Міністерство освіти і науки України
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА АКАДЕМІЯ»
Факультет інформатики
Кафедра мультимедійних систем

Магістерська робота

Освітній ступінь: магістр

на тему: **«СОЦІАЛЬНЕ ПРОСЛУХОВУВАННЯ ДЛЯ УПРАВЛІННЯ РЕПУТАЦІЄЮ БРЕНДУ НА ПРИКЛАДІ ПРОГРАМНОГО РІШЕННЯ (ПРОТОТИПУ), ЯКІ ДОПОМАГАЮТЬ КОМПАНІЯМ ВІДСТЕЖУВАТИ ТА КЕРУВАТИ РЕПУТАЦІЄЮ СВОГО БРЕНДУ В ІНТЕРНЕТІ ЗА ДОПОМОГОЮ АНАЛІЗУ СОЦІАЛЬНИХ МЕРЕЖ І ВІДСТЕЖЕННЯ НАСТРОЇВ»**

Виконав: студент 2-го року навчання,
Спеціальності
121 Інженерія програмного забезпечення
Волошко Максим Володимирович
Керівник _____Афонін А.О._____
доцент

Рецензент _____
Магістерська робота захищена
з оцінкою _____
Секретар ЕК _____
«_____» _____ 2024 р.

Київ 2024

Міністерство освіти і науки України
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА АКАДЕМІЯ»
Кафедра інформатики факультету інформатики

ЗАТВЕРДЖУЮ
Зав. кафедри інформатики,
к.ф.-м.н., доц. Гороховський С. С.

_____
(підпис)
«_____» _____ 2024 р.

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ
на магістерську роботу

студента 2 р. н. магістерської програми Інженерія програмного забезпечення
Волошка Максима Володимировича

Соціальне прослуховування для управління репутацією бренду на прикладі програмного рішення (прототипу), які допомагають компаніям відстежувати та керувати репутацією свого бренду в Інтернеті за допомогою аналізу соціальних мереж і відстеження настроїв

Зміст текстової частини до магістерської роботи:

Зміст
Анотація
Вступ
1 Аналіз предметної області
2 Дослідження аналізу тональності тексту
3 Аналіз моніторингу соціальних мереж
4 Розробка проекту системи-прототипу
Висновки
Список літератури

Дата видачі «___» _____ 2024 р.

Керівник (підпис)

Завдання отримав (підпис)

Тема: «Соціальне прослуховування для управління репутацією бренду на прикладі програмного рішення (прототипу), які допомагають компаніям відстежувати та керувати репутацією свого бренду в Інтернеті за допомогою аналізу соціальних мереж і відстеження настроїв»

Календарний план виконання роботи:

№ п/п	Назва етапу дипломного проекту (роботи)	Термін виконання етапу	Примітка
1.	Отримання завдання на дипломну роботу	18.10.2023	
2.	Огляд технічної літератури за темою роботи	30.10.2023	
3.	Провести аналіз предметної області	10.11.2023	
4.	Провести дослідження поняття аналізу тональності тексту, його методик та застосування у великих мовних моделях	14.12.2023	
5.	Провести дослідження можливостей моніторингу соціальних мереж	05.01.2024	
6.	Розробити систему-прототип для соціального прослуховування та керування репутацією бренду	25.03.2024	
7.	Написання пояснювальної записки	29.04.2024	
8.	Створення слайдів для доповіді та написання доповіді	05.05.2024	
9.	Аналіз отриманих результатів з керівником	08.05.2024	
10.	Попередній захист магістерської роботи	17.05.2024	
11.	Корегування роботи за результатами попереднього захисту	20.05.2024	
12.	Остаточне оформлення пояснювальної записки та слайдів	26.05.2024	
13.	Захист магістерської роботи (проекту)	згідно з розкладом ЕК	

Студент **Волошко Максим Володимирович**

Керівник **Афонін Андрій Олександрович**

«_____» _____

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАЧЕНЬ, ТЕРМІНІВ **5**

АНОТАЦІЯ **6**

ВСТУП **7**

РОЗДІЛ 1. УПРАВЛІННЯ РЕПУТАЦІЄЮ БРЕНДУ В СОЦІАЛЬНИХ МЕРЕЖАХ **9**

1.1. Поняття управління репутацією бренду у соціальних мережах 9
1.2. Значення репутації бренду в соціальних мережах 10
1.3. Стратегії управління репутацією бренду в соціальних мережах 11
Висновки до розділу 1 13

