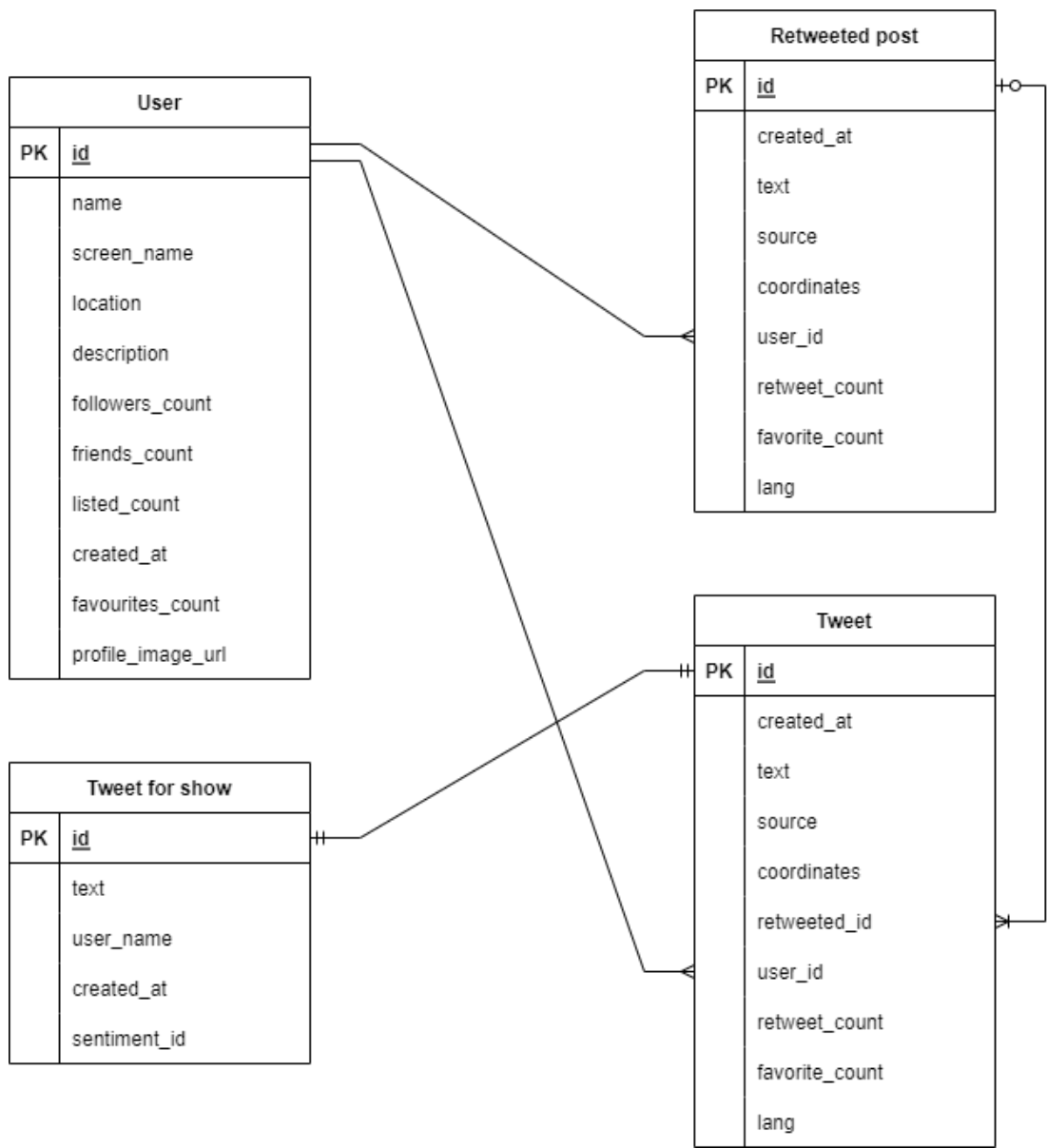


Рисунок 2.11 – Логічна модель даних

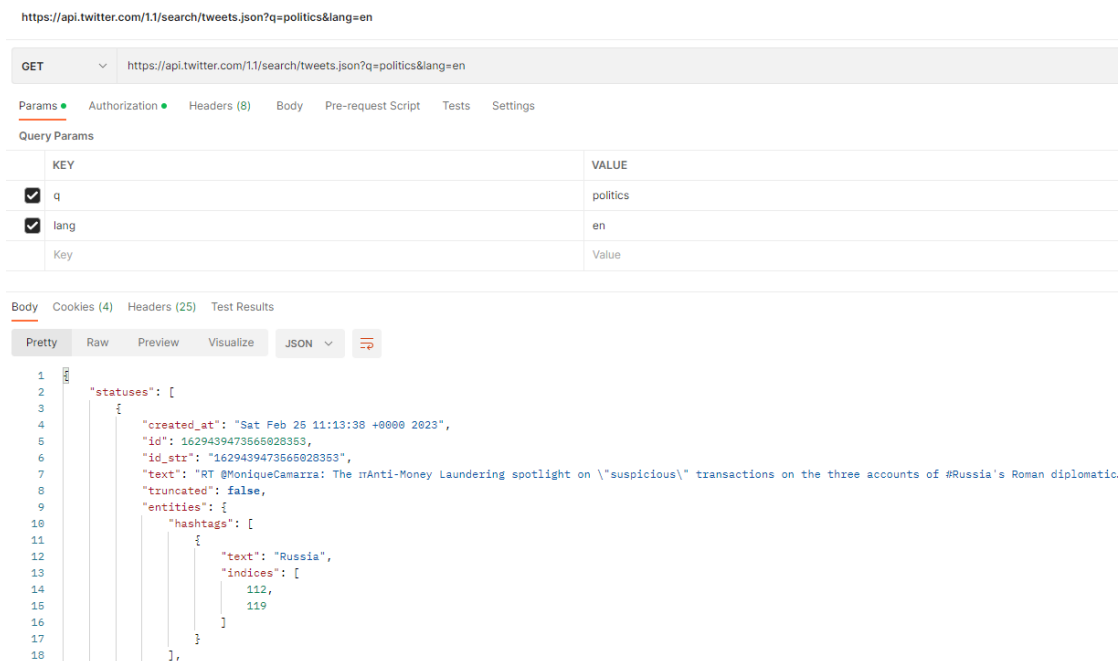


Рисунок 2.12 – Формат отримання даних

Отже, застосунок складається з чотирьох таблиць бази даних. Найбільш важлива з них – Пост. Саме вона несе інформацію про текст публікації, дату, мову та кількість вподобань. Пост може бути сформований за допомогою ретвіту іншої публікації. Тоді даний пост матиме id посту, ретвіт якого привів до його створення. За цим id є можливість прив'язати таблицю БД Ретвітнутий пост. Також пост зберігає id для визначення користувача, який зробив дану публікацію. За першим ним ми вже прив'язуємось до таблиці з інформацією про користувача.

Окремий зв'язок є з таблицею Пост для виведення. Вона утворюється зі всіх трьох описаних вище таблицю в процесі роботи нейронної мережі та містить лише ту інформацію, яка буде повертатись користувачу за його запитом.

2.3 Висновки

Другий розділ був присвячений розробці архітектури застосунку для визначення емоційного забарвлення постів у соціальних мережах. Було сформовано класифікаційні ознаки розроблюваної системи та побудовано карту процесів роботи застосунку. Також описано 2 основних бізнес-процеси: отримання даних за запитом користувача та процес навчання нейронної мережі. В

попередньому розділу було проаналізовано контекстну діаграму процесу ЯК-Є, тому тепер побудовано діаграму процесу ЯК-БУДЕ та виділено моменти, яких немає у застосунках, розглянутих раніше. Також було побудовано архітектуру інформаційної системи визначення емоційного забарвлення постів у соціальних мережах, визначено основні підсистеми та модулі роботи.

Ще одним важливим моментом було сформувавши інформаційне забезпечення розроблюваного застосунку для визначення емоційного забарвлення постів у соціальних мережах. В першу чергу це є словник даних з описом всіх змінних та яку інформацію вони в собі несуть. Визначено, що основним джерелом даних у застосунку буде API-запит до Twitter та результат роботи нейронної мережі. Також сформовано логічні та концептуальні моделі даних. В результаті база даних буде складатись з чотирьох таблиць: Користувач, Пост, Ретвітний пост та Пост для виведення.

3. Програмне забезпечення для визначення емоційного забарвлення постів у соціальних мережах

3.1 Обґрунтування вибору програмних засобів

