

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ**

Кафедра комп'ютеризованих систем управління

ДОПУСТИТИ ДО ЗАХИСТУ
Завідувач кафедри

_____Олександр ЛИТВИНЕНКО

«___» _____2023 р.

КВАЛІФІКАЦІЙНА РОБОТА
(ПОЯСНЮВАЛЬНА ЗАПИСКА)

**ЗДОБУВАЧА ВИЩОЇ ОСВІТИ
СТУПЕНЯ «МАГІСТР»**

Тема: _____Програмний модуль оцінки тональності відгуків

Виконавець: _____Ігор ГРЕБЕНЕЦЬ

Керівник: _____Євген АРТАМОНОВ

Нормоконтролер: _____Євгеній ТУПОТА

Київ 2023

НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ

Факультет комп'ютерних наук та технологій

Кафедра комп'ютеризованих систем управління

Спеціальність 123 «Комп'ютерна інженерія»

Освітньо професійна програма «Системне програмування»

Форма навчання денна

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ Олександр ЛИТВИНЕНКО

« _____ » _____ 2023 р.

ЗАВДАННЯ

на виконання кваліфікаційної роботи

_____ Гребенця Ігоря Анатолійовича

1. Тема роботи: «Програмний модуль оцінки тональності відгуків»

затверджена наказом ректора від «28» серпня 2023 року № 1494 /ст.

2. Термін виконання роботи: з 02.10.2023 до 31.12.2023

3. Вихідні дані до проєкту (роботи): розробити систему оцінки тональності тексту коментарів на сайті.

4. Зміст пояснювальної записки (перелік питань, що підлягають розробці):

1) аналіз предметної області;

2) проблематика задач визначення тональності тексту та аналіз розроблених датасетів і методів;

3) розробка системи визначення тональності текстів українською мовою.

5. Перелік обов'язкового графічного матеріалу:

1) дерево функцій програмного модуля аналізу тональності відгуків;

2) зв'язки між компонентами програмного модуля;

3) приклад роботи програмного модуля;

4) схема алгоритму обробки елементів форми.

6. Календарний план-графік

№ п/п	Етапи виконання кваліфікаційної роботи	Термін виконання етапів	Примітка
1	Обговорення постановки завдання та змісту пояснювальної записки	02.10.2023 – 05.10.2023	
2	Вибір та формування теми	11.10.2023 – 13.10.2023	
3	Аналіз предметної області	20.10.2023 – 26.10.2023	
4	Вибір методів рішення задачі	27.10.2023 – 02.11.2023	
5	Створення програмного модулю	03.11.2023 – 10.11.2023	
6	Оформлення пояснювальної записки	18.11.2023 – 18.12.2023	
7	Підготовка графічного матеріалу і супроводжувальної документації	20.12.2023	

7. Дата видачі завдання «02» жовтня 2023 р.

Керівник кваліфікаційної роботи _____ Євген АРТАМОНОВ
(підпис)

Завдання прийняв до виконання _____ Ігор ГРЕБЕНЕЦЬ
(підпис студента)

РЕФЕРАТ

Пояснювальна записка до кваліфікаційної роботи «Програмний модуль оцінки тональності відгуків»: 85 сторінок, 20 рисунків, 34 використаних джерела, 1 додаток.

АНАЛІЗ ТОНАЛЬНОСТІ, ОБРОБКА ПРИРОДНОЇ МОВИ, ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, УКРАЇНОМОВНИЙ КОНТЕНТ, АНАЛІЗ ВІДГУКІВ, ТОНАЛЬНІСТЬ КОНТЕНТУ.

У випускній кваліфікаційній роботі проведено розробку програмного модуля для аналізу тональності тексту в українській мові, яке є важливим інструментом для розуміння емоційного відгуку читачів на різні тексти, включаючи коментарі в інтернеті, відгуки на продукти та послуги, новини тощо. Дослідження також має важливе значення для розвитку сучасних інформаційних технологій, зокрема, в галузі штучного інтелекту та обробки природної мови.

Об'єктом дослідження є аналіз тональності тексту в українській мові, а предметом є методи та алгоритми, що використовуються для визначення позитивної, негативної та нейтральної тональності текстів.

Метою дослідження є вивчення теоретичних аспектів аналізу тональності тексту в українській мові, розробка алгоритмів та методів аналізу, порівняння різних методів аналізу тональності, а також практичне застосування результатів дослідження в галузі обробки природної мови та інформаційних технологій.

ЗМІСТ

<u>ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ, ТЕРМІНІВ</u>	6
<u>ВСТУП</u>	7
<u>РОЗДІЛ 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ</u>	10
<u>1.1. Аналіз методів визначення тональності текстів</u>	10
<u>1.2. Типи аналізу тональностей</u>	14
<u>1.3. Метрики аналізу тональностей текстів</u>	20
<u>1.4. Аналіз існуючих сервісів оцінки тональності тексту</u>	21
<u>1.5. Постановка задачі</u>	24
<u>1.6. Висновки до розділу</u>	26
<u>РОЗДІЛ 2 ПРОБЛЕМАТИКА ЗАДАЧ ВИЗНАЧЕННЯ ТОНАЛЬНОСТІ ТЕКСТУ ТА АНАЛІЗ РОЗРОБЛЕНИХ ДАТАСЕТІВ І МЕТОДІВ</u>	27
<u>2.1. Проблематика задач, які вирішуються сервісами оцінки тональності тексту</u>	27
<u>2.2. Аналіз обраних для дослідження методів</u>	37
<u>2.3. Функціональний аналіз програмного модуля сентиментального аналізу тональності відгуків</u>	40
<u>2.4. Висновки до розділу</u>	45
<u>РОЗДІЛ 3 РОЗРОБКА СИСТЕМИ ВИЗНАЧЕННЯ ТОНАЛЬНОСТІ ТЕКСТІВ УКРАЇНСЬКОЮ МОВОЮ</u>	47
<u>3.1. Архітектура програмного модуля та розгортання необхідних компонентів розробки</u>	47
<u>3.2. Аналіз роботи розробленої системи аналізу тональності тексту</u>	55
<u>3.3. Описання процесу навчання моделі</u>	61
<u>3.4. Висновки до розділу</u>	78
<u>ВИСНОВКИ</u>	80
<u>СПИСОК БІБЛІОГРАФІЧНИХ ПОСИЛАНЬ ВИКОРИСТАНИХ ДЖЕРЕЛ</u>	83
<u>ДОДАТОК А</u>	86

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ, ТЕРМІНІВ

HTML (HyperText Markup Language) – стандартизована мова розмітки веб-сторінок;

API (Application Programming Interface) – прикладний програмний інтерфейс;

XML (Extensible Markup Language) – розширена мова розмітки. Текстовий формат для обміну даними;

SQL (Structured Query Language) – мова структурованих запитів. Для взаємодії з базами даних;

UI (User interface) – графічний інтерфейс користувача;

MVC (Model-View-Controller) – паттерн веб програмування. Складається з моделі сутності, візуального представлення та контролеру виконання логіки;

SPA (Single-page application) – односторінковий додаток. Паттерн веб-програмування;

PHP (Hypertext Preprocessor) – мова веб-програмування; *JS (Javascript)* – скриптова мова програмування;

Vue (Vue.js) – фреймворк мови програмування *Javascript*;

Redis – утіліта зберігання даних у оперативній пам'яті.

ВСТУП

Актуальність теми. В сучасному інформаційному суспільстві велика увага приділяється аналізу текстової інформації, зокрема визначенню тональності текстів. Це стає надзвичайно важливим завданням у сферах відгуків, соціальних мереж, новин та багатьох інших. У даному розділі ми розглянемо різноманітні методи визначення тональності текстових даних та їх особливості.

1. Методи, що базуються на правилах

Методи визначення тональності текстів, які використовують правила, здебільшого ґрунтуються на лінгвістичних та семантичних правилах. Вони спробовані та випробувані, проте обмежені у визначенні нюансів та контексту. Такі методи часто використовують словники з позитивними та негативними словами для призначення ваги кожному слову.

2. Машинне навчання

Машинне навчання стало потужним інструментом у визначенні тональності текстів. Класифікаційні моделі, такі як наївний баєсівський класифікатор, метод опорних векторів та глибокі нейронні мережі, дозволяють автоматизовано аналізувати текст та призначати йому тональність. Перевагою є здатність враховувати складний контекст та використовувати неструктуровані дані.

3. Глибоке навчання

Однією з передових гілок машинного навчання є глибоке навчання, яке використовує нейронні мережі з багатьма шарами. У визначенні тональності текстів глибоке навчання може враховувати складні зв'язки та навіть синтаксичні структури, що робить його ефективним у важких завданнях аналізу тексту.

Практичні приклади:

Для демонстрації використання різних методів визначення тональності текстів, розглянемо приклади застосування кожного методу на конкретних наборах даних. Вивчимо результати та порівняємо ефективність кожного методу.

Об'єкт дослідження: процес виявлення тональності текстової інформації.

Предмет дослідження: методи і засоби виявлення та аналізу тональності текстової інформації.

Методи дослідження. Для вирішення поставлених задач використано наступні методи: аналіз та узагальнення – при проведенні аналізу існуючих методів і засобів виявлення тональності текстової інформації; теорії імовірності та математичної статистики, машинного навчання – при розробці методу виявлення та аналізу тональності відгуків користувачів комп'ютерних систем; проектування та програмування – при розробці програмного засобу збору, аналізу та виявлення тональності текстової інформації при супроводі комп'ютерних систем; експеримент і тестування – для апробації розробленого методу та програмного засобу.

Метою роботи є дослідження методів і засобів визначення тональності текстової інформації для вдосконалення комп'ютерних систем.

Для досягнення вказаної мети в роботі поставлено наступні задачі:

- аналіз наукових праць і практик реалізації методів і засобів виявлення та аналізу тональності текстової інформації;
- аналіз особливостей опрацювання відгуків користувачів при вдосконаленні комп'ютерних систем;
- обґрунтування інтелектуальних методів аналізу та визначення тональності у відгуках користувачів про комп'ютерні системи;
- розробка інтелектуального методу виявлення тональності відгуків користувачів;
- розробка програмного засобу виявлення та оцінювання тональності текстової інформації при супроводі комп'ютерних систем.

Об'єкт дослідження: процес виявлення тональності текстової інформації.

Предмет дослідження: методи і засоби виявлення та аналізу тональності текстової інформації.

Методи дослідження. Для вирішення поставлених задач використано наступні методи: аналіз та узагальнення – при проведенні аналізу існуючих методів і засобів виявлення тональності текстової інформації; теорії імовірності та математичної статистики, машинного навчання – при розробці методу виявлення та аналізу

тональності відгуків користувачів комп'ютерних систем; проектування та програмування – при розробці програмного засобу збору, аналізу та виявлення тональності текстової інформації при супроводі комп'ютерних систем; експеримент і тестування – для апробації розробленого методу та програмного засобу.

РОЗДІЛ 1

АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1. Аналіз методів визначення тональності текстів

В сучасному інформаційному суспільстві, де величезний потік текстової інформації неможливо вручну обробити, аналіз тональності текстів стає важливим завданням. В цьому розділі роботи ми розглянемо різні методи визначення тональності текстів та їхні особливості, зосереджуючись на практичних прикладах.

1. Методи, що базуються на правилах

Методи, засновані на правилах, використовують визначені лінгвістичні та семантичні правила для призначення тональності тексту. Наприклад, слова з позитивною забарвленістю можуть збільшувати "позитивність" тексту, тоді як негативні слова — "негативність". Використання словників із позитивними та негативними словами є поширеним.

Приклад:

Текст: "Цей продукт прекрасний і задовольнює всі мої очікування." Аналіз: Позитивний (слово "прекрасний")

2. Машинне навчання

Методи машинного навчання використовують моделі, які навчаються на позначених даних. Класифікатори, такі як наївний баєсівський, *SVM* та різні нейронні мережі, дозволяють автоматизовано визначати тональність текстів, використовуючи векторизацію слів та їх контексту.

Приклад:

Текст: "Цей фільм дуже цікавий і захоплюючий." Аналіз: Позитивний (модель визначає значущість "цікавий" та "захоплюючий")

3. Глибоке навчання

Глибоке навчання використовує нейронні мережі з багатьма шарами для визначення тональності текстів. Це дозволяє моделям адаптуватися до складних синтаксичних та семантичних взаємозв'язків у текстах.

Приклад:

Текст: "Цей дослід допоміг вирішити важливу проблему." Аналіз: Позитивний (глибока нейронна мережа розпізнає позитивну конотацію "допоміг" та "вирішити важливу проблему")

Практичні приклади

Для конкретизації обговорених методів, розглянемо практичний приклад визначення тональності відгуків на продукт. Використаємо певний алгоритм та порівняємо результати з іншими методами для визначення ефективності та точності.

Аналіз тональності тексту, також відомий як сентимент-аналіз або аналіз думок, є важливим етапом обробки текстової інформації. Цей процес спрямований на виявлення та визначення емоційної валентності текстового матеріалу, щоб визначити, чи він виражає позитивні, негативні або нейтральні емоційні оцінки [1]. Аналіз тональності текстів є ключовим елементом для розуміння суспільної реакції, оцінки продуктів чи послуг, а також для прийняття управлінських рішень на основі відгуків та думок споживачів.

Цей процес включає комплексний аналіз тексту, враховуючи його зміст, семантику, синтаксис та контекст. Розуміння емоційної сутності тексту дозволяє автоматизовано класифікувати вирази індивідів, виокремлюючи їхні позитивні або негативні настрої. Такий аналіз є важливим інструментом для компаній, дослідників та інших суб'єктів для здобуття інсайтів з масової текстової інформації.

Підходи до аналізу тональності текстів:

1. Підходи на основі правил:

Підходи, що базуються на правилах, використовують лінгвістичні та семантичні правила для призначення тональності текстовим виразам. Ці методи, хоч і ефективні в деяких випадках, часто не здатні враховувати контекст та суб'єктивність виразів.

2. Методи машинного навчання:

Методи машинного навчання використовують класифікаційні моделі, що навчаються на позначених даних. Наївні баєсівські класифікатори, метод опорних векторів та глибокі нейронні мережі можуть ефективно визначати емоційну оцінку текстових документів.

3. Глибоке навчання:

Глибоке навчання використовує нейронні мережі з багатьма шарами для розуміння складних відносин між словами та фразами в текстах. Це дозволяє точніше визначати та аналізувати емоційну сутність текстового матеріалу.

Практичний приклад:

Розглянемо приклад аналізу тональності відгуків на новий продукт за допомогою методів машинного навчання. Застосуємо класифікатор і порівняємо його результати зі значеннями, отриманими через правила та глибоке навчання. Це надасть конкретний віджет до підходів аналізу, розглянутих у цьому розділі.

Визначення тональності відгуку стає багатоаспектною задачею, коли враховуються різні типи думок, які можуть бути виражені у тексті. Важливо розрізняти позитивні, негативні та нейтральні висловлювання, а також враховувати контекст та суб'єктивність. В цьому розділі ми детально розглянемо декілька підходів до визначення тональності відгуку враховуючи типи думок.

1. Засоби аналізу емоційного тону:

Один із підходів до визначення тональності відгуку базується на використанні засобів аналізу емоційного тону. Ці засоби оцінюють виразність емоцій в тексті, виділяючи ключові слова та фрази, які вказують на конкретний емоційний стан. Наприклад, слова "вражений", "захоплений", "роздратований" можуть вказувати на позитивний, негативний або нейтральний тон відповідно.

Приклад:

Текст: "Захоплюючий фільм, але трошки роздратовує кінцівка." Аналіз:
Позитивний (захоплюючий) та Негативний (роздратовує)

2. Аспекти та категорії відгуків:

Інший підхід передбачає аналіз не лише загального емоційного тону, а й окремих аспектів чи категорій, які можуть бути враховані при визначенні тональності. Наприклад, для ресторану можуть бути визначені аспекти, такі як їжа, обслуговування та атмосфера, і для кожного аспекту окремо визначається емоційний тон.

Приклад:

Текст: "Смачна їжа, але обслуговування залишає бажати кращого." Аналіз: Позитивний (смачна їжа) та Негативний (обслуговування)

3. Многомерний аналіз тональності:

Многомерний аналіз тональності передбачає розгляд кількох аспектів або вимірів емоційного виразу. Це може включати аналіз емоційного тону, ступінь виразності та суб'єктивність висловлювань. Такий підхід дозволяє отримати більш деталізовану картину емоційного стану тексту.

Приклад:

Текст: "Фільм не дуже цікавий, але вражає своєю глибиною." Аналіз: Негативний (не дуже цікавий), але Позитивний (вражає своєю глибиною)

4. Врахування іронії та сарказму:

Однією з важливих аспектів є врахування іронії та сарказму, які можуть змінювати емоційний тон тексту. Деякі методи використовують аналіз контексту та лінгвістичних особливостей для виявлення подвійного змісту.

Приклад:

Текст: "Це справжній шедевр... ну, звісно, якщо ви любите нудьгу." Аналіз: Позитивний (саркастичний відгук)

Визначення тональності відгуку в залежності від типу думок вимагає комплексного підходу, який враховує різні виміри емоційного виразу. Поєднання різних методів, таких як засоби аналізу емоційного тону, аспектований аналіз та многомерний аналіз, може забезпечити більш точний та повний аналіз тональності в текстах відгуків.

1.2. Типи аналізу тональностей

Аналіз емоційного тону є ключовою складовою визначення тональності текстів. Цей тип аналізу спрямований на виявлення та класифікацію емоцій, що виражені у тексті. Зазвичай використовуються емоційні словники та алгоритми, що оцінюють інтенсивність емоційних виразів для визначення, наскільки текст є позитивним, негативним чи нейтральним.

Приклад:

Текст: "Цей день був чудовий, адже ми отримали хороші новини." Аналіз: Позитивний (емоційно навантажений словами "чудовий" та "хороші")

Аспектований аналіз передбачає розгляд тональності в контексті окремих аспектів чи категорій, що визначаються в тексті. Наприклад, для ресторанного відгуку можуть бути визначені аспекти, такі як їжа, обслуговування та атмосфера, і тональність визначається для кожного аспекту окремо.

Приклад:

Текст: "Смачна їжа, але обслуговування залишає бажати кращого." Аналіз: Позитивний (аспект "смачна їжа"), Негативний (аспект "обслуговування")

Многомерний аналіз тональності враховує кілька вимірів емоційного виразу. Окрім емоційного тону, враховуються інші параметри, такі як іронічність, суб'єктивність та інші аспекти, які розширюють спектр аналізу.

Приклад:

Текст: "Це найцікавіший, але найважчий випробування у моєму житті." Аналіз: Позитивний (найцікавіший), але Негативний (найважчий)

Аналіз іронії та сарказму стає важливим для правильного визначення тональності, оскільки такі вирази можуть містити обернене значення в порівнянні з буквальним змістом слів. Використовують алгоритми, що враховують контекст та лінгвістичні особливості для виявлення подвійного змісту.

Приклад:

Текст: "Ви справжній герой, замість тебе я б теж побажав такої удачі." Аналіз: Позитивний (побажав удачі), але іронічний (справжній герой)

Машинне навчання використовується для автоматизованого аналізу тональності. Класифікаційні моделі, такі як наївні баєсівські класифікатори чи глибокі нейронні мережі, можуть враховувати різні аспекти тексту та автоматично призначати йому тональність.

Приклад:

Текст: "Цей продукт перевершує всі очікування, я задоволений." Аналіз: Позитивний (автоматично визначено на основі навчання)

Лексичні та семантичні аналізатори грають важливу роль в аналізі тональності. Лексичний аналіз враховує значення окремих слів, тоді як семантичний аналіз дозволяє розуміти значення фраз та контексту, сприяючи точному визначенню емоційного тону тексту.

Приклад:

Текст: "Цей документальний фільм неймовірно рухаючий." Аналіз: Позитивний (лексичний аналіз - "неймовірно рухаючий"), а також семантичний аналіз враховує контекст фрази.