Для програмної реалізації задачі визначення емоційного забарвлення постів у соціальних мережах було обрано мову Python. Вона є однією з найбільш сучасних мов програмування, має дуже гарну продуктивність та портативність. Тобто є можливість запускати один і той самий код на будь-яких пристроях без додаткових коригувань. Також ця мова має дуже велику базу бібліотек для спрощення роботи. Python використовується для розробки онлайн і мобільних додатків, машинного навчання, при створенні ігор і зручний для автоматизації математичних розрахунків. Тому його можна застосовувати для вирішення практично будь-яких завдань[17].

Розглянемо більш детально бібліотеки Python, які було використано в курсовій роботі.

1. Pandas – бібліотека для маніпулювання та аналізу даними. Зокрема, вона пропонує структури даних та операції для маніпулювання числовими таблицями та часовими рядами. В задачі курсової дана бібліотека застосовується для обробки даних з вбудованою індексацією (dataFrame).

2. NumPy - бібліотека, яка підтримує роботу з великими, багатовимірними масивами та матрицями, а також має велику колекцію математичних функцій високого рівня для роботи з цими масивами. Таким чином вона частково вирішує проблему повільності.

3. Sklearn містить різні алгоритми класифікації, регресії та кластеризації, включаючи метод опорних векторів, випадкових лісів, градієнтного спуску, k-найближчих сусідів, і призначена для взаємодії з числовими та науковими бібліотеками NumPy та SciPy.

4. Tensorflow – бібліотека для машинного навчання та штучного інтелекту, має фокус саме на навчанні та висновках глибоких нейронних мереж. Вона була розроблена командою Google Brain для внутрішнього використання Google в

дослідженнях і виробництві. TensorFlow пропонує кілька рівнів абстракції, тому є можливість обрати той, який відповідає потребам задачі. Дана бібліотека значно полегшує створення нейронних мереж, покращує їх продуктивність та швидкість роботи. Вона має велику кількість класів для створення шарів нейронної мережі[18]. Для курсової було використано лише декілька з існуючих класів.

TextVectorizer перетворює пакет рядків (один приклад = один рядок) або в список індексів токенів (один приклад = 1D тензор цілочисельних індексів токенів), або в щільне представлення (один приклад = 1D тензор значень з плаваючою комою, що представляють дані про токени прикладу). Цей рівень призначений для обробки вхідних даних на природній мові. Словник для шару має бути або наданий при створенні, або вивчений за допомогою функції `adapt()`. Коли цей шар буде адаптовано, він проаналізує набір даних, визначить частоту окремих значень рядків і створить словник з них. Цей словник може мати необмежений розмір або бути обмеженим, залежно від параметрів конфігурації цього шару; якщо у вхідних даних більше унікальних значень, ніж максимальний розмір словника, то для створення словника будуть використані найбільш частотні терміни.