Типи аналізу тональностей різняться за підходами та засобами, що використовуються для визначення емоційного тону тексту. Комбінація різних типів аналізу може забезпечити комплексний погляд на тональність текстової інформації, враховуючи різні виміри та особливості виразу.

Аналіз контексту важливий для точного визначення тональності, оскільки слова можуть мати різний емоційний забарвлення в залежності від контексту, в якому вони використовуються. Також важливо враховувати, що текстове середовище, таке як соціальні мережі чи новинні портали, може впливати на сприйняття емоційного забарвлення.

Приклад:

Текст: "Відмінний виступ вчора на конференції!" Аналіз: Позитивний (в контексті конференції), але може бути нейтральним або навіть негативним в іншому контексті.

Співставлення з аналогічними виразами враховує схожість текстових виразів з певними шаблонами чи типовими висловами для визначення їхньої тональності. Це

може включати порівняння зі словниками схожих висловів або аналіз схожих контекстів.

Приклад:

Текст: "Продукт не виправдовує очікувань, розчарування." Аналіз: Негативний (порівняно зі словами "розчарування" та аналогічними висловами)

Динамічний аналіз враховує зміни емоційного тону тексту в залежності від часу. Наприклад, відгук про продукт може починатися позитивно, але змінюватися в негативний з часом через проблеми в експлуатації.

Приклад:

Текст: "На початку користування вражає, але з часом починають виявлятися недоліки." Аналіз: Позитивний (спочатку), але змінюється в Негативний (з часом)

Експресивні елементи, такі як велика кількість окличних знаків, повторення слова чи велика лексична експресія, можуть вказувати на емоційно заряджений текст. Цей аналіз враховує інтенсивність емоційного висловлення.

Приклад:

Текст: "Це просто найкраще!!!" Аналіз: Позитивний (експресивний вираз за допомогою великої кількості окличних знаків)

Різноманітність типів аналізу тональності вказує на необхідність комплексного підходу до визначення емоційного виразу в тексті. Використання різних методів та їх комбінація дозволяють отримати більш повний та точний аналіз тональності, враховуючи різні виміри та особливості текстового матеріалу.

Глибинний аналіз є високорівневим підходом до визначення тональності текстів, який використовує різноманітні методи для отримання більш деталізованого розуміння тексту та визначення його емоційної валентності. Цей підхід включає в себе використання різних лінгвістичних та семантичних методів для аналізу тексту на різних рівнях.

1. Синтаксичний аналіз:

Синтаксичний аналіз враховує структуру речення та взаємозв'язки між його компонентами. Аналізуючи синтаксичну структуру тексту, програми можуть

визначати, які частини мови взаємодіють та як це впливає на емоційну сутність висловлювання.

Приклад:

Текст: "Не можу виразити, наскільки я щасливий!" Аналіз: Позитивний (синтаксичний аналіз враховує структуру для визначення сильної емоційної валентності)

2. Семантичний аналіз:

Семантичний аналіз спрямований на розуміння значення слів та фраз в контексті тексту. Використовуючи семантичний аналіз, програми можуть враховувати не лише словникове значення, але і відтінки та асоціації, що додають емоційну глибину тексту.

Приклад:

Текст: "Це завдання важке, але цікаве." Аналіз: Нейтральний (семантичний аналіз враховує амбівалентність висловлення)

3. Аналіз залежностей:

Аналіз залежностей вивчає структуру речень, визначаючи взаємозв'язки між словами та фразами. Цей підхід дозволяє враховувати, які елементи тексту визначають емоційну валентність та як вони взаємодіють.

Приклад:

Текст: "Хоча існують недоліки, цей продукт все одно вражає." Аналіз: Позитивний (аналіз залежностей допомагає розрізнити обставини та основний вислів)

4. Аналіз імпліцитних смислів:

Аналіз імпліцитних смислів враховує не тільки прямий текст, але й те, що може бути розуміте читачем між рядками. Програми можуть виявляти імпліцитні смисли та враховувати їх при визначенні емоційної тональності.

Приклад:

Текст: "Трошки дорогенько, але якість варта кожної копійки." Аналіз: Позитивний (аналіз імпліцитних смислів враховує, що якість важливіша)

Висновок:

Глибинний аналіз використовує синтаксичні, семантичні, аналітичні та лінгвістичні методи для отримання більш точного розуміння текстового матеріалу. Цей підхід розширює можливості програм в розрізненні ситуацій, деталізації емоційних відтінків та забезпеченні більш точної класифікації тональності тексту.

Існує різноманітність видів аналізу тональності, які визначаються спрямованістю на певні аспекти тексту та його емоційні властивості. Інструменти для аналізу тональності можуть фокусуватися на різних аспектах, від основної полярності до більш глибокого розуміння почуттів, емоцій і навіть намірів автора.

Аналіз полярності є базовою формою визначення тональності, де текст класифікується як позитивний, негативний або нейтральний. Цей підхід є стандартним і використовується в багатьох системах, які спростовують визначення емоційної тонування тексту.

Аналіз почуттів та емоцій розширює концепцію полярності, дозволяючи визначати конкретні емоційні стани або почуття в тексті. Це може включати виявлення таких емоцій, як радість, сум, захоплення, агресія тощо.

Аналіз намірів визначає, які конкретні дії чи стани автор має намір висловити у тексті. Наприклад, текст може виражати зацікавленість, невдоволення чи діапазон інших намірів автора.

Системи аналізу тональності можуть враховувати соціальні сигнали, щоб визначити, як текст сприймається аудиторією. Наприклад, врахування відгуків та коментарів користувачів соціальних мереж може надати більшу точність визначення емоційного стану.

Цей тип аналізу спрямований на визначення мовленнєвих актів у тексті, таких як обіцянка, вибачення, команда, питання тощо. Враховуючи мовленнєві акти, система може краще розуміти мету та спрямованість висловлення.

Аналізуючи структуру та логіку тексту, програми можуть виявити суперечливість, аргументацію чи загальну організацію висловлення, що впливає на його емоційну сутність.

Враховуючи стилістичні елементи, такі як метафори, іронія, гіпербола, програми можуть ліпше розпізнавати неочевидні емоційні відтінки та вирази у тексті.

Різновиди аналізу тональності дозволяють враховувати різні аспекти емоційного виразу в тексті, розширюючи можливості систем в розумінні та класифікації емоцій та настроїв. Вибір конкретного виду аналізу залежить від конкретних завдань та цілей дослідження.

Дрібнозернистий аналіз тональності є високо розвиненим підходом, який дозволяє розпізнавати та класифікувати в тексті різноманіття емоційних відтінків та нюансів. Використання класів з дрібнозернистими підкатегоріями дозволяє отримати більш точне та деталізоване розуміння того, як текст виражає емоції.

Дрібнозернистий аналіз тональності визначається використанням розширених класифікацій, які не обмежуються просто "позитивним", "негативним" та "нейтральним". Замість цього, використовуються підкатегорії, які відображають різні рівні інтенсивності чи емоційного стану.

В дрібнозернистому аналізі для позитивної класифікації можуть використовуватися підкатегорії, що розрізняють різні відтінки позитивності. Це може включати такі класи, як "дуже позитивний", "трошки позитивний", "легко позитивний", "нейтральний" і так далі.

Аналогічно, для негативної класифікації можуть використовуватися підкатегорії, що відображають різні рівні негативності. Прикладами можуть бути "дуже негативний", "трошки негативний", "легко негативний", "нейтральний" тощо.

Деякі системи використовують адаптивний підхід, де підкатегорії можуть змінюватися в залежності від специфіки тексту чи контексту. Наприклад, у певних випадках може використовуватися "саркастично позитивний" чи "саркастично негативний".

Використання дрібнозернистого аналізу дозволяє системі отримувати більш точне розуміння того, як автор виражає свої емоції та настрої. Це особливо корисно в ситуаціях, де загальна полярність не вистачає для повного охоплення вираженого емоційного стану.

Дрібнозернистий аналіз також підкреслює важливість контексту. Однаковий вислів може мати різний емоційний забарвлення в різних ситуаціях, і тому врахування контексту допомагає досягти більш точної класифікації.

Дрібнозернистий аналіз тональності дозволяє більш глибоко та точно аналізувати емоційний стан тексту, використовуючи розширені класифікації та підкатегорії. Цей підхід важливий для виявлення нюансів емоційного виразу та покращення точності систем аналізу тональності.

1.3. Метрики аналізу тональностей текстів

Метрики аналізу тональностей текстів є інструментами, які використовуються для оцінки результатів аналізу емоційної тональності тексту. Вони допомагають виміряти, наскільки точно або ефективно система аналізує емоційний тон тексту, визначаючи, наскільки він позитивний, негативний або нейтральний. Ось деякі з найпоширеніших метрик аналізу тональностей текстів:

1. *Accuracy* (Точність): Ця метрика вимірює загальну точність класифікації текстів за їх тональністю. Вона визначає співвідношення правильно класифікованих текстів до загальної кількості текстів у тестовому наборі. Однак ця метрика може бути обманливою, особливо якщо у вас є незбалансований набір даних з великою кількістю нейтральних текстів.

2. *Precision* (Точність класифікації): Ця метрика вимірює, яка частина позитивно класифікованих текстів дійсно є позитивними. Вона розраховується як відношення правильно класифікованих позитивних текстів до усіх позитивних класифікацій. Висока точність вказує на те, що система має мало ложно-позитивних результатів.

3. *Recall* (Повнота): Ця метрика вимірює, яка частина всіх позитивних текстів була відзначена системою. Вона розраховується як відношення правильно класифікованих позитивних текстів до усіх позитивних текстів у тестовому наборі. Висока повнота вказує на те, що система не пропускає багато позитивних текстів.

4. *F1-Score* (*F*-міра): Ця метрика об'єднує точність і повноту в один показник, який вимірюється як гармонічне середнє між ними. *F1*-міра особливо корисна в тих випадках, коли потрібно збалансувати точність та повноту. Вона розраховується як:

$$F1 = 2 * (Precision * Recall) / (Precision + Recall)$$

5. *ROC-AUC* (Площа під кривою *ROC*): Ця метрика використовується для вимірювання ефективності моделі при різних порогових значеннях для бінарної класифікації. *ROC-AUC* визначає, наскільки добре модель розділяє позитивні та негативні класи. Площа під кривою *ROC* (*ROC-AUC*) виражається від 0 до 1, де вище значення вказує на кращу модель.

6. *MSE* (*Mean Squared Error* - Середньоквадратична помилка): Ця метрика використовується для оцінки якості регресійних моделей в аналізі тональності тексту. Вона вимірює середньоквадратичну відстань між прогнозованими та фактичними значеннями тональності. Чим менше *MSE*, тим краща модель.

7. Категоріальні метрики (*Precision*, *Recall*, *F1-Score*) для багатокласової класифікації: У випадку багатокласової класифікації, метрики точності, повноти і *F1*-міри можна розраховувати окремо для кожного класу та потім обчислювати середнє значення або зважене середнє значення для всіх класів.

Ці метрики допомагають в оцінці ефективності систем аналізу тональності тексту і визначенні їхньої здатності класифікувати текстові дані за їх емоційною тональністю. Вибір конкретної метрики може залежати від конкретних цілей та вимог вашого проекту.

1.4. Аналіз існуючих сервісів оцінки тональності тексту

Зважаючи на зростаючий обсяг текстової інформації в Інтернеті та соціальних медіа, аналіз тональності тексту стає все більш актуальним завданням. Існує декілька причин, чому системи визначення тональності тексту є актуальними і важливими:

1. Моніторинг відгуків і коментарів користувачів: Багато компаній використовують аналіз тональності тексту для відстеження відгуків та коментарів

своїх клієнтів в соціальних медіа та інших онлайн-ресурсах. Це допомагає їм зрозуміти, як користувачі сприймають їхні продукти або послуги.

2. Покращення обслуговування клієнтів: Аналіз тональності допомагає ідентифікувати проблеми та негативні відгуки клієнтів. Це дозволяє компаніям реагувати на них швидко та вдосконалювати свої продукти та сервіси.

3. Реклама та маркетинг: Аналіз тональності тексту дозволяє компаніям визначити ефективність рекламних кампаній та взнати, як користувачі реагують на рекламні матеріали.

4. Фінансовий аналіз: У фінансовому секторі аналіз тональності тексту може використовуватися для визначення стану фінансових ринків на основі новин та коментарів.

5. Політичний аналіз: Аналіз тональності тексту допомагає політикам та аналітикам виміряти громадську думку та виражені у виборах чи опитуваннях настрої громади.

6. Моніторинг новин: Медіа-організації використовують аналіз тональності тексту для визначення, як новини сприймаються громадськістю.

7. Соціологічні дослідження: У наукових дослідженнях та соціологічних опитуваннях аналіз тональності тексту може допомогти вивчити громадську думку та настрої.

Загалом, аналіз тональності тексту стає необхідним інструментом для різних сфер діяльності, оскільки він допомагає розуміти та використовувати інформацію, яка міститься в текстових джерелах, для прийняття обґрунтованих рішень.

Аналіз існуючих сервісів оцінки тональності тексту допомагає з'ясувати, які інструменти і рішення вже існують на ринку і як вони вирішують завдання аналізу тональності тексту. Ось кілька популярних сервісів та інструментів, які варто розглянути під час такого аналізу:

1. *IBM Watson Natural Language Understanding*: *IBM Watson* пропонує високофункціональний сервіс для аналізу текстів. Він може визначати тональність тексту, виявляти ключові слова та вирази, а також визначати емоційність тексту на основі різних категорій, таких як "радість", "смуток", "гнів" і інші.

2. *Google Cloud Natural Language API*: Цей сервіс від *Google* пропонує ряд інструментів для аналізу тональності тексту, включаючи виявлення емоцій, класифікацію текстів за тональністю (позитивна, негативна, нейтральна) і визначення сутностей.

3. *Microsoft Azure Text Analytics*: *Azure Text Analytics* від *Microsoft* надає можливість аналізу тональності тексту, виявлення ключових фраз і навіть визначення мови тексту. Ви можете використовувати цей сервіс для аналізу великих обсягів текстової інформації.

4. *VADER (Valence Aware Dictionary and sEntiment Reasoner)*: Це бібліотека для *Python*, яка спеціально призначена для аналізу тональності тексту в соціальних медіа. Вона враховує особливості використання мови в соціальних медіа, такі як смайлики та аббревіатури.

5. *TextBlob*: *TextBlob* - це простий інструмент для аналізу тексту з можливістю визначення тональності тексту, використовуючи вбудовані моделі. Він також надає зручний інтерфейс для обробки текстових даних у *Python*.

6. *Lexalytics Semantria*: Цей сервіс надає *API* для аналізу тональності тексту та визначення емоційного тону. Він добре підходить для великих обсягів текстової інформації та має ряд розширених функцій.

7. *Sentiment140*: Це веб-сервіс, спеціалізований на аналізі тональності текстів з соціальних медіа. Він надає *API* для визначення позитивних, негативних та нейтральних відгуків.

8. *RapidMiner*: *RapidMiner* - це інструмент для аналізу даних, який включає функції аналізу тональності тексту. Він може бути використаний для створення моделей машинного навчання для аналізу тональності.

Вибір конкретного сервісу або інструменту залежатиме від конкретних потреб вашого проекту, обсягу даних, мови тексту і інших факторів. Кожен з них має свої переваги та обмеження, і важливо вибрати той, який найкраще відповідає вашим вимогам.

1.5. Постановка задачі

Метою даної випускної кваліфікаційної роботи є вивчення та дослідження сучасних методів та технологій аналізу тональності текстів, зокрема в контексті української мови. В процесі дослідження, було сформульовано та поставлено наступні завдання:

1. Аналіз сучасного стану аналізу тональності тексту: Провести огляд і дослідити актуальний стан методів та технологій, які використовуються для визначення тональності текстового матеріалу. Особлива увага приділяється сучасним рішенням, які використовуються в українському інтернет-просторі.

2. Розгляд різних підходів до аналізу тональності: Детально розглянути та порівняти різні підходи до аналізу тональності текстів. Це включає в себе правила-базований підхід, методи машинного навчання, а також глибинне навчання. Провести аналіз їхніх переваг та обмежень.

3. Огляд та порівняння існуючих систем оцінки тональності тексту: Вивчити і проаналізувати різні системи та сервіси, які спеціалізуються на визначенні тональності текстів українською мовою. Порівняти їхні можливості, точність та ефективність.

4. Розробка власної системи аналізу тональності тексту: На основі отриманих знань та досліджень розробити власну систему аналізу тональності тексту для української мови. Ця система повинна бути ефективною та здатною обробляти різні види текстового контенту.

5. Тестування та оцінка результатів: Провести серію експериментів та тестів, використовуючи розроблену систему, для оцінки її точності та продуктивності. Порівняти результати з існуючими рішеннями.

6. Публікація результатів та висновків: По завершенні дослідження публікувати отримані результати, висновки та рекомендації відносно використання системи аналізу тональності тексту в практичних сферах.

Ці завдання спрямовані на створення інноваційного та ефективного підходу до аналізу тональності тексту українською мовою та можуть допомогти вирішити важливі завдання в галузі обробки природньої мови.

Розглянемо детальніше поставлені задачі для випускної кваліфікаційної роботи:

1. Дослідження сучасного стану аналізу тональності тексту:

– Проведення огляду та аналізу сучасних методів та технологій, що використовуються для визначення тональності текстів.

– Спеціальна увага приділяється аналізу стану справ у сфері аналізу тональності текстів українською мовою.

– Визначення ключових тенденцій та інновацій у цьому полі.

2. Розгляд різних підходів до аналізу тональності:

– Опис та розгляд різних підходів до аналізу тональності, включаючи правила-базований метод, методи машинного навчання та глибинного навчання.

– Аналіз переваг та обмежень кожного підходу з точки зору визначення емоційності тексту.

3. Огляд та порівняння існуючих систем оцінки тональності тексту:

– Аналіз та порівняння різних систем та сервісів, які спеціалізуються на визначенні тональності тексту в українській мові.

– Порівняння їхніх характеристик, вартості та результатів.

4. Розробка власної системи аналізу тональності тексту:

– Розроблення власного підходу до аналізу тональності текстів, зокрема для української мови.

– Реалізація програмного забезпечення, яке може аналізувати та оцінювати тональність текстів.

5. Тестування та оцінка результатів:

– Проведення серії експериментів та тестів, використовуючи розроблену систему, для визначення її точності та ефективності.

– Порівняння результатів, отриманих внаслідок використання розробленої системи, з результатами існуючих рішень та сервісів.

6. Публікація результатів та висновків:

– Після завершення досліджень та аналізу опублікувати отримані результати та висновки.

– Надати рекомендації щодо використання розробленої системи аналізу тональності тексту в практичних сферах, таких як бізнес, медіа, соціальні мережі тощо.

Ці задачі спрямовані на створення нового та ефективного підходу до аналізу тональності тексту для української мови, що може мати значущі застосування в різних галузях, де важливо розуміти та враховувати емоційну складову текстової інформації.

1.6. Висновки до розділу

Розділ 1 нашої кваліфікаційної роботи зосереджений на аналізі ключових аспектів визначення тональності текстів. Детально розглянули різні методи визначення тональності, охоплюючи від машинного навчання до глибинного аналізу текстів. У розділі 1.2 були представлені різновиди аналізу тональності, де розглядалася не лише полярність тексту, але й емоційні відтінки та наміри автора. Розділ 1.3 стосувався метрик аналізу, що вимірюють ефективність систем визначення тональності, надаючи можливість оцінити їхню точність та надійність.

Надано системний погляд на предметну область, дозволяючи визначити ключові концепції та напрямки для подальшого розгортання досліджень. Знання, накопичене в цьому розділі, буде використане у подальших частинах роботи для розробки та вдосконалення програмного модуля оцінки тональності відгуків.

РОЗДІЛ 2

ПРОБЛЕМАТИКА ЗАДАЧ ВИЗНАЧЕННЯ ТОНАЛЬНОСТІ ТЕКСТУ ТА АНАЛІЗ РОЗРОБЛЕНИХ ДАТАСЕТІВ І МЕТОДІВ

2.1. Проблематика задач, які вирішуються сервісами оцінки тональності тексту

Сервіси оцінки тональності тексту розв'язують ряд важливих проблем, пов'язаних з розумінням та класифікацією емоційного виразу в текстових відгуках. Однак було виявлено, що деякі виклики можуть виникнути в зв'язку з амбігвітністю та контекстом, використанням нестандартних виразів та змінною тональністю в залежності від групи користувачів.