Cond1D створює ядро згортки, яке згортається з вхідними даними шару в одному просторовому (або часовому) вимірі для отримання тензора виходів.

Dense реалізує операцію: $output = activation(dot(input, kernel) + bias)$, де `activation` - поелементна функція активації, передана як аргумент `activation`, `kernel` - матриця ваг, створена шаром, а `bias` - вектор зсуву, створений шаром (застосовується тільки якщо `use_bias` має значення `True`).

5. Django – це безкоштовний веб-фреймворк з відкритим вихідним кодом на основі мови Python, який слідує архітектурному шаблону "модель-шаблон-подання" (MTV)[13]. Основна мета - полегшити створення складних веб-сайтів, керованих базами даних. Фреймворк наголошує на багаторазовому використанні компонентів, меншій кількості коду, низькому рівню зв'язку, швидкій розробці та використовує принцип "не повторюйся". Python використовується всюди, навіть для налаштувань, файлів та моделей даних. Django також надає додатковий адміністративний інтерфейс створення, читання, оновлення та видалення, який

генерується динамічно шляхом самоаналізу та налаштовується за допомогою адміністративних моделей.

3.2 Структура програмного забезпечення

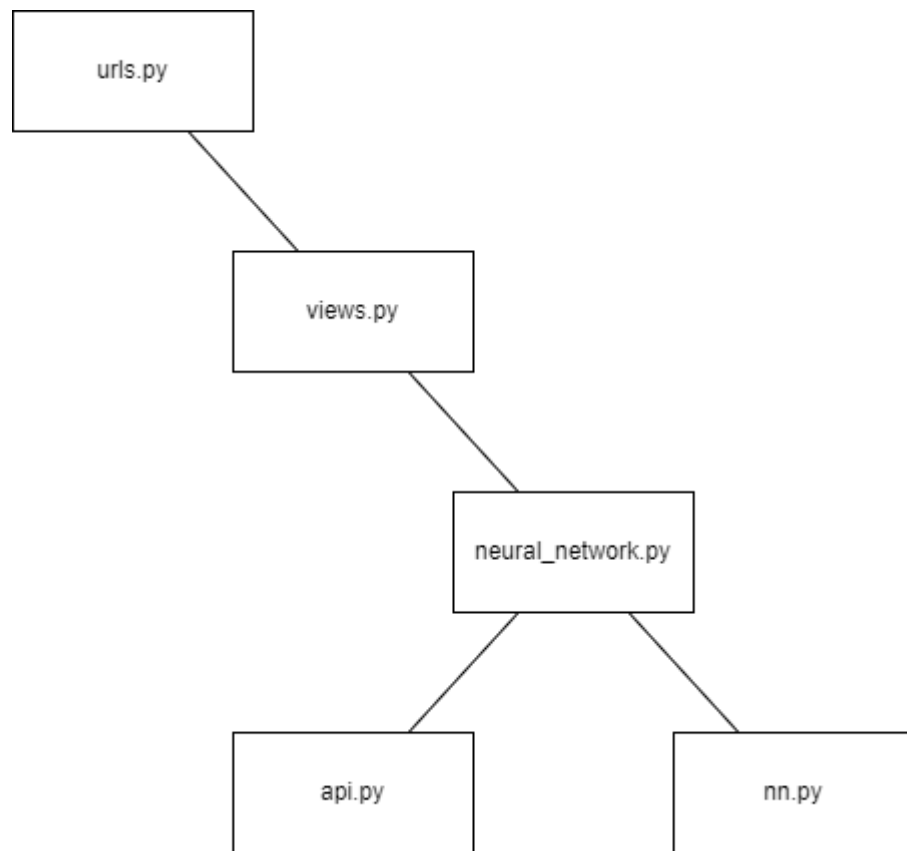


Рисунок 3.1 – Структурна схема програмних модулів

Таблиця 3.1 – Специфікація програмних модулів

Модуль	Опис
urls.py	Модуль для створення роутінга, де за кожною адресою закріплюється функція для отримання цієї сторінки з даними Вхідна інформація - Вихідна інформація Список сторінок
views.py	Модуль для об'єднання html файлу та необхідних даних. Тобто під час виклику передається запит, яка сторінка

	<p>цікавить, модуль оброблює цей запитом та повертає функцію для відтворення сторінки</p> <p>Вхідна інформація</p> <p>Необхідний настрій</p> <p>Вихідна інформація</p> <p>Відтворювана сторінка</p>
neural_network.py	<p>Модуль містить в собі функції для роботи з настроєм:</p> <ul style="list-style-type: none"> - за певним настроєм та запитом, на яку тему цікавлять пости, отримати відповідні пости - отримати id настрою за його назвою - отримати назву настрою за його id - отримати назву настрою для окремого посту за його тестом
api.py	<p>Модуль для отримання постів з твітеру за темою, яка цікавить користувача</p> <p>Вхідна інформація</p> <p>Тема постів</p> <p>Вихідна інформація</p> <p>Пости (без визначеного настрою)</p>
nn.py	<p>Модуль для роботи з нейронною мережею. Це є створення моделі, підготовка даних для навчання та тестування, саме навчання та збереження навченої НМ</p> <p>Вхідна інформація</p> <p>Дані для навчання</p> <p>Вихідна інформація</p> <p>Файл з навченою нейронною мережею</p>