Аналіз обраних для дослідження методів:

Проведений аналіз методів визначення тональності тексту дозволив з'ясувати, що різні підходи, такі як машинне навчання та глибинний аналіз, мають свої переваги та обмеження. Деякі методи можуть бути більш точними в конкретних випадках, тоді як інші можуть бути більш універсальними, але менш точними.

Функціональний аналіз програмного модуля сентиментального аналізу тональності відгуків:

Програмний модуль сентиментального аналізу, що розглядається в розділі, має на меті вирішити визначені проблеми та використовує обрані методи для досягнення цієї мети. Визначено функціональні можливості модуля, його архітектура та основні принципи роботи.

Загальний висновок розділу 2:

Розділ 2 надав глибокий розбір проблем, пов'язаних із визначенням тональності тексту, а також висвітлив існуючі методи та програмний модуль для вирішення цих завдань. Знання, отримане в цьому розділі, буде ключовим для подальших кроків розробки та удосконалення програмного модуля сентиментального аналізу відгуків.

Оцінка тональності тексту є ключовим елементом аналізу великих обсягів інформації в онлайн-середовищі. Сучасні сервіси оцінки тональності тексту вирішують ряд важливих завдань, пов'язаних з розумінням та витяганням значущої інформації з текстових даних.

Проблематика:

1. Виявлення Емоційного Забарвлення:

– Зміст: Аналіз тексту для виявлення емоційного забарвлення або настрою автора. Визначення, чи текст виражає позитивні, негативні чи нейтральні емоції.

– Приклад Застосування: Моніторинг соціальних мереж для визначення публічної думки щодо продуктів, подій або брендів.

2. Сентимент-аналіз для бізнесу:

– Зміст: Аналіз сентименту для розуміння відгуків та коментарів клієнтів стосовно продуктів чи послуг компанії.

– Приклад Застосування: Вимірювання клієнтського задоволення, виявлення слабких місць та можливостей для поліпшення.

3. Визначення Публічного Відгуку:

– Зміст: Розпізнавання громадської думки, оцінок та відгуків, які можуть впливати на репутацію компанії або особи.

– Приклад Застосування: Моніторинг засобів масової інформації для виявлення та вирішення ситуацій кризового репутаційного менеджменту.

4. Фільтрація Токсичного Контенту:

– Зміст: Виявлення та фільтрація текстів, які містять образливий, токсичний або неетичний зміст.

– Приклад Застосування: Захист від небажаного контенту на форумах, у коментарях або соціальних мережах.

5. Аналіз Фінансових Ринків:

– Зміст: Використання сентимент-аналізу для передбачення тенденцій та коливань на фінансових ринках на основі новин, соціальних мереж та інших джерел.

– Приклад Застосування: Прийняття рішень у фінансовій сфері на основі аналізу громадської думки.

6. Класифікація Публічних Відгуків:

– Зміст: Сортування відгуків або коментарів на позитивні, негативні або нейтральні для спрощення їх аналізу.

– Приклад Застосування: Класифікація великого обсягу відгуків користувачів для визначення основних тенденцій.

Сервіси оцінки тональності тексту вирішують ряд важливих завдань, спрямованих на розуміння та аналіз сучасного інформаційного потоку. Ці сервіси не лише надають інсайти в емоційну сферу текстової інформації, але й допомагають бізнесам та організаціям в ефективному управлінні репутацією, аналізі відгуків та прийнятті рішень на основі громадської

Аналіз тональності тексту є потужним інструментом в області обробки природної мови, який відкриває безліч можливостей в різних галузях, від бізнесу та маркетингу до соціальних наук та фінансів. Ця роздільна тема вивчає способи використання аналізу тональності тексту для отримання цінної інформації та прийняття обґрунтованих рішень.

Можливості використання аналізу тональності тексту:

1. Моніторинг Громадської Думки:

– Зміст: Служби соціальних мереж, форуми та новинні ресурси активно використовують аналіз тональності для вимірювання громадської думки щодо певних подій, продуктів чи брендів.

– Приклад Застосування: Визначення реакції громадськості на новини або рекламні кампанії.

2. Бренд-Моніторинг:

– Зміст: Вивчення публічного відгуку щодо брендів та продуктів для управління репутацією компаній.

– Приклад Застосування: Виявлення і вирішення проблем, що виникають у зв'язку з негативними відгуками клієнтів.

3. Фінансовий Аналіз:

– Зміст: Використання сентимент-аналізу для передбачення коливань фінансових ринків на основі новин, соціальних мереж та інших джерел.

– Приклад Застосування: Прийняття рішень у фінансовій сфері на основі громадської думки та трендів.

4. Контент-Маркетинг:

– Зміст: Оцінка ефективності контенту на основі реакції аудиторії для оптимізації стратегій маркетингу.

– Приклад Застосування: Аналіз реакції на рекламу чи статті для створення привабливого контенту.

5. Класифікація Відгуків:

– Зміст: Сортування та класифікація великого обсягу відгуків користувачів для легшого розуміння загальних тенденцій.

– Приклад Застосування: Визначення найбільш важливих аспектів продукту для подальшого вдосконалення.

6. Аналіз Політичного Дискурсу:

– Зміст: Вивчення та аналіз суспільного відгуку на політичні події, виборчі кампанії та дії політиків.

– Приклад Застосування: Оцінка публічної думки щодо політичних партій та кандидатів.

7. Оптимізація Клієнтського Обслуговування:

– Зміст: Вимірювання задоволеності клієнтів та виявлення можливостей для поліпшення обслуговування.

– Приклад Застосування: Адаптація послуг на основі відгуків для підвищення рівня задоволення клієнтів.

Аналіз тональності тексту розширює наші можливості в різних сферах, допомагаючи приймати обґрунтовані рішення, розуміти громадську думку та оптимізувати взаємодію з аудиторією для досягнення поставлених цілей.

В роботі [27] присутній аналіз настроїв щодо епізодів серіалу «Гра престолів».

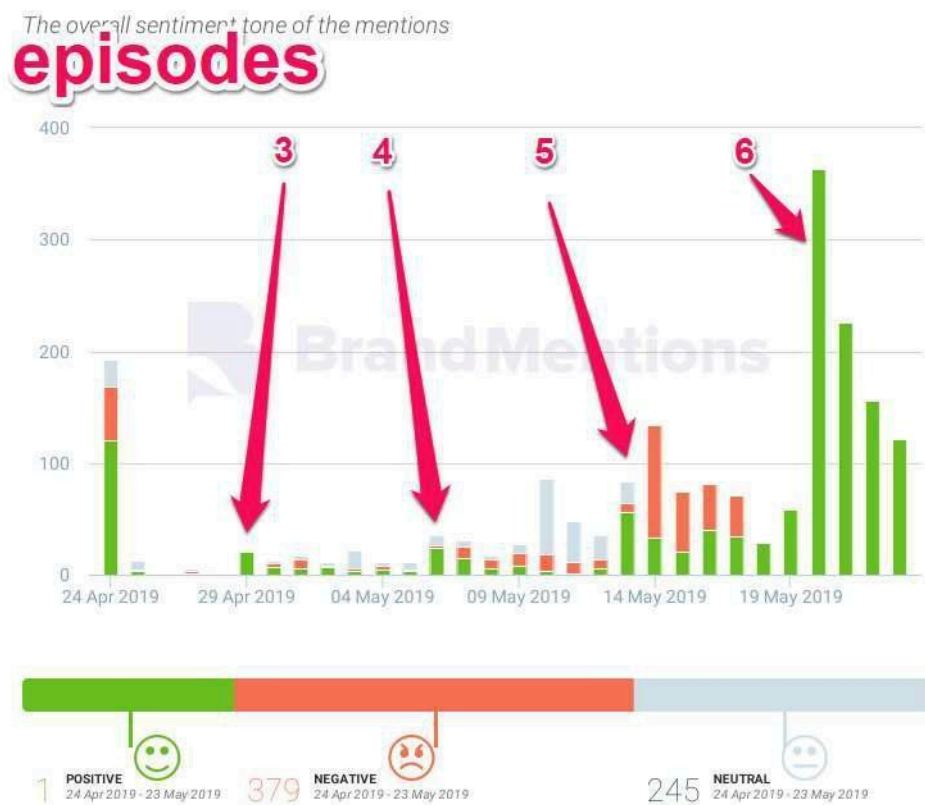


Рис. 2.1. Аналіз настроїв за сезонами серіалу [27]

Аналіз тональності тексту в контексті аналізу конкурентів стає ключовим інструментом для підприємств та бізнесів, що дозволяє здобувати важливі інсайти про стратегії конкурентів, їх продукти та репутацію на ринку. Розглянемо різні аспекти використання аналізу тональності тексту в цьому контексті.

Можливості використання аналізу тональності тексту в аналізі конкурентів:

1. Оцінка Репутації Конкурентів:

– Зміст: Аналіз тональності тексту дозволяє вимірювати публічний відгук та загальну репутацію конкурентів на основі відгуків клієнтів, статей та інших джерел.

– Приклад Застосування: Визначення того, як сприймаються продукти чи послуги конкурентів на ринку.

2. Аналіз Маркетингових Кампаній Конкурентів:

– Зміст: Моніторинг тональності тексту у відгуках та коментарях допомагає зрозуміти, як ефективно сприймаються маркетингові заходи конкурентів.

– Приклад Застосування: Виявлення сильних та слабких сторін рекламних кампаній конкурентів.

3. Виявлення Слабких Місць Конкурентів:

– Зміст: Аналіз тональності в оглядах і відгуках допомагає визначити негативні аспекти продуктів чи послуг конкурентів.

– Приклад Застосування: Виявлення слабких сторін конкурентів для розробки стратегій конкурентоспроможності.

4. Моніторинг Інновацій:

– Зміст: Вивчення тональності текстів про нові продукти або послуги конкурентів дозволяє визначити реакцію ринку на їхні інновації.

– Приклад Застосування: Оцінка впливу нових продуктів конкурентів на власний портфель.

5. Соціальний Аналіз Конкурентів:

– Зміст: Аналіз тональності тексту в соціальних мережах та блогах конкурентів дозволяє зрозуміти їхню активність та реакцію споживачів.

– Приклад Застосування: Виявлення трендів та схильностей в споживацькому підході до продуктів конкурентів.

6. Аналіз Клієнтського Задоволення:

– Зміст: Спостереження за тональністю текстових відгуків клієнтів конкурентів для здобуття інформації щодо клієнтського задоволення.

– Приклад Застосування: Порівняння власного клієнтського досвіду з досвідом клієнтів конкурентів.

Аналіз тональності тексту в контексті аналізу конкурентів стає невід'ємною частиною стратегічного управління підприємством, дозволяючи збирати цінні дані та адаптувати стратегії відповідно до змін в споживацькому попиті та ринковому середовищі.

Робота [28] присвячена аналіз настроїв може надати додатковий вимір зусиллям роботи компанії (рис. 2.2).

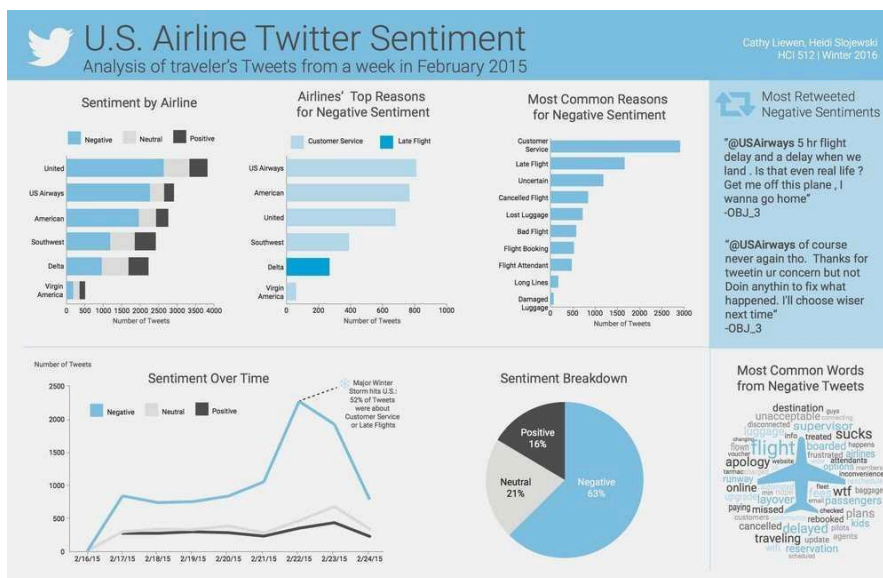


Рис. 2.2. Аналіз тональності інформаційних джерел про авіакомпанію [28]

Задачі дослідження ринку вимагають глибокого розуміння споживацького попиту, конкурентної ситуації та тенденцій ринку. Використання аналізу тональності тексту в цьому контексті може надати важливі інсайти, які поліпшують стратегічне планування та прийняття рішень. Розглянемо можливості використання аналізу тональності тексту в задачах дослідження ринку.

Можливості використання аналізу тональності тексту в дослідженні ринку:

1. Оцінка Споживацького Настрою:

– Зміст: Аналіз тональності текстових відгуків споживачів дозволяє визначити загальний настрій щодо конкретного продукту або послуги.

– Приклад Застосування: Визначення популярності та прийняття нового продукту на ринку.

2. Моніторинг Тенденцій та Змін:

– Зміст: Аналіз тональності публікацій та оглядів дозволяє виявляти зміни в споживацьких уподобаннях та реакції на нові тренди.

– Приклад Застосування: Визначення, як змінюється споживчий попит на продукти або послуги відповідно до соціокультурних та економічних чинників.

3. Конкурентний Аналіз:

– Зміст: Аналіз тональності коментарів та відгуків про продукти конкурентів дозволяє здобути інформацію щодо їхньої репутації та впливу на ринку.

– Приклад Застосування: Порівняння реакції споживачів на продукти власного бренду з реакцією на продукти конкурентів.

4. Ідентифікація Нових Можливостей:

– Зміст: Вивчення тональності текстових джерел дозволяє виявити нові можливості або ніші на ринку.

– Приклад Застосування: Визначення, як популярність певних характеристик продуктів впливає на споживачів.

5. Моніторинг Бренду:

– Зміст: Аналіз тональності згадок бренду в текстах дозволяє визначити рівень його популярності та стійкість репутації.

– Приклад Застосування: Слідкування за змінами у сприйнятті бренду на ринку.

6. Оцінка Маркетингових Кампаній:

– Зміст: Аналіз тональності текстів про маркетингові заходи дозволяє визначити ефективність та реакцію на рекламні кампанії.

– Приклад Застосування: Визначення того, як споживачі сприймають рекламу та промоакції.

Використання аналізу тональності тексту в задачах дослідження ринку допомагає компаніям здобувати глибоке розуміння споживацьких уподобань, визначати тренди та адаптувати свої стратегії відповідно до змін в ринковому середовищі.

Задачі розвитку бренду вимагають комплексного підходу для збереження та підвищення репутації компанії чи продукту. Аналіз тональності тексту стає важливим інструментом в цьому контексті, дозволяючи відслідковувати сприйняття бренду серед споживачів та реагувати на зміни вчасно. Розглянемо детально можливості використання аналізу тональності тексту в задачах розвитку бренду.

Можливості використання аналізу тональності тексту в розвитку бренду:

1. Слідкування за Репутацією:

– Зміст: Аналіз тональності відгуків, коментарів та згадок про бренд дозволяє визначити його репутацію в очах споживачів.

– Приклад Застосування: Моніторинг змін у репутації бренду на основі публікацій у соціальних мережах та відгуків на спеціалізованих платформах.

2. Аналіз Відгуків про Продукти:

– Зміст: Вивчення тональності відгуків про конкретні продукти допомагає зрозуміти, як споживачі сприймають індивідуальні елементи бренду.

– Приклад Застосування: Визначення, які характеристики продуктів найбільше впливають на задоволення споживачів.

3. Взаємодія з Клієнтами:

– Зміст: Аналіз тональності повідомлень та запитань споживачів дозволяє виявити проблеми та можливості для покращення обслуговування.

– Приклад Застосування: Визначення та вирішення негативних аспектів взаємодії з брендом, що може підвищити задоволеність клієнтів.

4. Кампанії зі Співпраці та Спонсорства:

– Зміст: Аналіз реакції споживачів на кампанії співпраці та спонсорства дозволяє оцінити ефективність заходів та визначити потреби аудиторії.

– Приклад Застосування: Оптимізація стратегій співпраці на основі реальних реакцій та вподобань споживачів.

5. Аналіз Трендів та Інновацій:

– Зміст: Спостереження за тональністю текстових джерел допомагає виявити нові тренди та інновації, які можуть вплинути на споживчі уподобання.

– Приклад Застосування: Прийняття змін у продуктах чи маркетингових стратегіях для адаптації до ринкових тенденцій.

6. Створення Іміджу та Брендкової Посадки:

– Зміст: Визначення тональності текстів призначених для позиціювання бренду дозволяє контролювати створення та утримання бажаного іміджу.

– Приклад Застосування: Підтримка позитивного сприйняття бренду шляхом аналізу реакцій на маркетингові заходи.

Заключення:

Використання аналізу тональності тексту в задачах розвитку бренду дозволяє компаніям активно взаємодіяти з аудиторією, керувати своєю репутацією та адаптувати стратегії відповідно до потреб споживачів.

Сучасні сервіси оцінки тональності текстів стали важливим інструментом для багатьох галузей, включаючи маркетинг, соціальні мережі, наукові дослідження та ряд інших. Однак існують проблеми, які ставлять під сумнів ефективність та точність таких сервісів. Розглянемо основні проблеми, які виникають у сучасних сервісах оцінки тональності текстів.

Основні проблемні задачі:

1. Полісемія та Амбігвітетність:

– Проблема: Слова, що мають кілька значень (полісемія) або неоднозначність в використанні (амбігвітетність), ускладнюють точне визначення тональності.

– Рішення: Розвиток алгоритмів, які враховують контекст та семантику для кращого розуміння значення слів.

2. Словники та Ньюанси Мови:

– Проблема: Багатозначність слів та нюанси в використанні мови можуть призводити до неточностей у визначенні емоційного забарвлення.

– Рішення: Застосування не лише словникових методів, а й машинного навчання для адаптації до варіацій мови та контексту.

3. Саркастичний та Іронічний Зміст:

– Проблема: Виявлення сарказму та іронії може бути викликаною завданням, оскільки текстові дані часто не передають змісту голосу та міміки.

– Рішення: Використання продвинутих алгоритмів, які враховують контекст та залежать від штучних нейронних мереж.

4. Взаємодія із Сленгом та Мовними Варіантами:

– Проблема: Велика кількість сленгових виразів та мовних варіантів може бути невідомою для стандартних моделей.

– Рішення: Використання актуальних баз даних, які включають в себе сленг та регіональні особливості мови.

5. Зміна Емоційного Стану в Часі:

– Проблема: Зміни в емоційному стані людини з часом можуть призводити до зміни тональності тексту.

– Рішення: Впровадження динамічних алгоритмів, які враховують зміни емоційного фону текстів у реальному часі.

6. Проблеми з Екстремальними Сприйняттями:

– Проблема: Деякі моделі можуть мати тенденцію до перебільшення негативних чи позитивних відгуків.

– Рішення: Перегляд та вдосконалення методів навчання для уникнення викривлення результатів.

Вирішення цих проблем допоможе покращити точність та надійність сучасних сервісів оцінки тональності текстів, забезпечуючи користувачам більш об'єктивні та корисні результати у різноманітних сценаріях використання.