NN
+ one_hot_encoder: OneHotEncoder + label_encoder: LabelEncoder + train_sentences: list + train_dataset: BatchDataset + valid_dataset: BatchDataset
+ get_lines_from_file(fileName): list[string] + get_lines_with_target(fileName): list + get_data_frame(fileName): DataFrame + get_train_labels_encrypt(targets): ndarray + get_valid_labels_encrypt(targets): ndarray + get_dataset(sentences, labels_encode): BatchDataset + get_text_vectorizer_layer(): any + get_token_embed_layer(text_vectorizer): any + prepair_data(train_filename, valid_filename): void + create_model(): void + run(num_epochs): void + save_model(dir): void + predict(text): any

Рисунок 3.2 – Діаграма класу NN

Таблиця 3.2 – Функції класу NN та їх призначення

Функція	Призначення
get_lines_from_file(filename): list[string]	Відкриває файл для читання та повертає порядково інформацію, яку описано у файлі
get_lines_with_target(filename): list	Рядки, які було витягнуто з файлу, діляться на текст посту та його настрої: (-1) – негативний 0 – нейтральний 1 – позитивний

<code>get_data_frame(filename): DataFrame</code>	Перетворення масиву рядків з настроєм в DataFrame (структура даних, яка легше піддається роботі в нейронних мережах)
<code>get_train_labels_encrypt(targets): ndarray</code>	Кодує емоційне забарвлення рядків з тренувальної вибірки в таке значення, яке буде зрозуміле програмі (якби текст мав характеристику «поганий», функція натомість призначила б йому деяке ціле число, наприклад, 1)
<code>get_valid_labels_encrypt(targets): ndarray</code>	Аналогічно попередній функції, але застосовується вже для тестової вибірки
<code>get_dataset(sentences, labels_encode): BatchDataset</code>	Пости все ще мають тип даних list, тому їх необхідно перевести в BatchDataset для швидшого завантаження та обробки
<code>get_text_vectorizer_layer(): any</code>	Попередня обробка тексту, при якому створюється словник слів і кожному слову дається певне числове значення. Чим частіше це слово використовується, тим менше значення воно матиме
<code>get_token_embed_layer(text_vectorizer): any</code>	Створення першого шару нейронної мережі, при якому кожне числове значення словника буде перетворюватися на щільний вектор фіксованого розміру (наприклад 4 -> [0.2, -0.41, 0.67])
<code>prepair_data(self, train_filename, valid_filename): void</code>	Включає в себе всю попередню роботу з даними, тобто виклик функцій <code>get_data_frame</code> , <code>get_train_labels_encrypt</code> , <code>get_valid_labels_encrypt</code> та <code>get_dataset</code>

<code>create_model(self): void</code>	Ця функція виконує попереднє створення моделі нейронної мережі: отримання Embedding шару, формування вхідного, згорткового, вихідного шару та компіляція всіх шарів з заданням функції оптимізації та списку метрик для оцінки
<code>run(self, num_epochs): void</code>	Виконується запуск створеної моделі нейронної мережі задану кількість ітерацій
<code>save_model(self, dir): void</code>	Збереження навченої нейронної мережі для подальшої роботи з нею вже в додатку
<code>predict(self, text): any</code>	Функція, в яку передається текст, а повертається матриця приналежності тексту до можливих настроїв. Наприклад, якщо вона поверне масив [0.15, 0.25, 0.6], це означає що з більшою вірогідністю цей текст належить до третього класу емоційного забарвлення, тобто «позитивний»

3.3 Керівництво користувача

Загалом застосунок має 2 сторінки: сторінка введення теми та сторінка з відсортованими постами. Початковою є сторінка введення теми. Вона містить поле для введення тексту та кнопку для збереження тексту та переходу на наступну сторінку. В полі для тексту користувачу необхідно ввести тему, за якою його цікавлять пости. Після чого натиснути на кнопку "Save". Після він автоматично переміститься на сторінку з відповідними постами.

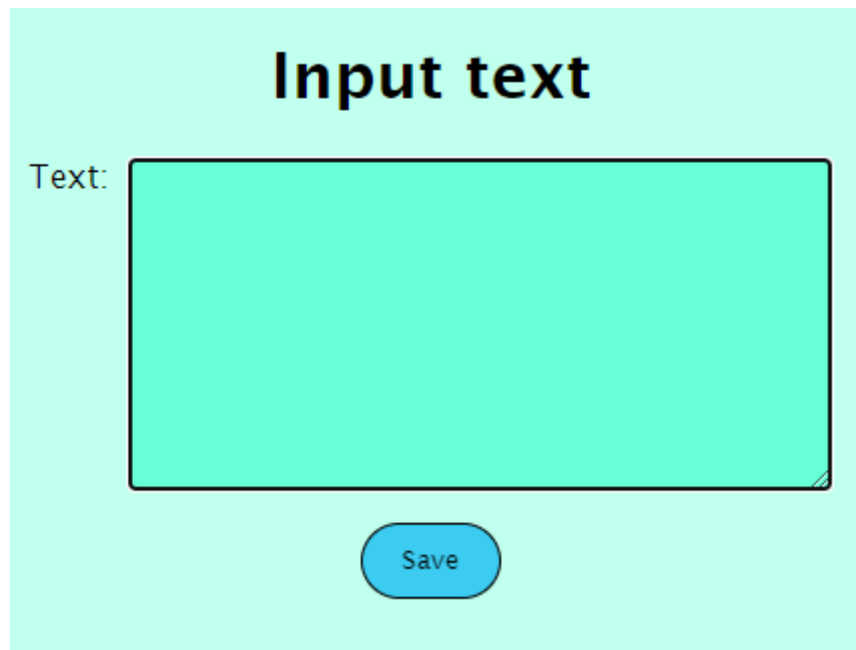


Рисунок 3.3 - Сторінка для введення

Друга сторінка відображує пости за введеним на попередній сторінці запитом користувача. Зверху можна побачити декілька вкладень: позитивний, нейтральний та негативний. За замовчуванням обрано "позитивний". Але при натисканні на якийсь з двох інших відбувається виведення постів відповідної категорії. Кожен виведений пост містить в собі аналогічну інформацію: дату публікації, автора, скорочений текст посту та посилання на першоджерело (в даному випадку відповідний пост в мережі Twitter). Поруч з вкладками для вибору відповідного настрою, є посилання "Return to input". Після кліку на нього користувач повертає до початкової сторінки з можливістю введення нового запиту. Користувач може ввести новий запит або ж повторити попередній, якщо його цікавлять нові публікації на ту саму тему.



Рисунок 3.4 - Сторінка з результатами фільтрації за настроєм

3.4. Оцінка точності роботи

Для навчання нейронної мережі, аби вона мала змогу достовірно визначати емоційне забарвлення тексту (в темі курсової роботи це є пости з соціальних мереж), було взято вибірку з 20 тисяч постів. Для тестування зазвичай беруть 10% даних від тої кількості, яку використовують для навчання, тому для перевірки було також взято 2 тисячі постів. Найбільш частовикористовуваним параметром для перевірки успішності алгоритму є Accuracy, який являє собою відношення кількості правильно передбачених спостережень до загальної кількості спостережень. Тож було вирішено слідкувати саме за цим параметром в процесі навчання нейронної мережі. На рисунку 3.5 наведено графік покращення точності на тестовій вибірці даних.

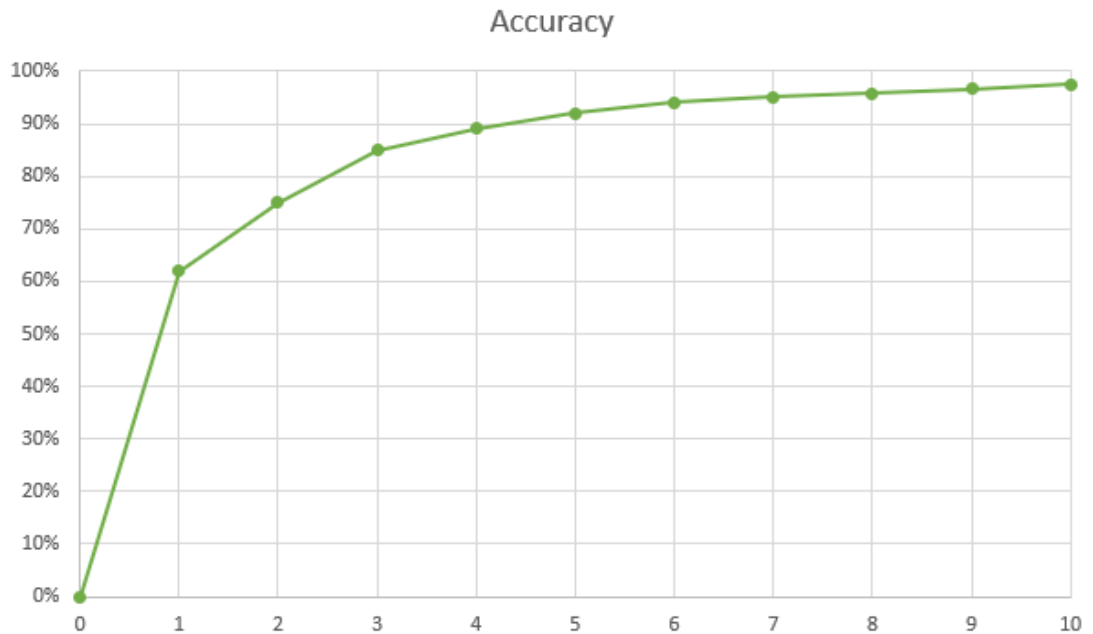


Рисунок 3.5 – Точність прогнозування на тестовій вибірці

В машинному навчанні існує таке поняття як перенавчання. В такому випадку нейронна мережа описує випадкову похибку або шум, замість взаємозв'язку, що лежить в основі даних. Тобто вона завчила тренувальний набір даних. Задля уникнення такої ситуації кількість ітерацій навчання даної моделі визначення емоційного забарвлення постів у соціальних мережах було обмежено до десяти. Бо при подальшому навчанні нейронна мережа продовжувала покращувати свою точність на тренувальній вибірці, але під час тестування, результати вже ставали нестабільні та в деякій мірі погіршувались. На рисунку 3.6 наведено графік покращення точності на тестовій вибірці даних при збільшенні кількості ітерацій.