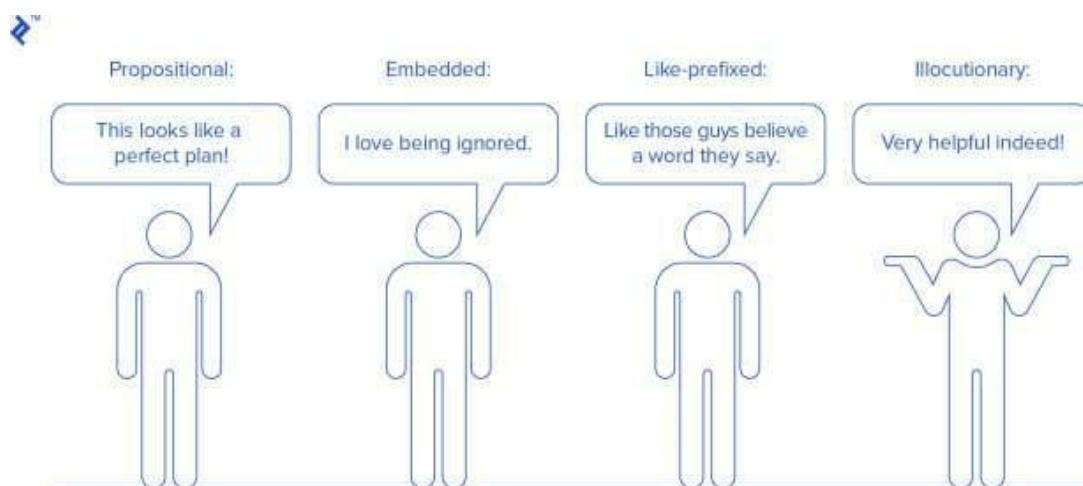


Рис. 2.3. Приклади сарказму

2.2. Аналіз обраних для дослідження методів

2.2.1 Згорткові нейронні мережі

Аналіз методів машинного навчання взявши до уваги технології обробки текстової інформації, зокрема аспекти визначення тональності тексту, розкриває важливість і ефективність застосування згорткових нейронних мереж (CNN) у цьому контексті.

Основні принципи:

1. Згорткові Шари (*Convolutional Layers*):

– Опис: Згорткові нейронні мережі використовують шари згортки для виявлення локальних особливостей у вхідних даних. Вони складаються з фільтрів, які рухаються по вхідному зображенні для виокремлення різних характеристик.

– Застосування: У визначенні тональності тексту, згорткові шари можуть виявляти ключові фрази чи слова, які вказують на емоційне забарвлення.

2. Шари Підтискання (*Pooling Layers*):

– Опис: Після згорткових шарів, шари підтискання використовуються для зменшення розмірності отриманих карт ознак, підсилюючи важливі особливості та зменшуючи кількість параметрів у моделі.

– Застосування: Використання шарів підтискання допомагає виділити найбільш значущі елементи тексту, сприяючи кращому розумінню емоційного відтінку.

3. Повністю З'єднані Шари (*Fully Connected Layers*):

– Опис: Повністю з'єднані шари використовуються для об'єднання виявлених характеристик та прийняття рішення щодо кінцевого виводу. Вони забезпечують просторову інформацію для аналізу взаємозв'язків між характеристиками.

– Застосування: У визначенні тональності, повністю з'єднані шари допомагають моделі робити узагальнені висновки щодо емоційного забарвлення тексту.

Переваги в контексті визначення тональності тексту:

1. Локалізація Особливостей:

– Згорткові шари дозволяють моделі ефективно виявляти емоційно забарвлені фрази чи висловлення в тексті.

2. Зменшення Обсягу Даних:

– Шари підтискання допомагають у скороченні обсягу даних, забезпечуючи при цьому збереження важливих емоційно забарвлених елементів.

3. Врахування Залежностей:

– Повністю з'єднані шари сприяють аналізу взаємозв'язків між словами та фразами, допомагаючи у визначенні загальної тональності тексту.

Використання згорткових нейронних мереж для визначення тональності тексту є перспективним напрямком, оскільки вони дозволяють ефективно розпізнавати емоційно забарвлені елементи та адаптуватися до широкого спектру мовленнєвих особливостей. Порівняно з традиційними методами, *CNN* показують високу точність та здатність узагальнення. (рис. 2.4).

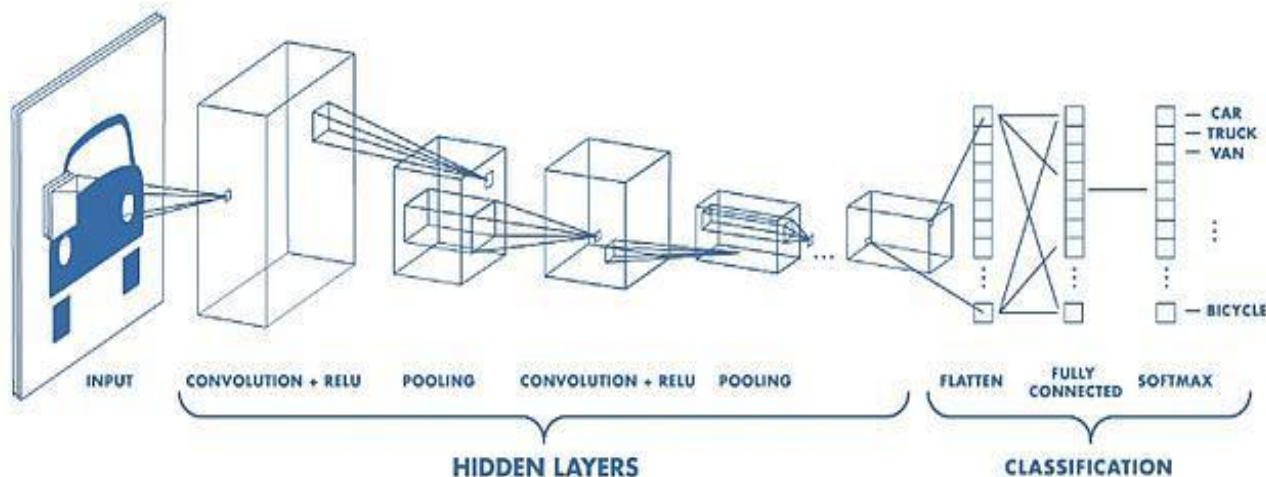


Рис. 2.4. Типова архітектура згорткової нейронної мережі

2.2.2 Рекурентна нейронна мережа – довга та короткочасна пам'ять

Рекурентні нейронні мережі (*RNNs*) є потужним інструментом у галузі обробки послідовностей, що включає в себе розпізнавання мови, машинний переклад, та визначення тональності тексту. Однією з основних варіацій *RNN* є мережа довгострокової та короткочасної пам'яті (*LSTM*), яка була вперше запропонована для вирішення проблеми зниклих та вибухаючих градієнтів, що є традиційними проблемами *RNN*.

Рекурентна нейронна мережа (*RNN*):

1. Опис:

– *RNN* є класом нейронних мереж, які працюють з послідовностями даних. Вони мають внутрішні стани, які дозволяють зберігати і використовувати інформацію з попередніх кроків для обробки поточного вхідного елемента.

2. Проблема зниклих та вибухаючих градієнтів:

– *RNN* мають тенденцію стикатися з проблемою зниклих та вибухаючих градієнтів, коли градієнти стають або надто малими, або надто великими під час зворотного поширення помилки. Це обмежує їх здатність враховувати далеке контекстуальне значення.

Мережа довгострокової та короткочасної пам'яті (*LSTM*):

1. Опис:

– *LSTM* є вдосконаленою версією *RNN*, яка вирішує проблему зниклих та вибухаючих градієнтів. Вона має додатковий механізм контролю потоку інформації через час.

2. Компоненти *LSTM*:

– Клітинний стан (*Cell State*): Зберігає інформацію з попередніх часових кроків та регулює, яка частина інформації повинна передаватись або забиратись.

– Ворота (*Gates*): Включають в себе ворота забуття, ворота входу та ворота виходу, кожне з яких регулює потік інформації.

3. Переваги в Контексті Аналізу Тональності Тексту:

– Довгострокова Залежність: *LSTM* здатні зберігати довгострокові залежності між словами та фразами, що особливо важливо в аналізі тональності.

– Зменшення Впливу Зниклих Градієнтів: Механізми воріт допомагають утримувати градієнти на прийнятному рівні, полегшуючи тренування мережі.

Використання мереж довгострокової та короткочасної пам'яті (*LSTM*) виявляється досить ефективним для завдань аналізу тональності тексту, дозволяючи моделі ефективно враховувати контекст та виконувати більш складні завдання в порівнянні з традиційними рекурентними нейронними мережами.

2.3. Функціональний аналіз програмного модуля сентиментального аналізу тональності відгуків

Сентимент-аналіз, або аналіз тональності, - це галузь обробки природної мови, яка визначає емоційний характер тексту. Програмний модуль сентимент-аналізу призначений для визначення та класифікації емоційного відтінку текстових

документів, таких як відгуки, коментарі, тощо. Нижче наведений детальний функціональний аналіз такого модуля.

Основні функції модуля:

1. Попередня обробка тексту:

– Очищення від шуму: Виявлення та вилучення непотрібних символів, цифр, аббревіатур, щоб поліпшити точність аналізу.

– Токенізація: Розбиття тексту на окремі слова (токени) для подальшого аналізу.

2. Визначення тональності:

– Бінарна класифікація: Визначення, чи є відгук позитивним чи негативним.

– Множинна класифікація: Класифікація відгуку на кілька категорій емоційного відтінку (позитивний, негативний, нейтральний).

3. Використання алгоритмів машинного навчання:

– Тренування моделей: Використання наборів даних для тренування класифікатора на основі алгоритмів машинного навчання, таких як наївний Баєсівський класифікатор, метод опорних векторів, глибокі нейронні мережі тощо.

4. Аналіз контексту:

– Робота зі словниками емоцій: Визначення важливих слів та фраз, які можуть вказувати на конкретний емоційний стан.

5. Робота з нейтральними відгуками:

– Виявлення нейтральних відгуків: Класифікація відгуків, що не несуть вираженої позитивної або негативної емоційної оцінки.

6. Оцінка точності:

– Валідація та тестування: Використання валідаційних та тестових наборів даних для перевірки точності та ефективності моделі.

7. Інтерпретація результатів:

– Забезпечення зрозумілості: Пояснення результатів аналізу для кінцевого користувача, можливо, за допомогою виведення ключових слів чи речень, що визначають емоційний відтінок.

Функціональний аналіз програмного модуля sentiment-аналізу визначає ключові етапи обробки та аналізу текстових відгуків, щоб визначити їхню емоційну тональність. Це важливий інструмент для бізнесу та наукових досліджень для розуміння сприйняття користувачами продуктів чи послуг.

Дерево функцій програмного модуля аналізу тональності українськомовних відгуків (рис. 2.5):

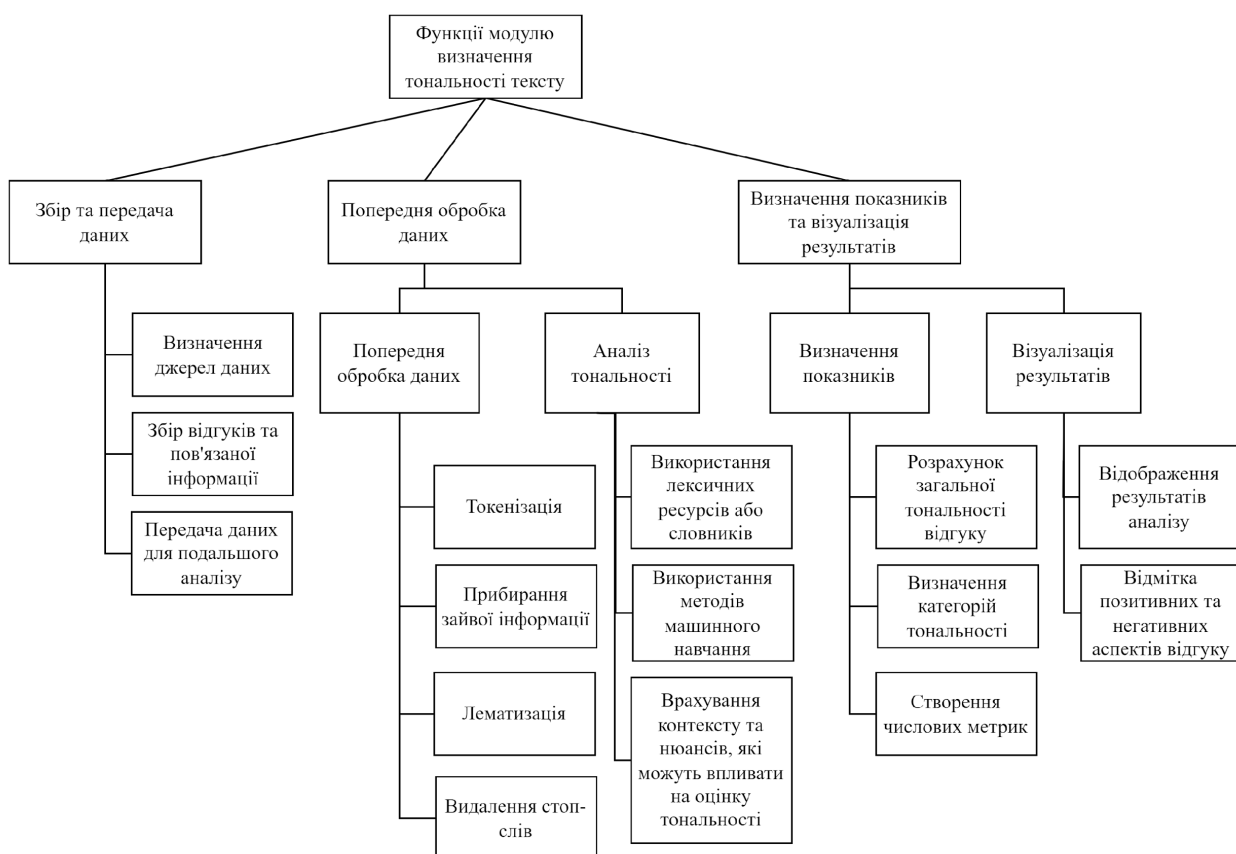


Рис. 2.5. Дерево функцій програмного модуля сентиментального аналізу тональності відгуків

Діаграма *IDEF0 (Integration Definition for Function Modeling)* "ЯК БУДЕ" – це графічний інструмент, що використовується для моделювання функцій та їх взаємозв'язків у процесах або системах. Для створення діаграми "ЯК БУДЕ" використовуються блоки, стрілки та тексти для відображення функцій, вхідних та вихідних даних, управління та залежностей між ними (рис. 2.6).

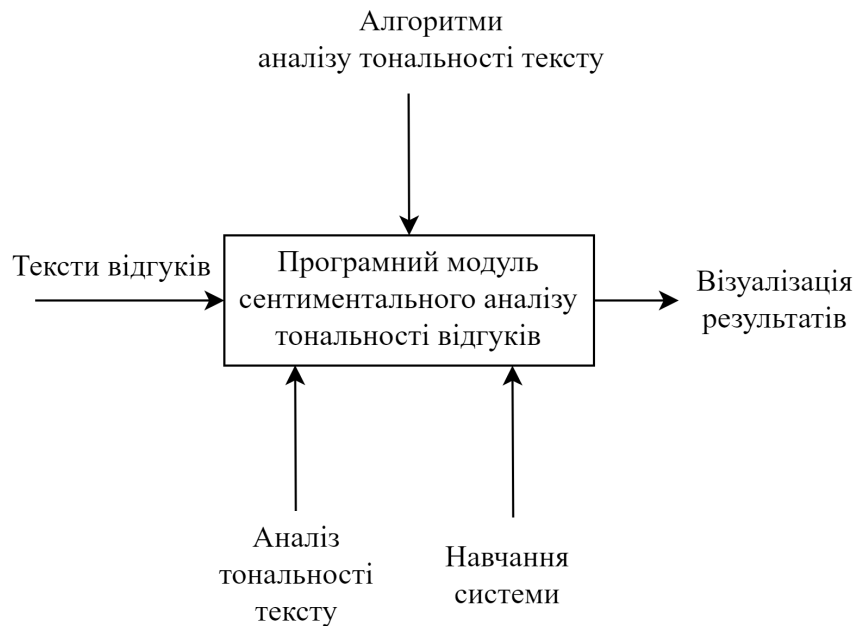


Рис. 2.6. Діаграма «ЯК БУДЕ» для програмного модуля сентиментального аналізу тональності відгуків

На зображенні ви можете побачити діаграму з назвою "Рис. 2.6. Діаграма «ЯК БУДЕ» для програмного модуля сентиментального аналізу тональності відгуків". Ця діаграма представляє процес роботи програмного модуля, який призначений для аналізу емоційного забарвлення (сентименту) текстів, зокрема відгуків користувачів. Процес розпочинається з "Авторитетного зразка тональності тексту", який служить вхідними даними для "Програмного модуля сентиментального аналізу тональності відгуків".

Основні компоненти діаграми та їх взаємодія:

1. Авторитетний зразок тональності тексту - це, можливо, набір даних або корпус текстів, який використовується для навчання або калібрування програмного модуля, щоб він міг коректно розпізнавати емоційне забарвлення в текстах.

2. Програмний модуль сентиментального аналізу - це головний компонент системи, який обробляє текстові дані і визначає їх сентимент, тобто чи є загальний тон відгуку позитивним, негативним, чи нейтральним.

3. Текстові відгуки - вхідні дані для програмного модуля, які можуть надходити від користувачів або бути зібрані з різних джерел.

4. Аналіз тональності тексту і Настроювання системи - ці процеси можуть представляти внутрішні механізми програмного модуля, де аналіз виконується на основі певних алгоритмів, а настроювання системи дозволяє керувати параметрами для покращення точності аналізу.

5. Виведення результатів - це кінцевий етап, на якому програмний модуль видає результати аналізу, які можуть використовуватися для прийняття рішень, оцінювання продуктів або послуг, аналізу настрою споживачів тощо.

Ця діаграма, ймовірно, використовується для ілюстрації плану розробки або удосконалення програмного забезпечення, що здійснює сентиментальний аналіз.

Опис діаграми *IDEF0* "ЯК БУДЕ" може бути представлений як детальний огляд складових частин системи аналізу тональності тексту. Кожна функція та її декомпозиція відображає конкретні етапи обробки вхідних даних і вихідних результатів:

1. Функція "Вхідні дані":

- Прийняття вхідних даних: Збір відгуків чи текстів для аналізу.
- Перевірка формату та достовірності: Валідація вхідних даних на предмет відповідності вимогам аналізу.

2. Функція "Аналіз тональності тексту":

- Токенізація тексту: Розбиття тексту на слова або фрази для аналізу.
- Лематизація/стемінг: Приведення слів до основної форми для уніфікації.
- Видалення стоп-слів: Виключення з тексту неважливих слів.
- Аналіз за допомогою словників/машинного навчання: Визначення емоційного забарвлення.

- Контекстуалізація: Врахування контексту тексту для точнішого аналізу.

3. Функція "Визначення показників":

- Загальна тональність: Визначення основного емоційного забарвлення тексту.
- Категорії тональності: Класифікація тексту за позитивними, негативними або нейтральними категоріями.

- Числові метрики: Кількісне вираження результатів аналізу.

4. Функція "Візуалізація результатів":

– Підготовка даних: Організація даних для візуалізації.

– Графічне відображення: Створення діаграм, графіків та інших візуальних елементів.

– Передача результатів: Подання візуалізованих даних для кінцевого користувача.

5. Функція "Вихідні дані":

– Форматування результатів: Встановлення структури вихідних даних.

– Виведення результатів: Представлення аналізу кінцевим користувачам або системам.

Стрілки на діаграмі *IDEFO* вказують на потік інформації та зв'язки між функціями. Відповідно, вхідні дані перетворюються в процесі аналізу та врешті-решт візуалізуються і подаються як вихідні результати. Ця діаграма є кроком до створення інтуїтивно зрозумілої та ефективної системи для кінцевих користувачів.

2.4. Висновки до розділу

У даному розділі представлено структуру системи аналітики емоційного забарвлення текстів українською мовою. Вона включає набір модулів для розробки: збір та первинна обробка даних, трансформація тексту в векторну форму, алгоритми класифікації тональності та модуль для графічного представлення результатів.