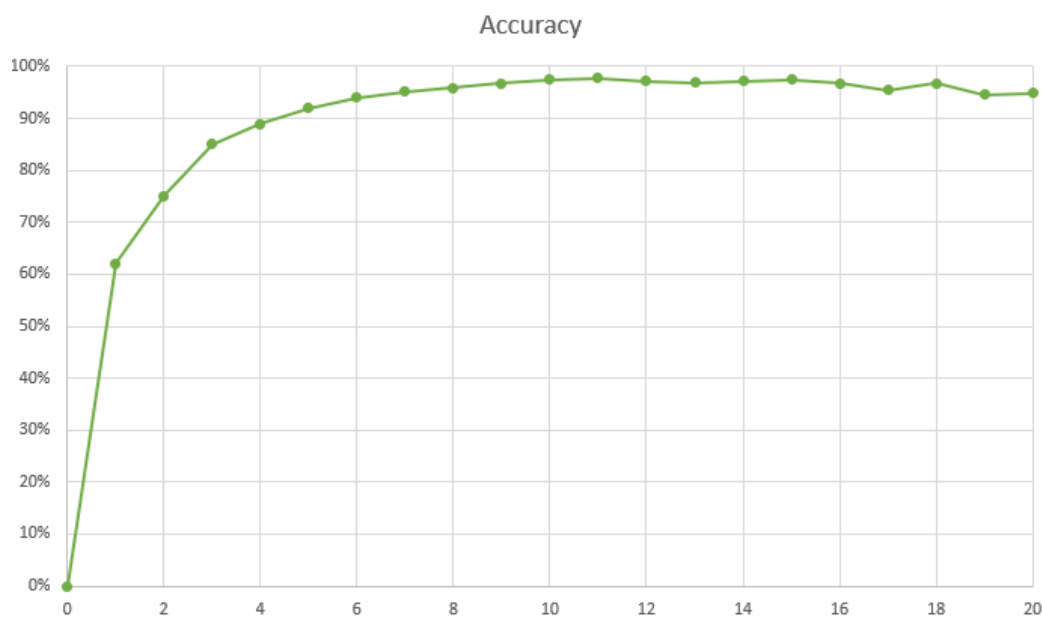


Рисунок 3.6 – Результати збільшення кількості ітерацій

3.5 Огляд процесу тестування

Для початку тестування необхідно ввести назву теми, яка цікавить. Для першого випадку було обрано тему “media”.

Рисунок 3.7 - Вибір теми постів для першого тесту

В результаті було отримано список з постами. Необхідно проаналізувати для кожної теми хоча б по одному посту.

Posts from Twitter - media

Sentiment: Positive

[Return to input](#) [Positive](#) [Neutral](#) [Negative](#)

Big 🗨️

May 19, 2023, 4:38 p.m.

Social media have you thinking everybody getting money til you hang with them

[👉 Link on original](#)

Ankur Sharma

May 19, 2023, 4:38 p.m.

Ekam Sanatan Bharat Dal had booked a free show of "The Kerala Story" on 18.05.2023 for Girl Students of Sanskrit Un...

[Link on original](#)

Alexandra Klinnik

May 19, 2023, 4:38 p.m.

Plenel says. "You are the first journalist to call me," he says over the phone, speaking to Variety. "That's intere...

[Link on original](#)

Светлана Флора 🇲Х

May 19, 2023, 4:38 p.m.

The racist in the white wrangler, Tracey Robert Blackwell III, wasn't hard to find at all, especially not when soci...

[Link on original](#)

Ron Filipkowski

May 19, 2023, 4:38 p.m.

The entire story that has been all over Fox and right-wing social media every day for a week of homeless veterans b...

[Link on original](#)

Рисунок 3.8 - Пости з позитивним настроєм на тему медіа

Серед позитивних постів було обрано другий пост, "Ekam Sanatan Bharat Dal booked a free show of "The Kerala Story" on 18.05.2023 for Girl of Sanskrit University". В перекладі це означає: "Екам Санатан Бхарат Дал замовив безкоштовний показ фільму "Історія Керали" на 18.05.2023 для студенток Санскритського університету". Це без сумнівів позитивний пост, адже в ньому розповідається про те як людина займається благодичністю, допомагає іншим.

Posts from Twitter - media

Sentiment: Neutral
 Return to input Positive Neutral Negative

May 19, 2023, 4:56 p.m.

अमरदास भल्ला
@adhicutting Do u think anybody believes media now? Even modi supoorters bhakt will run to exchange notes. [Link on original](#)

May 19, 2023, 4:56 p.m.

Jane E Fitzgerald
@twaddict03 I didn't say no-one was swayed, I said I don't know anyone who was swayed by the media. I teach resea... [Link on original](#)

May 19, 2023, 4:56 p.m.

Sanguinefemme ❤️
@MadcapGaffes @Concreteisland2 @AyoCaesar Calling activists irrational? I can see the media has worked its magic on you. [Link on original](#)

May 19, 2023, 4:56 p.m.

M.
Dhoni literally pushed a person who came to talk to him. Imagine if Kohli would have done this media would have ca... [Link on original](#)

May 19, 2023, 4:56 p.m.

YG FAMILY
#JISOO [ME] OFFLINE FAN SIGNING EVENT ▶ [Link on original](#)

Рисунок 3.9 - Пости з нейтральним настроєм на тему медіа

Для аналізу постів з нейтральним настроєм було обрано перший пост, “@adhicutting Do u think anybody believes media now? Even modi supoorters bhakt will run to exchange notes”. В перекладі це буде: “@adhicutting Думаєте, зараз хтось вірить ЗМІ? Навіть моді-супортери бхакт побіжать обмінюватися нотатками”. Це є просто роздуми деякої людини, які не несуть ні позитивного, ні негативного змісту, тобто загалом цей пост дійсно буде нейтральним.

Posts from Twitter - media

Sentiment: Negative
 Return to input Positive Neutral Negative

May 19, 2023, 4:56 p.m.