Було здійснено вивчення функціональності створеної системи для аналізу емоційного забарвлення текстів, в ході якого оцінено результати тестувань, точність та продуктивність системи.

Детально описано етапи тренування моделі, з акцентом на відбір та адаптацію даних для тренування, розробку та тонке налаштування моделі.

З урахуванням аналізу результатів тестувань можна стверджувати, що створена система для аналізу емоційного відтінку текстів українською мовою є ефективним засобом для визначення емоційних реакцій користувачів. Важливо зазначити, що система володіє значним потенціалом для розширення та удосконалення в рамках

подальших досліджень у галузі обробки природної мови та розвитку інформаційних технологій.

РОЗДІЛ 3

РОЗРОБКА СИСТЕМИ ВИЗНАЧЕННЯ ТОНАЛЬНОСТІ ТЕКСТІВ УКРАЇНСЬКОЮ МОВОЮ

3.1. Архітектура програмного модуля та розгортання необхідних компонентів розробки

Архітектура системи визначення тональності текстів українською мовою включає ряд додаткових модулів, які є необхідними для визначення тональності на рис. 3.1.



Рис. 3.1. Зв'язки між компонентами програмного модуля обробки текстової інформації для визначення тональності

Архітектура програмного модуля для Системи Визначення Тональності Текстів українською мовою:

1. Мета Системи:

– Визначення емоційного відтінку тексту, виявлення позитивних, негативних та нейтральних відгуків або стверджень українською мовою.

2. Компоненти Системи:

– Введення Тексту:

– Забезпечення можливості введення тексту через інтерфейс користувача або завантаження текстових файлів.

– Процес Аналізу Тексту:

– Токенізація:

– Розділення тексту на окремі токени (слова чи фрази) для подальшого аналізу.

– Модель Оцінки Тональності:

– Використання алгоритмів машинного навчання чи правилкових методів для визначення емоційного забарвлення кожного токена.

– Навчання моделі на великому корпусі текстів для досягнення високої точності.

– Локалізація Тексту:

– Забезпечення можливості обробки тексту українською мовою.

– Переклад тексту або використання спеціалізованих моделей для аналізу україномовного контенту.

– Виведення Результатів:

– Подання результатів аналізу у зрозумілій формі, можливість визначення загальної емоційної тону тексту та інших статистичних даних.

3. Технології та Мови Програмування:

– Мова Програмування:

– Використання *Python*, який є популярним для обробки текстів та машинного навчання.

– Бібліотеки та Фреймворки:

– Використання бібліотек для обробки тексту, таких як *NLTK* або *SpaCy*.

– Використання фреймворків для машинного навчання, наприклад, *TensorFlow* або *PyTorch*.

4. Розгортання Компонентів:

- Інтерфейс Користувача:

- Створення графічного або консольного інтерфейсу для зручної взаємодії з користувачем.

- Сервер або Хмарні Рішення (опціонально):

- Можливість розгортання системи на сервері або хмарних платформах для доступу із різних пристроїв.

5. Тестування та Валідація:

- Тестування Моделі:

- Застосування тестових наборів даних для перевірки точності та стабільності моделі.

- Тестування Інтерфейсу:

- Перевірка зручності та ефективності взаємодії з користувачем.

6. Оновлення та Підтримка:

- Оновлення Моделі:

- Регулярні оновлення навченої моделі для врахування нових трендів та забезпечення актуальності результатів.

- Підтримка Мов:

- Розширення підтримки для інших мов для більш широкого використання.

7. Безпека:

- Захист Від Атак:

- Розробка заходів безпеки для захисту системи від можливих атак або використання зловмисного контенту.

- Можливість фільтрації шкідливих текстів та ідентифікації потенційно небезпечних відгуків.

Створюємо нове середовище, для чого потрібно в *Anaconda Navigator* натиснути на клавішу імпорт та вибрати файл з розширенням *.yaml*, щоб встановити потрібні нам бібліотеки (рис. 3.2).

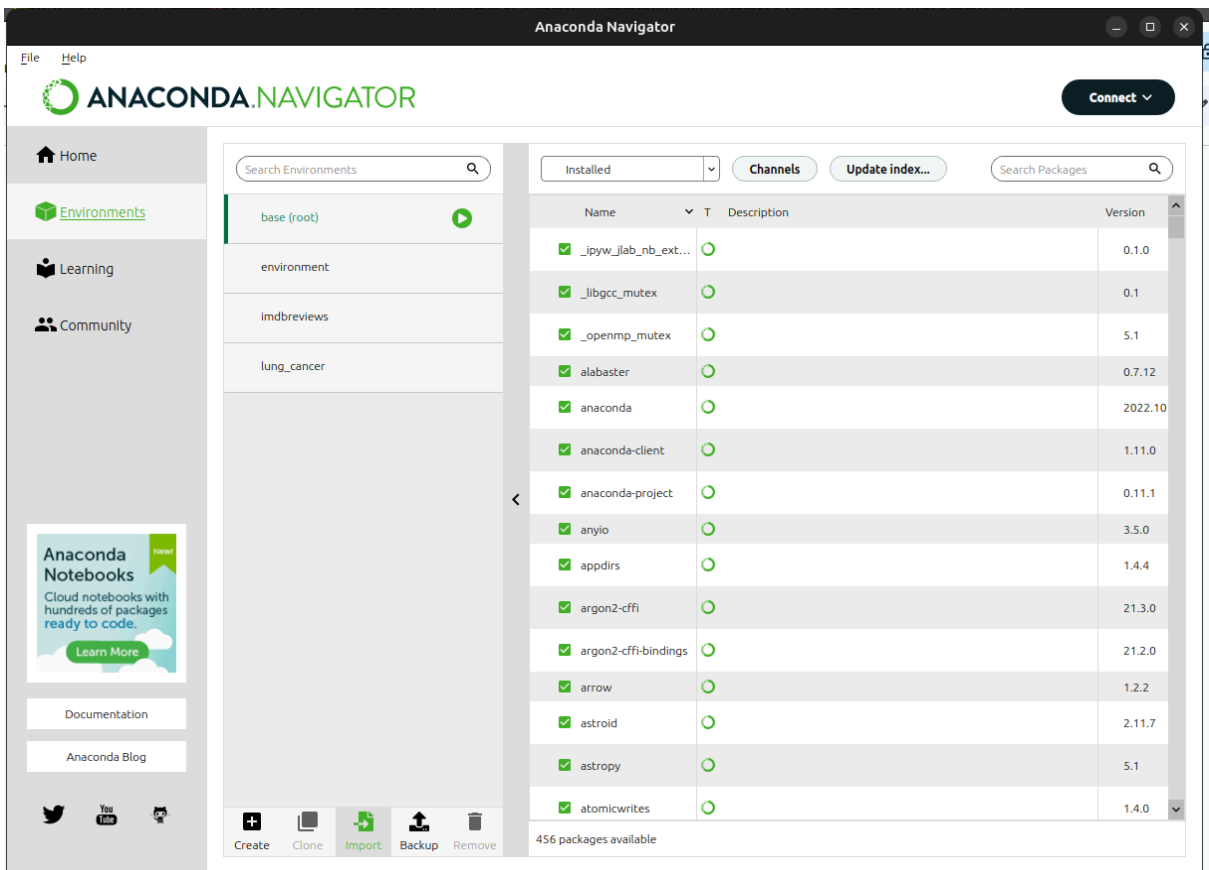


Рис. 3.2. Налаштування проєкту в *Anaconda Navigator*

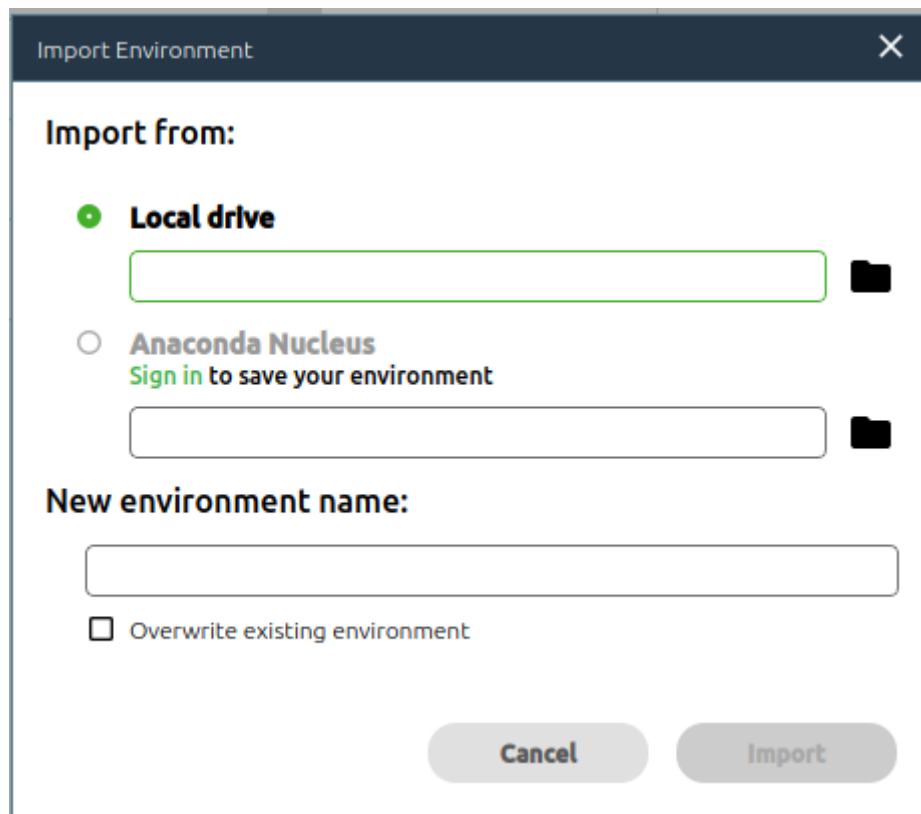


Рис. 3.3. Вибір розміщення проєкту в *Anaconda Navigator*

Детальний опис Розгортання Середовища Розробки з Використанням *Anaconda* та *Anaconda Navigator*:

1. *Anaconda*:

– Опис:

– *Anaconda* - це вільний та відкритий дистриб'ютив *Python* та *R*, який призначений для обробки та аналізу даних, наукового моделювання, великих обчислень та машинного навчання.

– Кроки встановлення:

– Завантажте встановний файл *Anaconda* з офіційного веб-сайту.

– Виконайте встановлення, слідуючи інструкціям майстра встановлення.

– Виберіть опцію додавання *Anaconda* до шляху системи для забезпечення зручного доступу.

2. *Anaconda Navigator*:

– Опис:

– *Anaconda Navigator* - це графічний інтерфейс для управління та використання різних інструментів та середовищ, що постачаються з *Anaconda*.

– Кроки встановлення:

– Після встановлення *Anaconda* використовуйте *Anaconda Navigator* для управління вашими середовищами та пакетами.

– Запустіть *Anaconda Navigator*, використовуючи ярлик або команду в командному рядку.

– Для роботи з *Python* середовищем виберіть "Home" та перейдіть до відповідної вкладки, де ви зможете створювати нові середовища, встановлювати пакети, запускати *Jupyter Notebook* тощо.

3. Створення та Управління Середовищами:

– Опис:

– *Anaconda* дозволяє створювати ізольовані середовища для різних проектів, кожне з яких може мати свої власні версії *Python* та пакетів.

– Кроки створення:

– Використовуйте *Anaconda Navigator* або командний рядок для створення нового середовища.

– Встановіть необхідні пакети в кожному середовищі, щоб забезпечити ізоляцію та управління залежностями проекту.

4. Використання *Jupyter Notebook*:

– Опис:

– *Jupyter Notebook* - це інтерактивна оболонка, яка дозволяє об'єднати текст, код, графіку та результати виконання в одному документі.

– Кроки використання:

– Запустіть *Jupyter Notebook* через *Anaconda Navigator* або з командного рядка.

– Створіть новий ноутбук та почніть працювати, використовуючи *Python* код, коментарі та інші елементи.

Завдяки *Anaconda* та *Anaconda Navigator*, ви можете легко управляти вашими проектами, встановлювати необхідні пакети та ефективно використовувати ізольовані середовища для різних завдань.

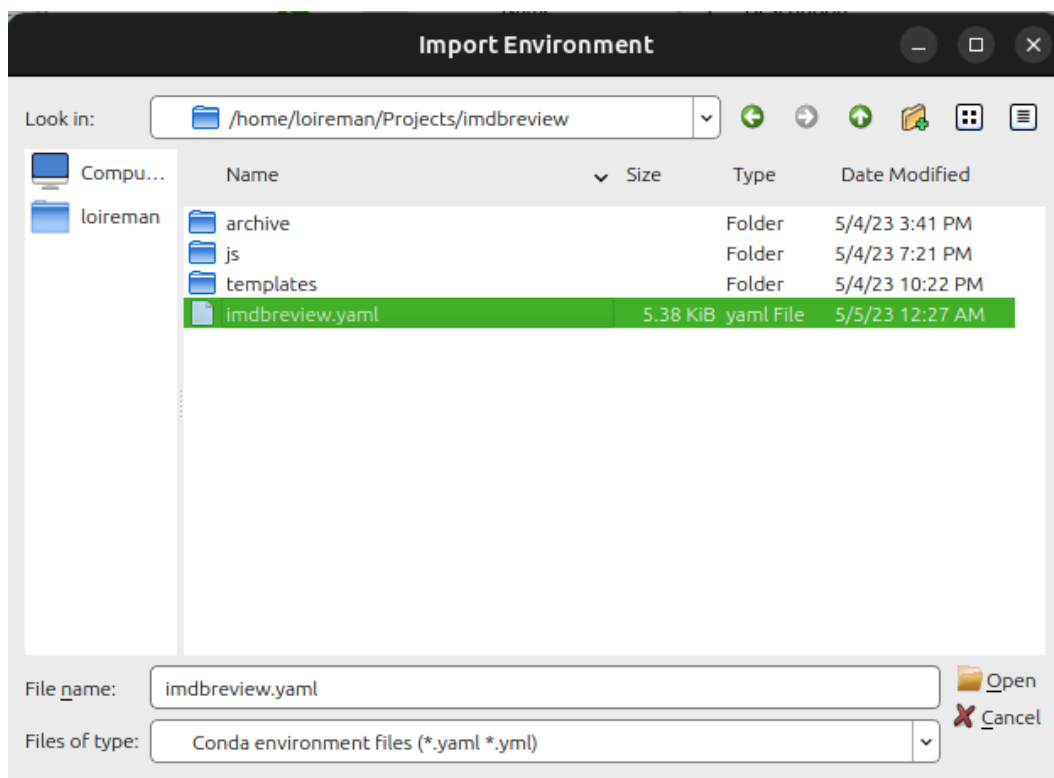


Рис. 3.4. Основний модуль проекту в *Anaconda Navigator*

5. Встановлення Бібліотек:

– Опис:

– Після налаштування *Anaconda* можна встановлювати необхідні бібліотеки для розробки та тренування моделі. Зазвичай для мовного моделювання та машинного навчання використовують бібліотеки, такі як *TensorFlow*, *PyTorch*, *scikit-learn* та інші.

– Кроки встановлення:

– Відкрийте *Anaconda Navigator* та використовуйте інтерфейс для встановлення бібліотек через "*Environments*".

– Або використовуйте командний рядок та пакетний менеджер (*pip* або *conda*) для встановлення бібліотек.

6. Пояснення Коду Тренування Моделі:

– Опис:

– Код для тренування моделі визначення тональності тексту включає в себе кілька етапів, таких як підготовка даних, визначення архітектури моделі, тренування та оцінка.

– Етапи Коду:

– Підготовка Даних:

– Завантаження та підготовка набору даних для тренування. Розділення на навчальний та тестовий набори.

– Визначення Архітектури Моделі:

– Створення архітектури моделі, наприклад, використовуючи бібліотеки типу *TensorFlow* або *PyTorch*. Це може бути рекурентна нейронна мережа (*RNN*), конволюційна нейронна мережа (*CNN*) або інша структура.

– Тренування Моделі:

– Використання навчального набору для тренування моделі. Визначення гіперпараметрів, вибір оптимізатора та функції втрат.

– Оцінка Моделі:

– Використання тестового набору для оцінки продуктивності моделі. Визначення метрик, таких як точність, втрата тощо.

– Приклад Коду:

– *pythonCopy code*

Приклад використання *TensorFlow* для створення та тренування моделі

```
import tensorflow as tf from tensorflow.keras.models
```

```
import Sequential from tensorflow.keras.layers
```

```
import Embedding, LSTM, Dense
```

```
# Підготовка даних
```

```
# ...
```

```
# Визначення архітектури моделі
```

```
model = Sequential()
```

```
model.add(Embedding(input_dim=vocab_size, output_dim=embedding_dim,  
input_length=max_length))
```

```
model.add(LSTM(100)) model.add(Dense(1, activation='sigmoid'))
```

```
# Компіляція та тренування моделі
```

```
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
```

```
model.fit(train_data, train_labels, epochs=10, validation_data=(test_data,  
test_labels))
```

Зазначений код є загальним прикладом та може варіюватися в залежності від обраної бібліотеки та архітектури моделі.

Функція *accurac_model* визначена для обчислення середньої точності (*accuracy*) за допомогою перехресної перевірки для кожної з моделей, перелічених у словнику *models*. Вона приймає чотири аргументи:

1. *x*: Це словник, де ключами є імена моделей, а значеннями - вектори ознак для кожного набору даних.

2. *y*: Це масив міток (класів) для кожного набору даних, який відповідає векторам ознак у словнику *x*.

3. *models*: Це словник, де ключами є імена моделей (наприклад, "*MultinomialNB*", "*SVC*", тощо), а значеннями - об'єкти моделей класифікації, які ви хочете оцінити.

4. *not_sparse*: Цей параметр вказує, чи потрібно перетворити розріджені матриці у масиви перед передачею їх у функцію *cross_val_score*.

Основною операцією, яку виконує ця функція, є виклик функції *cross_val_score* з бібліотеки *Scikit-learn* для кожної моделі та набору даних. Функція *cross_val_score* використовує перехресну перевірку для оцінки точності моделі.

Результатом роботи функції *accurac_model* є словник, де ключами є імена моделей, а значеннями - середні значення точності для кожної моделі на відповідних наборах даних.

Наприклад, якщо *models* містить дві моделі, "*MultinomialNB*" та "*SVC*", і *x* містить вектори ознак для двох наборів даних, функція *accurac_model* поверне словник, де ключами будуть імена моделей, а значеннями - середні значення точності для кожної моделі на відповідних наборах даних.

Функція *report* може бути викликана для виведення результатів оцінки точності моделей, отриманих за допомогою функції *accurac_model*, у зручному для аналізу форматі. Вона може виводити середню точність для кожної моделі, а також стандартне відхилення (якщо потрібно).

Ці функції можуть бути корисними для порівняння різних моделей класифікації на вашому корпусі текстів і визначення того, яка модель працює найкраще для вашої задачі аналізу тональності.

```
(imdbreview) loireman@cobain:~$ cd /home/loireman/Projects/imdbreview/  
(imdbreview) loireman@cobain:~/Projects/imdbreview$ python app.py  
2023-05-05 07:12:41.459619: I tensorflow/core/platform/cpu_feature_guard.cc:193]  
This TensorFlow binary is optimized with the NVIDIA RTX 3090. To enable more GPU acc...
```

Рис. 3.5. Команда запуску веб-інтерфейсу

Після цього у нас завантажиться сайт за адресою *http://localhost:5000*

```
Skipping registering GPU devices...  
Loaded model from disk  
* Serving Flask app 'app'  
* Debug mode: off  
WARNING: This is a development server. Do not use it in a production deployment.  
Use a production WSGI server instead.  
* Running on http://127.0.0.1:5000  
Press CTRL+C to quit  
█
```

Рис. 3.6. Формування сайту за визначеною адресою

3.2. Аналіз роботи розробленої системи аналізу тональності тексту

Застосунок при запуску має декілька параметрів: поле вводу тексту, та кнопка “Завантажити” (рис. 3.7).

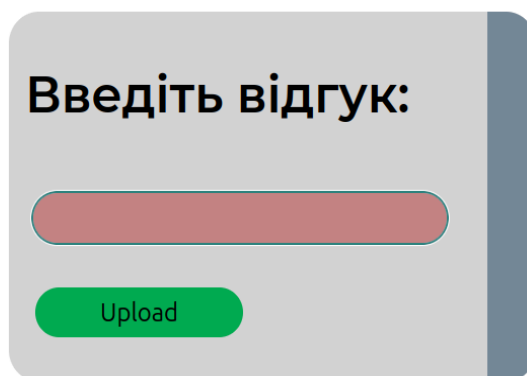


Рис. 3.7. Інтерфейс розробленої системи

На зображенні ви бачите інтерфейс користувача для застосунку, який, судячи з контексту, призначений для аналізу відгуків або тексту. Інтерфейс має мінімалістичний дизайн з наступними елементами:

1. Заголовок: "Введіть відгук:". Цей заголовок інструктує користувача дією, яку потрібно виконати — введення тексту відгуку.

2. Поле вводу тексту: Велике поле для вводу тексту виділено рожевим кольором. Це основний елемент інтерфейсу, куди користувачі можуть ввести текст відгуку, який вони хочуть аналізувати. Розмір поля вказує на те, що очікується введення довшого тексту.

3. Кнопка “*Upload*”: Зелена кнопка розташована нижче поля вводу тексту. Вона має назву “*Upload*”, що зазвичай вказує на дію завантаження файлу або даних. Однак у контексті аналізу відгуків ця кнопка може бути використана для подальшого відправлення тексту, введеного в поле вище, до системи для обробки.

Інтерфейс розроблено з акцентом на функціональність і простоту використання. Користувачам не потрібно розбиратися зі складними меню або

налаштуваннями — вони просто вводять текст і натискають кнопку для його обробки. Цей інтерфейс може бути частиною більшої системи, яка аналізує емоційний забарвлення текстів, як зазначено у вашому попередньому повідомленні.

Якщо у поле (позначене червоним) не ввести значення та натиснути завантажити – виникне наступне повідомлення (рис. 3.8).

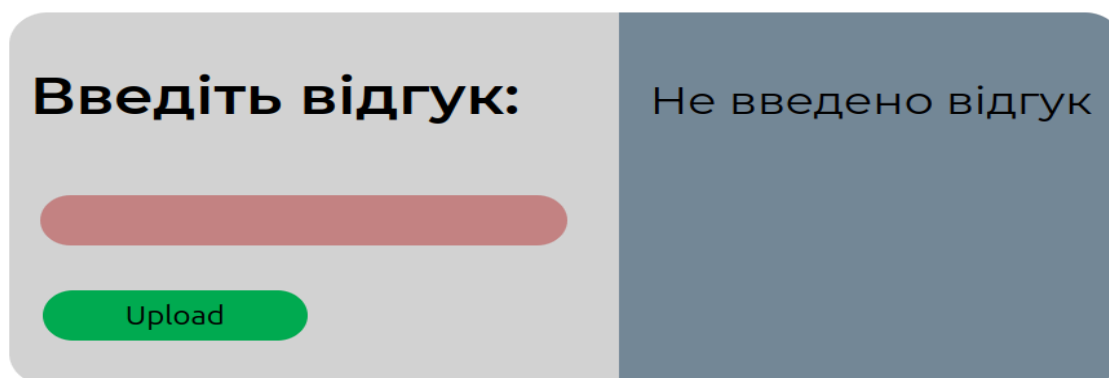


Рис. 3.8. Обробка помилок в системі

В іншому випадку, на основі текст відгуку отримаємо оцінку емоційного забарвлення відгуку (погано/задовільно/добре), яку можна замінити на більшу кількість оцінки забарвлення від 1 до 10 (дана властивість налаштовується у коді).

Також будується переклад тексту, що зумовлено тим, що базові набори даних *IMDB* було взято з відкритих джерел і вони були англійською мовою, а вже потім проводилась локалізація (рис. 3.9).

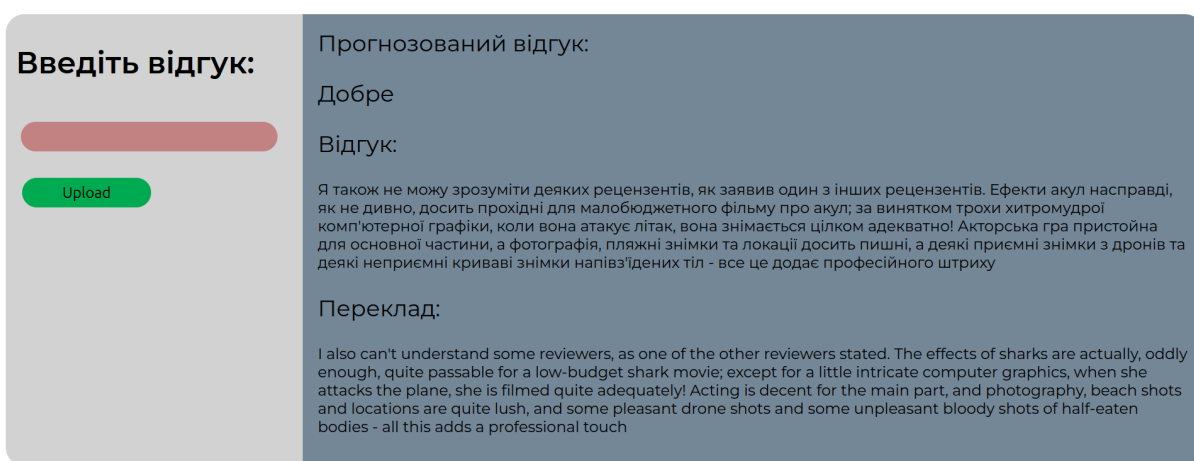


Рис. 3.9. Приклад роботи визначення тональності тексту

На зображенні представлено інтерфейс користувацького застосунку, де можна ввести текст відгуку для аналізу. Якщо користувач залишить поле для тексту (позначене рожевим кольором) порожнім та спробує натиснути кнопку "Upload", система видасть повідомлення про помилку, яке говорить "Не введено відгук". Це повідомлення, розміщене у сірому прямокутнику справа від поля вводу, слугує індикатором, що користувач мусить ввести текст перед завантаженням для аналізу.

У випадку введення тексту відгуку та натискання на кнопку, система проаналізує текст і надасть оцінку емоційного забарвлення відгуку. За замовчуванням, оцінка може виражатися у трьох категоріях: "погано", "задовільно", "добре". Проте, існує можливість налаштувати систему таким чином, щоб оцінка емоційного забарвлення представлялася числовим значенням від 1 до 10, що дає більш диференційоване розуміння тональності відгуку. Це налаштування здійснюється в кодї програми, дозволяючи адаптувати механізм оцінювання відповідно до конкретних потреб користувачів або вимог проекту.

Детальний опис модулів для Системи Визначення Тональності Тексту:

1. Модуль Визначення Мови Тексту:

– Мета:

– Визначення мови, у якій написано введений текст.

– Опис Функціоналу:

– Використання статистичних методів, мовних моделей чи алгоритмів класифікації для визначення мови.

– Розрізнення між різними мовами для подальшої коректної обробки тексту.

2. Модуль Створення Графу Сутностей:

– Мета:

– Побудова графу сутностей для аналізу зв'язків та контексту між різними елементами тексту.

– Опис Функціоналу:

– Виділення ключових слів та термінів, які можуть представляти сутності.

– Побудова графу, де вузли - це сутності, а ребра - взаємозв'язки між ними.

3. Модуль Розпізнавання Іменованих Сутностей:

– Мета:

– Визначення та класифікація іменованих сутностей (осіб, місць, організацій тощо) у тексті.

– Опис Функціоналу:

– Використання методів машинного навчання (*NER - Named Entity Recognition*) для виявлення та класифікації іменованих сутностей.

– Застосування базових мовних моделей та глибокого навчання для підвищення точності розпізнавання.

4. Модуль Визначення Тональності Тексту:

– Мета:

– Встановлення емоційного забарвлення тексту (позитивне, негативне, нейтральне).

– Опис Функціоналу:

– Використання алгоритмів машинного навчання для класифікації тексту за емоційними категоріями.

– Навчання моделі на збалансованому корпусі текстів з позитивними та негативними відгуками.

– Врахування внутрішньотекстового контексту та графу сутностей для підвищення точності визначення тональності.

Кожен модуль взаємодіє з іншими для створення комплексної системи визначення тональності тексту. Наприклад, результат розпізнавання іменованих сутностей може впливати на аналіз тональності, дозволяючи системі враховувати конкретні аспекти тексту при визначенні емоційного забарвлення.

У наступних двох рядках використовуються методи *ylabel* та *xlabel* з бібліотеки *Matplotlib* для встановлення підписів до осей *Y* та *X* відповідно.

Побудований графік представлено на рис. 3.10.

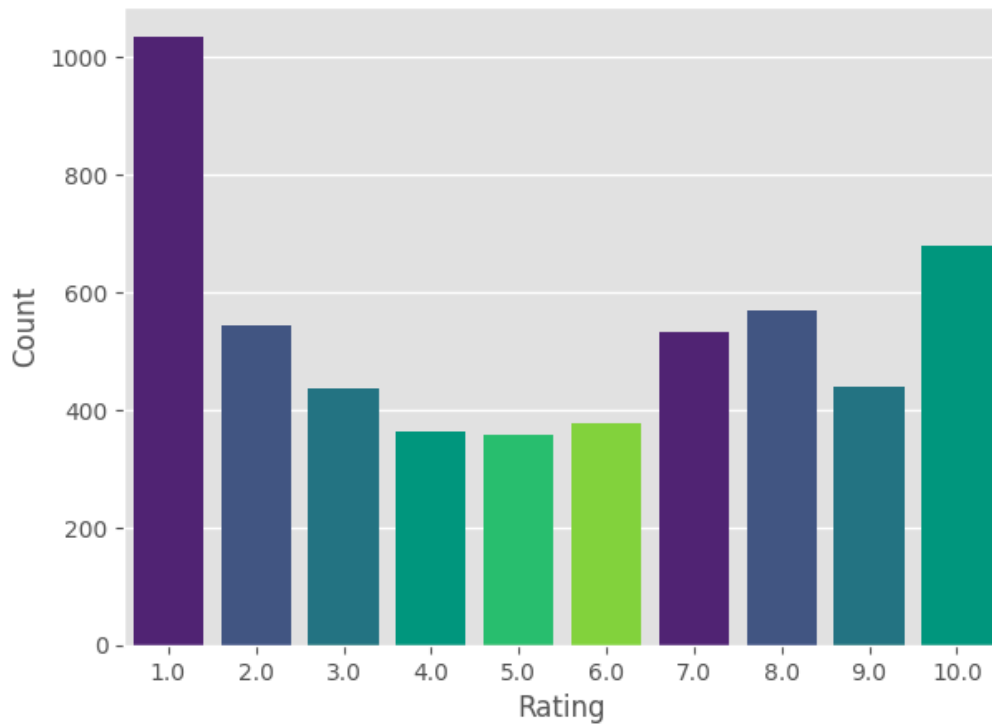


Рис. 3.10. Відношення тексту то шкали тональності

Послідовність обробки даних передбачає застосування функції *ravel*, яка є відповідальною за трансформацію багатовимірного масиву *ax* в одновимірну структуру. Ця операція необхідна для спрощення подальшої обробки даних. Далі, протягом трьох окремих операцій, функція *Significant_words* задіюється для аналізу та графічного представлення 15 ключових слів. Кожна з цих операцій спеціалізується на різних категоріях відгуків: вона виділяє і візуалізує найбільш вагомні слова, що зустрічаються у заголовках відгуків загалом, а також виокремлює слова з позитивними та негативними відгуками. Такий підхід дозволяє зробити порівняльний аналіз лексики, яка використовується користувачами при формулюванні їхніх вражень, та виявити терміни, що найчастіше асоціюються з певними емоційними оцінками (рис. 3.11).

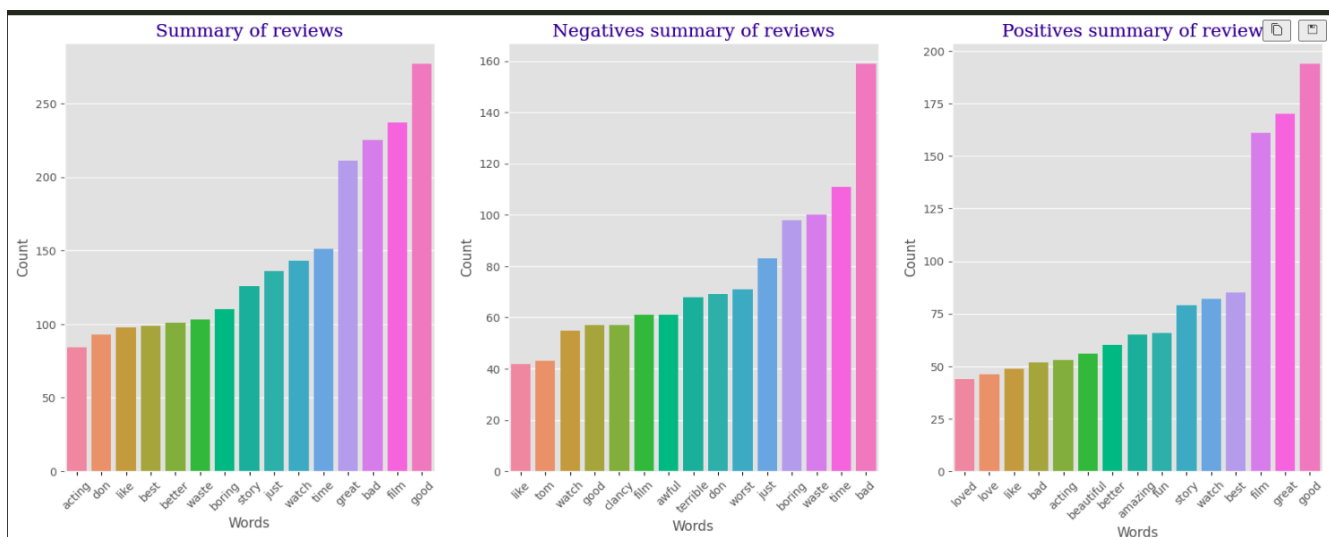


Рис. 3.11. Приклад побудованих графіків для візуалізації визначення тональності тексту

Аргументи `vec_title`, `data.title`, `data[data.rating<5].title` та `data[data.rating>5].title` передаються як аргументи `Vectorizer` та `Data` функції `Significant_words`. Аргументи `title` передаються як аргументи `title` функції `Significant_words`. Аргументи `ax[0]`, `ax[1]` та `ax[2]` передаються як аргументи `ax` функції `Significant_words`.

3.3. Описання процесу навчання моделі

Процес навчання моделі у контексті визначення тональності тексту включає кілька важливих етапів, що визначають якість та ефективність результуючої системи. Нижче подано детальний опис кожного етапу цього процесу:

1. Підготовка Даних:

– Збір та Очищення Даних: На цьому етапі здійснюється збір великого обсягу текстових даних, що містять відгуки та коментарі. Дані можуть бути отримані з різних джерел, таких як веб-сайти, соціальні мережі тощо. Після цього проводиться очищення та підготовка даних, включаючи видалення шуму, обрізку текстів, та перетворення їх у формат, придатний для навчання моделі.

– Розподіл Даних: Дані поділяються на навчальний та тестовий набори для ефективного оцінювання відповіді моделі. Також може використовуватися валідаційний набір для налаштування параметрів моделі.

– Векторизація Тексту: Текст перетворюється в числовий формат через векторизацію. Цей процес може включати токенізацію (поділ тексту на слова або токени), векторне представлення слів (наприклад, за допомогою *Word Embeddings*), або використання передових методів, таких як *BERT*.

2. Вибір Моделі:

– Вибір Архітектури Моделі: Визначається архітектура моделі, яка може включати в себе різні типи шарів, наприклад, рекурентні, сверточні або трансформерні, залежно від завдань та характеристик даних.

– Визначення Гіперпараметрів: Обираються гіперпараметри моделі, такі як розмір пакету, швидкість навчання, кількість шарів тощо. Ці параметри можуть впливати на швидкість навчання та якість моделі.

3. Навчання Моделі:

– Формування Вихідного Шару: Визначається формат вихідного шару моделі, який відповідає кількості класів тональності (позитивний, негативний, нейтральний).

– Визначення Функції Втрат: Обирається функція втрат, яка визначає, наскільки віддалені передбачені значення від фактичних.

– Навчання та Оптимізація: Модель навчається на тренувальних даних, але при цьому слід уникати перенавчання. Застосовуються оптимізатори та методи регуляризації для підтримки загальної здатності моделі.

4. Оцінка та Покращення Моделі:

– Оцінка на Тестових Даних: Модель оцінюється на тестових даних для визначення її точності, вибіркової та інших метрик ефективності.

– Файнтюнінг та Оптимізація: Здійснюється файнтюнінг моделі для покращення її роботи на нових даних. Можлива оптимізація гіперпараметрів.

5. Валідація та Впровадження:

– Валідація Результатів: Результати оцінюються на валідаційному наборі для перевірки узгодженості та коректності.

– Впровадження в Продукцію: У разі успішного навчання та валідації модель готова для впровадження в реальні умови. Важливо забезпечити моніторинг та підтримку моделі в робочому середовищі.

Загальний успіх процесу навчання моделі залежить від правильного підбору даних, архітектури моделі, гіперпараметрів та правильного оцінювання результатів. Тільки в комплексі ці етапи дозволяють створити надійну та ефективну систему визначення тональності текстів.

У даному підрозділі буде викладено детальний аналіз функціональності та елементів інтерфейсу розробленої системи аналізу тональності тексту.

1. Інтерфейс Користувача:

– Система має простий та зрозумілий інтерфейс, що складається з поля вводу тексту та кнопки "Завантажити", як зображено на рис. 3.7. Користувач може ввести текст для аналізу та завантажити його для отримання результатів.

2. Обробка Помилки:

– У разі, якщо користувач не ввів значення у поле (позначене червоним), та спробував завантажити, система виведе відповідне повідомлення про помилку, як показано на рис. 3.8. Цей механізм допомагає виключити можливість виникнення некоректних запитань до системи.

3. Визначення Тональності та Переклад Тексту:

– У випадку введення тексту користувача, система визначає оцінку емоційного забарвлення відгуку (погано/задовільно/добре). Також можлива зміна кількості оцінки забарвлення від 1 до 10, що визначається у коді системи. Дана функція дозволяє налаштовувати точність та гранулярність аналізу.

– З метою роботи з базовими даними *IMDB*, які походять з англійської мови, система використовує переклад тексту. Це необхідно, оскільки було взято зразок даних у вигляді відгуків, які були в англійському варіанті. Процес перекладу сприяє розширенню функціональності системи та полегшує користувачам роботу з різномовним контентом.

4. Візуалізація Результатів:

– Для наглядності та візуалізації результатів аналізу, розроблено фрагмент коду, який використовує бібліотеки *Matplotlib* та *Seaborn*. Графік, побудований з використанням цього коду, візуалізує відношення тексту до різних тонів, що надає користувачам чітке уявлення про розподіл тональності в текстових відгуках.

– Наведений фрагмент коду налаштовує стиль графіка, створює полотно та графік, використовуючи *Seaborn*. Отриманий графік дозволяє досліджувати частоту виникнення різних тональностей у текстових відгуках.

Класифікація тексту - це завдання машинного навчання, в якому текстові дані призначаються до одного або декількох попередньо визначених класів або категорій. У завданні аналізу тональності тексту основна мета - визначити, чи має текст позитивний, негативний або нейтральний настрій.

Важливі поняття:

1. Корпус текстів: Для навчання класифікатора потрібен набір текстових даних, який називається корпусом. Цей корпус має містити тексти, які вже мають позначку щодо їх тональності (позитивний, негативний, нейтральний).