Willi023
@ProjectLincoln Fact is Fact Hitler was chosen as newsmaker of the year on Times magazine..a democrat run media outlet [Link on original](#)

May 19, 2023, 4:56 p.m.

God bless the USA! us 🇺🇸
@NSFWPoliticsNow @keithmp22 @tedcruz You people are insane. Again, it was weather-related. STEP AWAY FROM MAINSTREAM MEDIA. [Link on original](#)

Рисунок 3.10 - Пости з негативним настроєм на тему медіа

Серед негативних публікацій також можна проаналізувати перший пост, “@ProjectLincoln Fact is Fact Hitler was chosen as newsmaker of the year on Times magazine..a democrat run media outlet”. Він містить таке повідомлення: “@ProjectLincoln Факт є факт Гітлер був обраний ньюсмейкером року за версією журналу "Таймс" - видання, яким керують демократи”. Будь-яка людина, яка хоч трохи знайома з історією світу, розуміє що згадка про Гітлера в такому ключі точно не є чимось гарним. Скоріш за все, як і користувача, який виставив цей допис, у інших ця інформація викличе обурення або розпач, що є негативними почуттями.

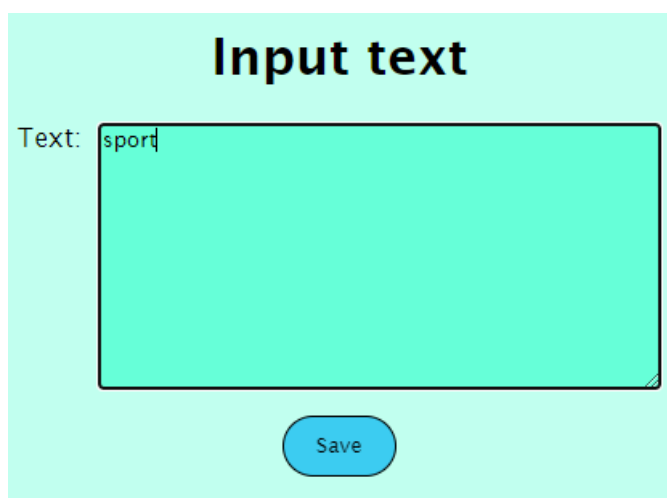
The image shows a web form with a light blue background. At the top, the text "Input text" is displayed in a bold, black font. Below this, there is a text input field with the label "Text:" to its left. The input field contains the word "sport". At the bottom center of the form, there is a blue, rounded rectangular button with the word "Save" written on it in white.

Рисунок 3.11 - Вибір теми постів для другого тесту

Для другого тестування було обрано тему “sport”.



Рисунок 3.8 - Пости з позитивним настроєм на тему спорту

Серед позитивних публікацій було обрано четвертий пост, “Support artists and writers by streaming the shows while available and buying what you can directly from them”. В перекладі це означає: “Підтримайте художників і письменників, переглядаючи їхні шоу, поки вони доступні, і купуючи все, що зможете, безпосередньо у них”. Користувач закликає робити добрі справи, тому це точно є позитивним дописом.

Posts from Twitter - sport

Sentiment: Neutral

Return to input Positive Neutral Negative

अमरदास भल्ला

May 19, 2023, 4:56 p.m.

@adhicutting Do u think anybody believes media now? Even modi supoorters bhakt will run to exchange notes. [Link on original](#)

Jane E Fitzgerald

May 19, 2023, 4:56 p.m.

@twtaddict03 I didn't say no-one was swayed, I said I don't know anyone who was swayed by the media. I teach resea... [Link on original](#)

Sanguinefemme ❤️

May 19, 2023, 4:56 p.m.

@MadcapGaffes @Concreteisland2 @AyoCaesar Calling activists irrational? I can see the media has worked its magic on you. [Link on original](#)

M.

May 19, 2023, 4:56 p.m.

Dhoni literally pushed a person who came to talk to him. Imagine if Kohli would have done this media would have ca... [Link on original](#)

YG FAMILY

May 19, 2023, 4:56 p.m.

#JISOO [ME] OFFLINE FAN SIGNING EVENT ▶ [Link on original](#)

Рисунок 3.9 - Пости з нейтральним настроєм на тему спорту

Для аналізу постів з нейтральним настроєм було обрано другий пост, “@twtaddict03 I didn't say no-one was swayed, I said I don't know anyone who was swayed by the media.”. В перекладі це буде: “@twtaddict03 Я не казав, що ніхто не був під впливом, я сказав, що не знаю нікого, хто був би під впливом ЗМІ”. Цей пост скоріш за все є продовженням діалогу між двома особами, який для інших користувачів не несе абсолютно ніяких емоцій. Тому його дійсно можна вважати постом з нейтральним настроєм.



Рисунок 3.10 - Пости з негативним настроєм на тему спорту

Серед негативних публікацій можна проаналізувати четвертий пост, “As a teacher, I’d say, what about the gaslighting of the OTHER children in the class to pretend a boy is a girl or vice versa?”. Він містить таке повідомлення: “Як вчитель, я б запитав, а як щодо того, що інші діти в класі видають себе за хлопчика чи дівчинку, або навпаки?”. Те, що описано в цьому пості не є нормою, тому його без сумнівів можна віднести до негативних.

3.5 Висновки

Третій розділ був присвячений розробці програмного забезпечення для визначення емоційного забарвлення постів у соціальних мережах. Для застосунку було обрано мову програмування Python та основні бібліотеки Tensorflow та Django. Перша допомагає у створенні нейронної мережі, друга komponує інтерфейс у вигляді веб-сайту. Також було описано структурну схему програмних модулів. Основними файлами є api.py для роботи з Twitter API та nn.py для взаємодії зі створеною нейронною мережею. Окрім цього, було сформовано діаграму класу NN

з детальним поясненням кожної функції класу, так як саме вони відповідають за побудову і роботу всієї нейронної мережі.