2. Перетворення тексту в числові вектори: Текст необхідно перетворити в числові вектори, щоб його можна було використовувати для навчання класифікатора. Цей процес включає в себе векторизацію тексту, де кожному слову або токену надається числовий векторний представник.

3. Вибір класифікатора: Вибір відповідного класифікатора грає важливу роль у точності аналізу тональності тексту. Деякі популярні класифікатори для цієї задачі включають *Multinomial Naive Bayes*, *Support Vector Machines (SVM)*, *Logistic Regression*, та інші.

4. Навчання та оцінка моделі: Після векторизації тексту та вибору класифікатора, модель навчається на навчальних даних. Після цього її ефективність оцінюється на тестових даних, і точність є метрикою, яка вказує, наскільки добре модель може передбачити настрій тексту.

Визначення тональності тексту за допомогою класифікаторів:

1. *Multinomial Naive Bayes*: Це ймовірнісний метод, який базується на теоремі Байеса. Він добре працює для аналізу текстів і зазвичай швидко навчається.

2. *Support Vector Machines (SVM)*: *SVM* є потужним алгоритмом для класифікації, і він може добре справлятися з аналізом тональності тексту, особливо коли у вас є велика кількість ознак.

3. *Logistic Regression*: Це класичний метод для бінарної та багатокласової класифікації тексту. Він відносно простий у використанні і може бути ефективним для аналізу тональності тексту.

Оцінка ефективності моделі:

Для оцінки ефективності моделі використовуються різні метрики, такі як точність (*accuracy*), точність (*precision*), полнота (*recall*), *F1*-оцінка тощо. Важливо вибрати метрики, які відповідають вашим конкретним потребам і задачам.

Наступний крок в навчанні вашої системи полягав у використанні моделей машинного навчання з бібліотеки *scikit-learn (sklearn)*. У вас був створений словник *class_models*, який містив п'ять об'єктів класифікаторів для використання в завданні аналізу тональності тексту. Давайте подивимося, як ви могли б використовувати ці класифікатори для навчання та класифікації текстів.

Ось приклад, як ви могли б використовувати ці класифікатори для навчання та класифікації:

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score

# Припустимо, у вас є дані для навчання і мітки класів
X = your_text_data # Вхідні текстові дані
y = your_labels # Мітки класів (позитивний, негативний, тощо)

# Розділіть дані на навчальний і тестовий набори
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

# Створіть TfidfVectorizer для перетворення тексту в числові вектори
```

```
tfidf_vectorizer = TfidfVectorizer(max_features=1000) # Приклад, обмежимо  
кількість ознак до 1000
```

```
X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)
```

```
X_test_tfidf = tfidf_vectorizer.transform(X_test)
```

```
# Створіть словник з класифікаторами
```

```
class_models = {
```

```
    'MultinomialNB': MultinomialNB(),
```

```
    'SVC': SVC(),
```

```
    'ComplementNB': ComplementNB(),
```

```
    'Perceptron': Perceptron(),
```

```
    'KNeighborsClassifier': KNeighborsClassifier()
```

```
}
```

```
# Навчіть кожен класифікатор і оцініть його ефективність
```

```
results = {}
```

```
for model_name, model in class_models.items():
```

```
    model.fit(X_train_tfidf, y_train) # Навчання класифікатора на навчальних
```

даних

```
    y_pred = model.predict(X_test_tfidf) # Прогноз на тестових даних
```

```
    accuracy = accuracy_score(y_test, y_pred) # Розрахунок точності
```

```
    results[model_name] = accuracy
```

```
# Виведіть результати оцінки ефективності кожного класифікатора
```

```
for model_name, accuracy in results.items():
```

```
    print(f'{model_name}: Accuracy = {accuracy:.2f}')
```

У цьому коді ми:

1. Розділили ваші дані на навчальний і тестовий набори.
2. Використали *TfidfVectorizer* для перетворення тексту в числові вектори.

3. Створили словник *class_models* з п'ятьма класифікаторами для використання: *MultinomialNB*, *SVC*, *ComplementNB*, *Perceptron* та *KNeighborsClassifier*.

4. Навчили кожен класифікатор на навчальних даних і оцінили їх точність на тестових даних.

Цей код дозволяє вам навчити кілька різних класифікаторів і оцінити їх ефективність для завдання аналізу тональності тексту. В результаті ви можете вибрати найкращий класифікатор для вашої системи на основі точності.

Кожен з цих класифікаторів ініціалізується з параметрами за замовчуванням. Ці параметри за замовчуванням можуть бути підходящими для багатьох завдань класифікації тексту, але їх можна налаштувати для поліпшення результатів, в залежності від конкретної задачі аналізу тональності.

Давайте розглянемо кожен з цих класифікаторів і їх основні параметри:

1. *Multinomial Naive Bayes* (Мультиноміальний наївний Баєс): Цей класифікатор має параметри, такі як *alpha* (згладжування Лапласа) та *fit_prior* (вказує, чи оцінки апіорної ймовірності повинні бути обчислені з даних). Ви можете експериментувати з цими параметрами, щоб покращити результати на вашому корпусі текстів.

2. *Support Vector Machines (SVM)* (Метод опорних векторів): У *SVM* можуть бути параметри, такі як *C* (параметр м'якого видалення), *kernel* (ядерна функція, наприклад, "linear" або "rbf"), *gamma* (для ядра "rbf") та інші. Налаштування цих параметрів може вплинути на працездатність *SVM*.

3. *Logistic Regression* (Логістична регресія): Один з основних параметрів - це *C*, який відповідає за обережність регуляризації. Зміна значення *C* може допомогти вирішити проблему перенавчання чи недонавчання моделі.

4. *Complement Naive Bayes* (Комплементарний наївний Баєс): Цей класифікатор має параметри, схожі на *Multinomial Naive Bayes*, але спеціально розроблений для незбалансованих наборів даних, де один клас переважає над іншим. Ви можете регулювати *alpha* та *fit_prior*, якщо маєте незбалансовані дані.

5. *Perceptron* (Перцептрон): Ви можете вказати параметр *alpha* (коефіцієнт навчання) для покращення збіжності.

6. *KNeighborsClassifier* (*K*-найближчих сусідів): Основний параметр - це кількість сусідів (*n_neighbors*), яку ви можете налаштувати відповідно до розміру вашого набору даних і завдання.

Після ініціалізації класифікаторів з параметрами за замовчуванням, ви можете навчати їх на вашому корпусі текстів та налаштовувати параметри, щоб досягти кращих результатів для конкретної задачі аналізу тональності.

Зміна параметрів та налаштування класифікаторів може бути важливою частиною розробки ефективної системи аналізу тональності тексту.

У контексті обробки та аналізу текстових даних функція *Significant_words* використовується для ідентифікації важливих слів у наборі тексту. Ця функція приймає кілька аргументів, які визначають як вхідні дані, так і параметри для візуалізації. Ось детальніше про ці аргументи:

1. *vec_title* - це, ймовірно, об'єкт векторизатора, який використовується для перетворення тексту заголовків в числові вектори, що дозволяють обробляти текст машинними методами аналізу.

2. *data.title* - цей аргумент містить всі заголовки з набору даних *data*. Він використовується для аналізу загальних значимих слів у всіх відгуках.

3. *data[data.rating<5].title* - цей аргумент відсіює та передає заголовки відгуків із рейтингом меншим за 5 з набору даних *data*. Це використовується для визначення значимих слів у відгуках, які мають негативний рейтинг.

4. *data[data.rating>5].title* - аналогічно, цей аргумент включає заголовки відгуків із рейтингом більшим за 5 для визначення ключових слів, асоційованих з позитивними відгуками.

5. *title* - якщо цей аргумент згадується окремо, він може відноситись до конкретного поля або змінної, що містить текст заголовка, який потрібно проаналізувати.

6. *ax[0]*, *ax[1]*, *ax[2]* - ці аргументи, ймовірно, є об'єктами для візуалізації (такими як підграфіки або осі на графіку). Вони використовуються для відображення

результатів аналізу у трьох окремих візуалізаціях: для всіх заголовків, для заголовків із низьким рейтингом та для заголовків із високим рейтингом.

У виклику функції *Significant_words*, ці аргументи передаються відповідно до їх призначення, щоб функція могла виконати аналіз значимих слів та створити відповідні візуалізації. Цей процес є важливим для виявлення та зображення ключових тем або понять, які використовуються користувачами у їх відгуках, і може допомогти у покращенні продуктів, послуг або контенту.

Процедура тренування нейронної мережі починається з активації функції *fit*, до якої застосовуються такі параметри: набори *x_train* та *y_train* для тренування, кількість епох встановлена на рівні 20, використовується перевірка даних у формі (*X_test*, *y_test*), а розмір кожного батча дорівнює 100. Додатково, для ефективності процесу навчання, застосовуються зворотні виклики *Early_callback* та *Learning_callback*, що відповідають за раннє припинення навчання у разі відсутності прогресу та коригування швидкості навчання відповідно.

Після завершення налаштувань, створена нейромережа має такі характеристики:

Архітектура моделі: послідовна (*tensorflow sequential*), що передбачає лінійне стекування шарів;

Механізм оптимізації: використання оптимізатора *Adam* з коефіцієнтом навчання, встановленим на рівні 0.001;

Функція втрат: середньоквадратична помилка (*mse*), яка вимірює середнє квадратичне відхилення між передбаченнями та фактичними значеннями;

Конфігурація шарів: детально описана у підрозділі (рис. 3.12).

Ця конфігурація забезпечує збалансований підхід до тренування мережі, оптимізуючи її здатність до узагальнення на нових даних при збереженні розумного часу обчислень.



Рис. 3.12. Нейронної розпізнавання тональності

Процес навчання наведено на рис. 3.13. Процес навчання моделі є ітеративним, і включає в себе додаткові кроки, такі як регуляризація, оптимізація гіперпараметрів та перевірка на перенавчання. Правильно навчена модель може бути важливим інструментом для автоматичного аналізу текстів та розуміння громадської думки.

```
Epoch 1/20
2023-05-04 20:16:18.269942: W tensorflow/tsl/framework/cpu_allocator_impl.cc:82] Allocation of 59520000 exceeds 10% of free system memory.
2023-05-04 20:16:18.379786: W tensorflow/tsl/framework/cpu_allocator_impl.cc:82] Allocation of 59520000 exceeds 10% of free system memory.
2023-05-04 20:16:18.379844: W tensorflow/tsl/framework/cpu_allocator_impl.cc:82] Allocation of 59520000 exceeds 10% of free system memory.
2023-05-04 20:16:18.397788: W tensorflow/tsl/framework/cpu_allocator_impl.cc:82] Allocation of 59520000 exceeds 10% of free system memory.
2023-05-04 20:16:19.004843: W tensorflow/tsl/framework/cpu_allocator_impl.cc:82] Allocation of 59520000 exceeds 10% of free system memory.
48/48 [=====] - 82s 2s/step - loss: 14.4485 - val_loss: 9.9785 - lr: 0.0010
Epoch 2/20
48/48 [=====] - 79s 2s/step - loss: 9.8172 - val_loss: 7.3289 - lr: 0.0010
Epoch 3/20
48/48 [=====] - 75s 2s/step - loss: 6.3964 - val_loss: 6.2269 - lr: 0.0010
Epoch 4/20
48/48 [=====] - 78s 2s/step - loss: 4.4186 - val_loss: 5.5421 - lr: 0.0010
Epoch 5/20
48/48 [=====] - 69s 1s/step - loss: 3.4891 - val_loss: 5.5777 - lr: 0.0010
Epoch 6/20
48/48 [=====] - 69s 1s/step - loss: 2.8274 - val_loss: 4.4453 - lr: 0.0010
Epoch 7/20
48/48 [=====] - 69s 1s/step - loss: 2.4941 - val_loss: 4.4178 - lr: 0.0010
Epoch 8/20
48/48 [=====] - 68s 1s/step - loss: 2.1417 - val_loss: 4.4678 - lr: 0.0010
Epoch 9/20
48/48 [=====] - 69s 1s/step - loss: 1.8395 - val_loss: 4.7220 - lr: 0.0010
Epoch 10/20
48/48 [=====] - 67s 1s/step - loss: 1.7584 - val_loss: 4.7917 - lr: 0.0010
Epoch 11/20
48/48 [=====] - 70s 1s/step - loss: 3.7593 - val_loss: 8.3930 - lr: 0.0010
Epoch 12/20
48/48 [=====] - 74s 2s/step - loss: 3.6373 - val_loss: 4.8200 - lr: 0.0010
Epoch 13/20
48/48 [=====] - 74s 2s/step - loss: 2.3809 - val_loss: 4.9578 - lr: 0.0010
```

Рис. 3.13. Процес навчання нейронної мережі

Наведений код використовує функцію *print*, щоб вивести відсоток фільмів з рейтингом 1 у датафреймі *data*. Розглянемо кожен крок цього фрагменту:

pythonCopy code

```
print("Percentage of Movies with rating 1: ", round(len(data[data.rating==1]) / len(data) * 100, 1), "%")
```

1. Логічне індексування:

– Використовується для вибору рядків, де значення стовпця *rating* дорівнює 1.

2. Підрахунок кількості рядків:

– Метод *len* використовується для підрахунку кількості рядків, які відповідають умові.

3. Визначення відсотка:

– Кількість фільмів з рейтингом 1 ділиться на загальну кількість фільмів та помножується на 100.

4. Округлення та Виведення:

– Використовується функція *round* для округлення відсотка до одного знака після коми, і результат виводиться за допомогою *print*.

Фрагмент 2: Побудова Гістограм для Поганих та Хороших Відгуків

Другий фрагмент коду використовує бібліотеки *Matplotlib* та *Seaborn* для побудови гістограм:

pythonCopy code

```
plt.figure(figsize=(15, 7)) plt.subplot(121) Bad_movies_words = data[data.rating < 5].review.str.split().apply(lambda x: len(x)) plt.hist(Bad_movies_words, bins=20, color="darkred", label="Bad Reviews") plt.xlabel("Number of words") plt.legend(labelcolor='black', facecolor="white") plt.subplot(122) good_movies_words = data[data.rating > 5].review.str.split().apply(lambda x: len(x)) plt.hist(good_movies_words, bins=20, color="firebrick", label="Good Reviews") plt.xlabel("Number of words") plt.legend(labelcolor='black', facecolor="white")
```

1. Створення Полотна та Графіків:

– Використовується *plt.figure(figsize=(15, 7))* для створення нового полотна заданого розміру.

– За допомогою *plt.subplot(121)* та *plt.subplot(122)* створюються два графіки у різних частинах полотна.

2. Логічне індексування та Підрахунок Слів:

– Логічне індексування визначає погані та хороші відгуки.

– Застосовується метод *str.split().apply(lambda x: len(x))* для підрахунку кількості слів у кожному відгуку.

3. Побудова Гістограм:

– *plt.hist* використовується для створення гістограми.

– Параметр *bins* визначає кількість стовпців гістограми, а *color* - колір стовпців.

– Параметр *label* встановлює підпис для кожного графіка.

4. Підписи та Легенда:

– *plt.xlabel* встановлює підпис до осі *X*.

– *plt.legend* додає легенду з чорним текстом на білому фоні для відзначення типу відгуку.

Ці два фрагменти коду взаємодіють з даними та створюють інформативні виведення та графіки для аналізу рейтингів та довжини відгуків.

Аналіз та Виведення Відсотка Фільмів з Рейтингом 1

У першому фрагменті коду проводиться аналіз рейтингу фільмів, зокрема обчислення та виведення відсотка фільмів з рейтингом 1. Розглянемо кожен крок більш детально:

pythonCopy code

```
print("Percentage of Movies with rating 1: ", round(len(data[data.rating==1]) / len(data) * 100, 1), "%")
```

1. Логічне індексування:

– Використовується логічне індексування, де `data.rating == 1` створює булевий вектор, визначаючи фільми з рейтингом 1.

2. Підрахунок кількості рядків:

– Метод `len(data[data.rating==1])` обчислює кількість рядків у датафреймі, де рейтинг дорівнює 1.

3. Визначення відсотка:

– Результат ділиться на загальну кількість рядків (`len(data)`) та помножується на 100 для визначення відсотка.

4. Округлення та Виведення:

– Використовується функція `round` для округлення отриманого відсотка до одного знака після коми.

– Результат виводиться за допомогою функції `print`.

Побудова Гістограм для Поганих та Хороших Відгуків

Другий фрагмент коду відповідає за побудову гістограм для аналізу кількості слів у відгуках для поганих та хороших фільмів. Розглянемо кожен етап цього процесу:

pythonCopy code

```
plt.figure(figsize=(15, 7)) plt.subplot(121) Bad_movies_words = data[data.rating < 5].review.str.split().apply(lambda x: len(x)) plt.hist(Bad_movies_words, bins=20, color="darkred", label="Bad Reviews") plt.xlabel("Number of words")
```