Так як фінальний результат дипломної роботи можна вважати повноцінним застосунком, для нього необхідно було сформулювати керівництво користувача з детальним поясненням кожного кроку та взаємодії. І звичайно, перш ніж передавати цей додаток на використання, було проведено перевірку його коректної роботи, розглянуто декілька тест-кейсів, проаналізовано результат та визначено справність роботи застосунку.

Висновки

В даній дипломній роботі було досліджено суть задачі визначення емоційного забарвлення постів у соціальних мережах, її проблематику.

Перший розділ був присвячений аналізу процесу визначення емоційного забарвлення постів у соціальних мережах. Розглянуто роботи Роберта Стіна та Бінг Люя. Було наведено вже існуючі інформаційні системи для визначення емоційного забарвлення тексту, описано процес роботи одного з них з використанням методології IDEF0 та DFD. Також визначено основні вимоги до вхідних даних, функціональні та нефункціональні вимоги.

У другому розділі була робота над розробкою архітектури застосунку. Сформовано класифікаційні ознаки розроблюваної системи та побудовано карту процесів роботи застосунку. Також описано 2 основних бізнес-процеси: отримання даних за запитом користувача та процес навчання нейронної мережі. Побудовано діаграму процесу ЯК-БУДЕ та виділено моменти, яких немає у застосунках, розглянутих раніше. Також було побудовано архітектуру інформаційної системи визначення емоційного забарвлення постів у соціальних мережах, визначено основні підсистеми та модулі роботи. Останнім моментом цього розділу роботи було сформувавши інформаційне забезпечення розробленого застосунку для визначення емоційного забарвлення постів у соціальних мережах. В першу чергу це був словник даних з описом всіх змінних та яку інформацію вони в собі несуть.

Третій розділ був присвячений розробці програмного забезпечення для визначення емоційного забарвлення постів у соціальних мережах. Для застосунку було обрано мову програмування Python та основні бібліотеки Tensorflow та Django. Також було описано структурну схему програмних модулів та діаграму класу NN з детальним поясненням кожної функції класу. І звичайно, вже для готового застосунку необхідно було сформувавши керівництво користувача та провести перевірку його коректної роботи. В останній пункт було включено формування тест-кейсів, аналіз їх результат та визначення справності роботи застосунку.

В результаті виконання дипломної роботи було розроблено застосунок, який дозволяє визначити емоційне забарвлення постів у соціальних мережах.

Список використаних джерел

1. Antony Samuels. News Sentiment Analysis / Antony Samuels, John Mcgonical // University of Southern California, Caltech. – 2017.
2. Liu Bing. Sentiment Analysis and Opinion Mining, 2012.
3. Muhammad Irzam Liaqat. Sentiment analysis techniques, challenges, and opportunities: Urdu language-based analytical study / Muhammad Irzam Liaqat, Muhammad Awais Hassan, Muhammad Shoaib, Syed Khaldoon Khurshid, Mohamed A. Shamseldin // PeerJ Comput Sci. – 2022.
4. Phil Picton. What is a Neural Network? / Faculty of Technology, The Open University, UK – 1994.
5. Stine Robert A. Sentiment Analysis. Department of Statistics, University of Pennsylvania, Philadelphia, 2018.
6. Turney Peter D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In Proceedings of Annual Meeting of the Association for Computational Linguistics, 2002.
7. Walaa Medhat. Sentiment analysis algorithms and applications: A survey / Walaa Medhat, Ahmed Hassan, Hoda Korashy // School of Electronic Engineering, Canadian International College, Cairo Campus of CBU, Egypt – 2014.
8. Widodo Budiharto. Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis / Widodo Budiharto, Meiliana Meiliana // Journal of Big Data volume 5, Article number: 51 – 2018.
9. Xing Fang. Sentiment analysis using product review data / Xing Fang, Justin Zhan // Journal of Big Data volume 2, Article number: 5 – 2015.
10. Yelena Mejova. Sentiment Analysis: An Overview / Computer Science Department, University of Iowa – 2014.

Електронні ресурси

11. 15 of The Best Sentiment Analysis Tools [Електронний ресурс] – Режим доступу до ресурсу: <https://monkeylearn.com/blog/sentiment-analysis-tools/>

12. Brandwatch [Электронный ресурс] – Режим доступа до ресурсу:
<https://www.brandwatch.com/>
13. Django [Электронный ресурс] – Режим доступа до ресурсу:
<https://www.djangoproject.com/>
14. Lexalytics. Sentiment Analysis Explained [Электронный ресурс] – Режим доступа до ресурсу: <https://www.lexalytics.com/technology/sentiment-analysis/>
15. MeaningCloud [Электронный ресурс] – Режим доступа до ресурсу:
<https://www.meaningcloud.com/>
16. MonkeyLearn [Электронный ресурс] – Режим доступа до ресурсу:
<https://monkeylearn.com/>
17. Python 3.11.3 documentation [Электронный ресурс] – Режим доступа до ресурсу: <https://docs.python.org/3/>
18. Tensorflow [Электронный ресурс] – Режим доступа до ресурсу:
<https://www.tensorflow.org/about>
19. Social Searcher [Электронный ресурс] – Режим доступа до ресурсу:
<https://www.social-searcher.com/>