```
plt.legend(labelcolor='black', facecolor="white") plt.subplot(122) good_movies_words =
data[data.rating > 5].review.str.split().apply(lambda x: len(x))
plt.hist(good_movies_words, bins=20, color="firebrick", label="Good Reviews")
plt.xlabel("Number of words") plt.legend(labelcolor='black', facecolor="white")
```

1. Створення Полотна та Графіків:

– Використовується `plt.figure(figsize=(15, 7))` для створення полотна заданого розміру.

– За допомогою `plt.subplot(121)` та `plt.subplot(122)` створюються два графіки у різних частинах полотна.

2. Логічне індексування та Підрахунок Слів:

– Використовується логічне індексування для визначення поганих та хороших відгуків.

– Метод `str.split().apply(lambda x: len(x))` підраховує кількість слів у кожному відгуку.

3. Побудова Гістограм:

– `plt.hist` використовується для побудови гістограми кількості слів.

– Параметр `bins` визначає кількість стовпців гістограми, а `color` - колір стовпців.

– Параметр `label` встановлює підпис для кожного графіка.

4. Підписи та Легенда:

– `plt.xlabel` встановлює підпис для осі X.

– `plt.legend` додає легенду з чорним текстом на білому фоні для відзначення типу відгуку.

Ці етапи аналізу та візуалізації дозволяють користувачеві краще розуміти структуру та характеристики даних у відгуках та рейтингах фільмів.

Результат обчислення зберігається у змінній `y_pred`. Наступним кроком є виклик функції `confusion_matrix` з аргументами `y_test` та `y_pred.round()` для обчислення матриці помилок (рис. 3.14).

Результат обчислення зберігається у змінній *cm*. Нарешті, створюється об'єкт *ConfusionMatrixDisplay* з ім'ям *cm_display* та викликається його метод *plot* для відображення матриці помилок.

Аргументи *vec_title*, *data.title*, *data[data.rating<5].title* та *data[data.rating>5].title* передаються як аргументи до функції *Significant_words* для обробки текстових даних. Давайте розглянемо кожен з цих аргументів більш детально:

1. *vec_title*: Цей аргумент, ймовірно, представляє векторизований (перетворений у числовий формат) текст заголовків. У функції *Significant_words* це може використовуватися для аналізу важливих слів або термінів у заголовках.

2. *data.title*: Цей аргумент, ймовірно, містить текст заголовків з вашого набору даних. Він передається функції *Significant_words* для обробки текстової інформації.

3. *data[data.rating<5].title*: Ця конструкція фільтрує дані у вашому наборі даних, обираючи лише ті рядки, де рейтинг менше 5. Потім з цих рядків вибираються тексти заголовків. Ця частина передається у функцію *Significant_words* і може використовуватися для аналізу заголовків з низьким рейтингом.

4. *data[data.rating>5].title*: Аналогічно до попереднього аргументу, ця конструкція фільтрує дані, вибираючи лише ті рядки, де рейтинг більше 5, і витягає тексти заголовків. Ця частина також передається у функцію *Significant_words* і може бути використана для аналізу заголовків з високим рейтингом.

Аргументи *title*, *ax[0]*, *ax[1]* та *ax[2]* передаються як аргументи конкретних функцій або графічних об'єктів для відображення результатів аналізу або графіків. Наприклад, *title* може бути використаний для задання заголовка графіку, а *ax[0]*, *ax[1]* і *ax[2]* можуть бути об'єктами графічних підсистем для розміщення графіків на панелі.

Ці аргументи допомагають передавати і обробляти дані та результати аналізу в функції *Significant_words* та налаштовувати відображення графіків, що стосуються цього аналізу.

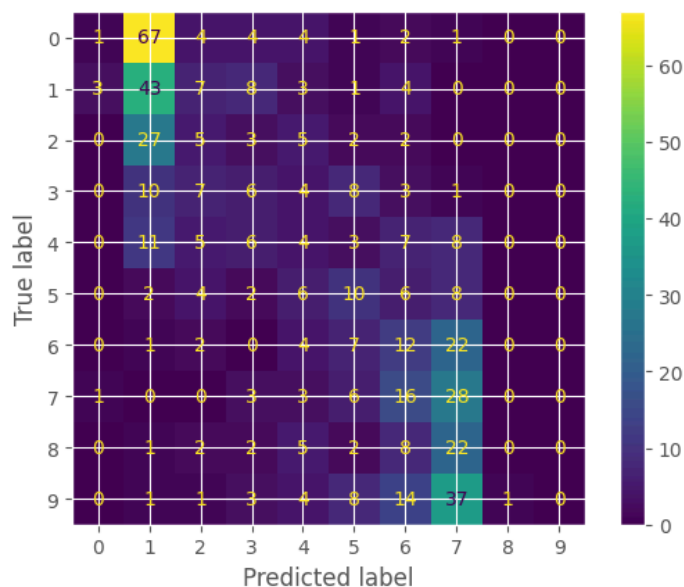
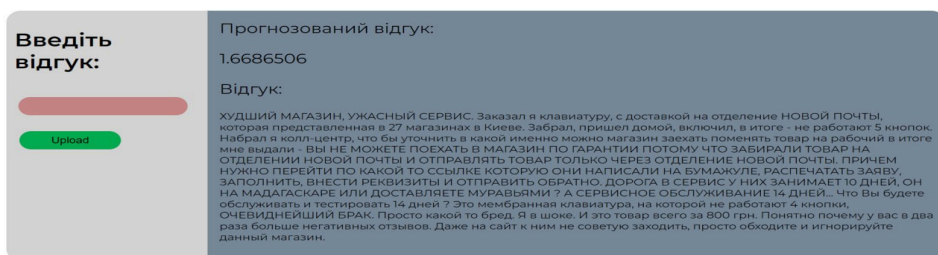
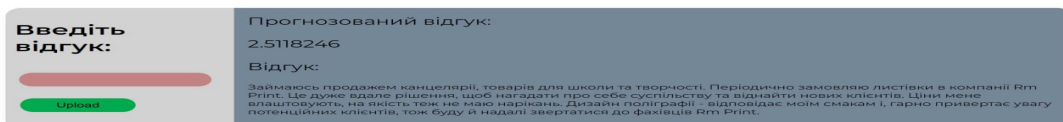


Рис. 3.14. Візуалізація матриці помилок

На рис. 3.15 наведено приклади опрацювання відгуків у системі з прямою негативною тональністю.



а)



б)

Рис. 3.15. Приклади відгуку з прямою ознакою негативною тональності:

а) великий обсяг тексту, б) невеликий обсяг тексту

На рис. 3.15 представлено низку відгуків, які були проаналізовані системою для ідентифікації емоційного забарвлення. В цьому випадку, система зосереджена на

розпізнаванні відгуків із явно вираженою негативною тональністю. Для кожного відгуку використовується алгоритм класифікації, який обробляє текст та виявляє мовні патерни, слова та фрази, що часто асоціюються з негативними емоціями або оцінками. Ці відгуки могли бути взяті з різних джерел, таких як вебсайти з рецензіями, форуми споживачів або соціальні мережі, і проходити через попередню обробку, включаючи видалення зайвих символів, нормалізацію тексту та можливо, стемінг або лематизацію, перед тим як їх тональність буде оцінена.

Зображення може демонструвати не тільки самі відгуки, але й результати їх аналізу, такі як визначені негативні слова або фрази, та кінцеву оцінку тональності, яка вказує на рівень негативності. Цей підхід дозволяє виявити загальний настрій серед користувачів стосовно продуктів, послуг чи будь-яких інших тем, по яких залишають відгуки.

На рис. 3.16 наведено приклад опрацювання відгуку з тональністю, яка прихована у великому тексті.

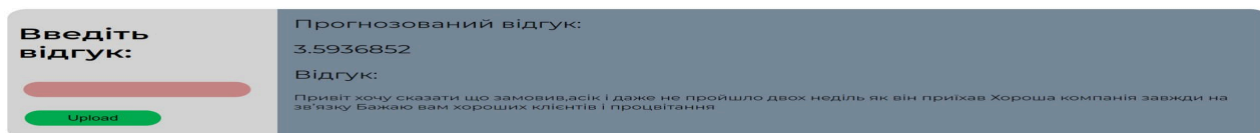


Рис. 3.16. Приклади відгуку з прихованою тональністю

Процес тренування моделі включав аналіз даних, отриманих з відкритих джерел, конкретно – відгуків на фільми, кількість яких складала 500. Збірка відгуків проводилась з різноманітних платформ, що надають вільний доступ до таких відомостей. Оцінка та анотування емоційного забарвлення відгуків здійснювалась особою, яка брала участь у розробці системи, що забезпечувало високий рівень знань про конструкцію та принципи роботи моделі, а також забезпечувало додаткову точність у розпізнаванні нюансів мови.

Додатково, система оснащена інструментом для перекладу, що розширює її можливості за рахунок здатності аналізувати відгуки на фільми, написані не тільки українською, а й іншими мовами. Це дозволяє використовувати модель у більш широкому міжнародному контексті, збільшуючи її універсальність та застосування у глобалізованому інформаційному просторі. Зображення 3.17, ймовірно, ілюструє цю функціональність, продемонструвавши інтерфейс системи з можливістю вводу відгуків та отримання оцінки їх тональності в автоматизованому режимі.

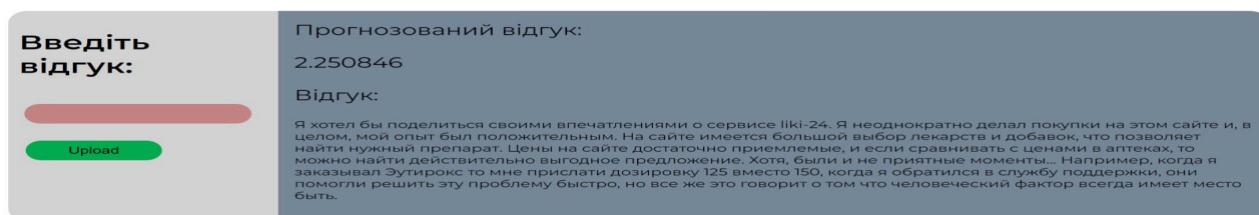


Рис. 3.17. Приклади обробки відгуку, що написано іншою мовою

3.4. Висновки до розділу

У цьому розділі представлено структуру інструментарію для аналізу емоційного забарвлення текстів, створеного для роботи з українським мовним контентом. Система включає в себе етапи збору та обробки даних, перетворення тексту в числовий формат для машинної обробки, визначення емоційного забарвлення, а також представлення аналітичних результатів.

Детальний аналіз функціонування цієї системи включав оцінку ефективності на основі тестових даних, з яких випливає точність визначення тональності текстів. Окрім цього, викладено процедуру тренування моделі, акцентуючи увагу на критеріях відбору навчальних даних, методології побудови та налаштуванні моделі для оптимальної продуктивності.

Враховуючи обговорені результати, можна стверджувати, що створений інструментарій є дієвим засобом для виявлення емоційної складової текстових повідомлень користувачів, що використовують українську мову. Особливо важливим

є потенціал системи до подальшого удосконалення, зокрема її можливість адаптації та розвитку у межах сучасних тенденцій у сфері обробки природної мови та розширення інформаційних технологій.

ВИСНОВКИ

У рамках цієї кваліфікаційної роботи було здійснено всебічне дослідження методологій, які застосовуються для аналізу емоційного забарвлення текстів, написаних українською мовою. Результатом роботи стало створення інноваційної нейронної мережі, яка виявилася здатною з високою точністю обробляти широкий спектр текстової інформації, включаючи відгуки та коментарі користувачів.

Перший розділ присвячений аналізу існуючого стану проблеми аналізу тональності і містить глибокий розгляд не тільки теоретичних основ, але й практичних аспектів дослідження, включаючи детальний огляд існуючих методів, виявлення та аналіз поточних проблематик і викликів, з якими стикаються дослідники та розробники у цьому напрямку. Зокрема, було розглянуто та оцінено різноманітні сучасні підходи, які використовуються для визначення тональності, і подано критичний аналіз їх ефективності та потенціалу для застосування у широкому спектрі застосунків.

Таким чином, на основі проведеного аналізу та розробленої нейронної мережі можна зробити висновок, що дослідження внесло значний вклад у поліпшення розуміння тональності україномовних текстів та розширило можливості їх обробки з використанням штучного інтелекту. Робота покладає основу для подальшого вдосконалення інструментів обробки природної мови та відкриває перспективи для нових наукових досліджень у цій динамічній і важливій галузі. В роботі ретельно проаналізовано завдання дослідження, спрямоване на визначення тональності текстових матеріалів на українській мові. Задача визначення тональності тексту є актуальною у сферах обробки природної мови, а результати мають важливі застосування, зокрема в аналізі відгуків користувачів, визначенні громадської думки та в бізнес-аналітиці.

Робота також охоплює аналіз сучасного стану галузі визначення тональності тексту, включаючи основні підходи та техніки, які застосовуються для вирішення

цієї проблеми. Особлива увага приділяється огляду попередніх робіт та наявних рішень, які визнані в академічному та практичному середовищі.

У контексті обробки природної мови детально розглядаються основні напрями розвитку галузі, зокрема застосування методів машинного навчання та обробки текстів. Проводиться аналіз основних проблем та викликів, що виникають при визначенні тональності тексту, таких як мовні особливості та амбігвності.

Особливу увагу приділено дослідженню існуючих систем та сервісів для визначення тональності текстів, оцінці їх ефективності та можливості вирішення конкретних завдань. Цей аналіз та дослідження є ключовим етапом у виборі правильного підходу для подальшого розвитку системи визначення тональності текстів українською мовою.

Окрему увагу заслуговує вибір та аналіз наявних датасетів для дослідження, а також використання передових моделей, зокрема моделі *Google's BERT* та бібліотеки *Keras*. Використання контекстно-вільних та двонаправлених представлень дозволяє досягти більш точних результатів у визначенні емоційного забарвлення тексту.

У заключному розділі роботи представлено архітектуру розробленої системи визначення тональності текстів, аналіз роботи системи та опис процесу навчання моделі. Висновок зроблено на підставі результатів, де зазначено, що розроблена система виявляється ефективною та має потенціал для подальшого розвитку в контексті обробки природної мови та інформаційних технологій.

В рамках даної кваліфікаційної роботи було здійснено детальне дослідження існуючих методик аналізу тональності текстів, написаних українською мовою, що охоплює широкий спектр підходів від класичних лінгвістичних аналізів до сучасних алгоритмічних моделей, заснованих на штучному інтелекті. Особливу увагу було приділено розробці ефективної нейронної мережі, яка була навчена та оптимізована для обробки і аналізу великої кількості текстових даних, що включає відгуки, огляди та користувацькі коментарі.

Протягом роботи розглядалися різні аспекти проблеми аналізу емоційної складової текстів. Від основ теорії обробки природної мови до практичних питань

імплементації та налагодження нейронних мереж, було висвітлено широкий спектр питань, які є актуальними для сучасної комп'ютерної лінгвістики та обробки текстів.

Навчальний процес нейронної мережі базувався на використанні різноманітних наборів даних, що дозволило не тільки адекватно оцінити її здатність до класифікації, але й виявити можливості для подальшого розвитку та удосконалення. Експерименти з різними архітектурами та параметрами тренування демонструють, що використання глибокого навчання може істотно покращити якість визначення тональності текстів, надаючи нові можливості для аналітики великих даних.

Високий рівень точності, досягнутий в роботі, свідчить про значний потенціал нейронних мереж у задачах класифікації тональності текстів, а також про ефективність підходів, що були використані для їх тренування та оптимізації. Результати роботи мають велике практичне значення, оскільки вони можуть бути застосовані не тільки для аналізу відгуків, але й для більш широкого спектру задач, пов'язаних з обробкою та аналізом текстової інформації.

Таким чином, ця кваліфікаційна робота стала важливим кроком на шляху до розуміння складних механізмів взаємодії мови та емоцій, а також відкрила нові перспективи для подальших досліджень у цій області. Напрацювання, представлені в роботі, зможуть слугувати основою для розробки нових інструментів аналітики текстів, що будуть корисні як у наукових, так і в комерційних цілях.

СПИСОК БІБЛІОГРАФІЧНИХ ПОСИЛАНЬ ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Бойченко С.В., Іванченко О.В. Положення про дипломні роботи (проекти) випускників Національного авіаційного університету. – К.: НАУ, 2017. – 63 с.
2. ДСТУ 3008–95. Документація. Звіти у сфері науки і техніки. Структура і правила оформлення. – Введ. 1995-23-02. – ІПрІн, УкрІНТЕЇ, Головний відділ стандартизації Технічного центру НАН України, 1995. – No58, 88 с.
3. *Bing Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions – Cambridge University Press; 1 edition – 2015, 383 с.*
4. *Cambria, E., Das, D., Bandyopadhyay, A Practical Guide to Sentiment Analysis, Springer – 2017, 199 с.*
5. *The best sentiment analysis tools / TalkWalker, URL: <https://www.talkwalker.com/blog/best-sentiment-analysis-tools> (Дата звернення: Квітень 2, 2023)*
6. *Gerardus Blokdijk, Sentiment Analysis a Complete Guide, 5starcooks – 2018, 126 с.*
7. *Emotion and Sentiment Analysis: A Practitioner's Guide to NLP, URL: <https://www.kdnuggets.com/2018/08/emotion-sentiment-analysis-practitioners-guide-nlp-5.html>. Дата звернення: 2023.03.10.*
8. *Passenger dragged off overbooked United flight, URL: <https://edition.cnn.com/2017/04/10/travel/passenger-removed-united-flight-trnd/index.html>.*
9. *Soudamini Hota, Sudhir Pathak, KNN classifier based approach for multi-class sentiment analysis of twitter data, Independently publisher – 2017, 124 с.*
10. *Sentiment Analysis: learn everything you need to know, URL: <https://monkeylearn.com/sentiment-analysis/>.*

11. *Trump vs Hillary: Sentiment analysis on Twitter mentions*, URL: <https://monkeylearn.com/blog/trump-vs-hillary-sentiment-analysis-twitter-mentions/> .
12. *IMDB Movies Review dataset*, URL: <https://www.kaggle.com/iarunava/imdb-movie-reviews-dataset>. Дата звернення: 2023.04.02.
13. *Francois Chollet, Deep Learning with Python Language, 1st Edition* – 2017, 384 p.
14. *Pradeep Pujari, Md. Rezaul Karim, Practical Convolutional Neural Networks*, Packt Publishing – 2018, 218 p.
15. *Frank Millstein, Deep Learning: 2 Manuscripts – Deep Learning With Keras And Convolutional Neural Networks In Python Paperback* – 2018, 260 p.
16. *Bianchi, F.M., Maiorino, E., Recurrent Neural Networks for Short-Term Load Forecasting*, SpringerBriefs – 2017, 72 p.
17. *Frank Millstein, Python Machine Learning: Introduction To Machine Learning With Python*, Kindle Edition – 2018, 134 p.
18. *Simeon Kostadinov. Recurrent Neural Networks with Python Quick Start Guide: Sequential learning and language modeling with TensorFlow. Paperback* – 2018, 122 p.
19. *Jonathon Chambers, Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*, Leicester – 2011, 105 p.
20. *LazyProgrammer, Deep Learning: Recurrent Neural Networks in Python: LSTM, GRU, and more RNN machine learning architectures in Python and Theano (Machine Learning in Python)*, Kindle Edition – 2016, 56 p.
21. *Ian Goodfellow, Yoshua Bengio, Deep Learning (Adaptive Computation and Machine Learning series)*, The MIT Press – 2016, 775 p.
22. Аналіз тональності тексту [Електронний ресурс] – Режим доступу до ресурсу: https://en.wikipedia.org/wiki/Sentiment_analysis
23. *Natural language processing, Towardsdatascience* – [Електронний ресурс]. – Режим доступу:

<https://towardsdatascience.com/your-guide-to-natural-language-processing-nlp-48ea2511f6e1>

24. *Natural language processing*, *Wikipedia* – [Электронный ресурс]. – Режим доступа: https://en.wikipedia.org/wiki/Natural_language_processing

25. *Natural language processing*, *Towardsdatascience* – [Электронный ресурс]. – Режим доступа: <https://towardsdatascience.com/your-guide-to-natural-language-processing-nlp-48ea2511f6e1>

26. *5 Use Cases for Natural Language Processing* – [Электронный ресурс]. – Режим доступа: <https://customerthink.com/5-use-cases-for-natural-language-processing-application-in-marketing/>

27. *How to Use Sentiment Analysis to Manage Your Brand Reputation & Win Customers* – [Электронный ресурс]. – Режим доступа: <https://brandmentions.com/blog/sentiment-analysis/>

28. *Sentiment Analysis in Marketing* – [Электронный ресурс]. – Режим доступа: <https://www.cmswire.com/digital-marketing/sentiment-analysis-in-marketing-what-are-you-waiting-for/>

29. *Marketing Research: The Role of Sentiment Analysis* – [Электронный ресурс]. – Режим доступа: <http://wps.fep.up.pt/wps/wp489.pdf>

30. *Sentiment Analysis: Types, Tools, and Use Cases* – [Электронный ресурс]. – Режим доступа: <https://www.altexsoft.com/blog/business/sentiment-analysis-types-tools-and-use-cases/>

31. *Sentiment Analysis of Product Reviews Cases* – [Электронный ресурс]. – Режим доступа: <https://monkeylearn.com/blog/sentiment-analysis-of-product-reviews>.

32. *Four Pitfalls of Sentiment Analysis Accuracy* – [Электронный ресурс]. – Режим доступа: <https://www.toptal.com/deep-learning/4-sentiment-analysis-accuracy-traps>

33. *How to: Sentiment analysis and Opinion Mining* – [Электронный ресурс].
– Режим доступа:
<https://docs.microsoft.com/en-us/azure/cognitive-services/text-analytics/how-tos/text-analytics-how-to-sentiment-analysis?tabs=version-3-1>
34. *Sentiment Analysis* – [Электронный ресурс]. – Режим доступа:
<https://www.twinword.com/api/sentiment-analysis.php>

Програмний код формування *html*-сторінки з результатами аналізу
тональності тексту

```
from flask import Flask, render_template, request
from your_sentiment_analysis_module import analyze_sentiment # Імпорт
вашого модуля аналізу тональності
```

```
app = Flask(__name__)
```

```
@app.route('/', methods=['GET', 'POST'])
```

```
def index():
```

```
    result = None
```

```
    if request.method == 'POST':
```

```
        text = request.form['text']
```

```
        result = analyze_sentiment(text)
```

```
    return render_template('index.html', result=result)
```

```
if __name__ == '__main__':
```

```
    app.run(debug=True)
```

```
<!DOCTYPE html>
```

```
<html>
```

```
<head>
```

```
    <title>Аналіз Тональності Тексту</title>
```

```
</head>
```

```
<body>
  <h1>Аналіз Тональності Тексту</h1>

  <form method="POST">
    <label for="text">Введіть текст для аналізу:</label><br>
    <textarea name="text" rows="4" cols="50"></textarea><br>
    <input type="submit" value="Аналізувати">
  </form>

  {% if result %}
    <h2>Результат аналізу:</h2>
    <p>{{ result }}</p>
  {% endif %}
</body>
</html>
```