РОЗДІЛ 2. АНАЛІЗ ТОНАЛЬНОСТІ ТЕКСТУ **15**

2.1. Поняття та значення аналізу тональності тексту 15
2.2. Методології та підходи до аналізу тональності тексту 16
2.2.1. Підходи на основі правил *16*
2.2.2. Автоматизовані підходи *16*
2.2.3. Гібридні підходи *17*
2.3. Виклики аналізу тональності тексту 18
2.4. Застосування великих мовних моделей для аналізу тональності тексту 19
Висновки до розділу 2 21

РОЗДІЛ 3. АНАЛІЗ МОЖЛИВОСТЕЙ МОНІТОРИНГУ СОЦІАЛЬНИХ МЕРЕЖ **22**

3.1. Поняття та значення моніторингу соціальних мереж 22
3.2. Огляд можливостей API популярних соціальних мереж для моніторингу 24
3.2.1. Meta Graph API *24*
3.2.2. Twitter/X API *27*
3.2.3. TikTok API *29*
Висновки до розділу 3 30

РОЗДІЛ 4. ОПИС РОЗРОБЛЕНОЇ СИСТЕМИ-ПРОТОТИПУ **32**

4.1. Схема даних 32
4.2. Архітектура системи 35
4.3. Hasura 37
4.4. Сервіси-інтеграції Facebook та Instagram 39
4.4.1. Правила синхронізації даних платформ з системою *40*
4.4.2. Механізм роботи з хештегами платформи Instagram *41*
4.5. Сервіси-аналізатори GPT, Claude та Gemini 42
4.6. Механізм роботи черги повідомлень 46
4.6.1. Черга аналізу *46*
4.6.2. Черга затримки *46*
4.6.3. Черга непридатних повідомлень *47*
4.7. Користувацький портал 48
4.7.1. Сторінка процесу адаптації *49*
4.7.2. Вкладка підсумку *50*
4.7.3. Вкладка інформації про пости *52*
4.7.4. Вкладка аналізу *54*
4.7.5. Оповіщення про негативний сентимент *56*
4.8. Висновки до розділу 4 58

ВИСНОВКИ **59**

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ **61**

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАЧЕНЬ, ТЕРМІНІВ

API – application programming interface (англ. прикладний програмний інтерфейс)

SA – sentiment analysis (англ. аналіз тональності)

OM – opinion mining (англ. аналіз думок)

NLP – natural language processing (англ. природна обробка мови)

LLM – large language model (англ. велика мовна модель)

HTTP – hypertext transfer protocol (англ. прокол передачі гіпертексту)

REST – representational state transfer (англ. передача репрезентативного стану)

TTL – time-to-live (англ. час існування)

GPT – generative pre-trained transformer (англ. генеративний попередньо навчений трансформатор)

АНОТАЦІЯ

Ця робота має на меті розробити та оцінити прототип програмного рішення для соціального прослуховування, який допомагає компаніям відстежувати та керувати репутацією свого бренду в інтернеті за допомогою аналізу соціальних мереж і відстеження настроїв. Перша частина зосереджена на аналізі методів та інструментів управління репутацією бренду в соціальних мережах, включаючи сучасні підходи до сентимент-аналізу та можливості моніторингу соціальних мереж на прикладі Facebook, Instagram, Twitter (X) та TikTok. Друга частина розглядає методології аналізу тональності тексту, включаючи автоматизовані та гібридні підходи, та застосування великих мовних моделей для аналізу контенту. Третя частина містить огляд можливостей моніторингу соціальних мереж, зокрема через використання API популярних платформ для збору та аналізу даних. У четвертій частині представлено огляд розробленого прототипу системи для моніторингу соціальних мереж, описуючи його архітектуру, інтеграційні платформи, алгоритми обробки даних та інструменти для здійснення сентимент-аналізу.

Ключові слова: Соціальне прослуховування, Управління репутацією бренду, Сентимент-аналіз, Моніторинг соціальних мереж, Великі мовні моделі

ВСТУП

У сучасному цифровому світі, де соціальні мережі займають важливе місце в житті людей та бізнесу, управління репутацією бренду стало критичним елементом успіху компаній. З поширенням платформ, таких як Facebook, Instagram, Twitter/X та TikTok, споживачі мають можливість висловлювати свої думки та впливати на сприйняття брендів у реальному часі. Це створює як можливості, так і виклики для компаній у підтриманні свого позитивного іміджу та взаємодії зі своєю аудиторією.

Це дослідження має на меті розробку та оцінку прототипу програмного рішення для соціального прослуховування, яке допомагає компаніям відстежувати та керувати репутацією свого бренду в інтернеті за допомогою аналізу соціальних мереж і відстеження настроїв.

Основною метою цього дослідження є створення системи-прототипу для моніторингу соціальних мереж, зокрема Facebook та Instagram, яка включає в себе засоби для здійснення сентимент-аналізу контенту та надання аналітичної інформації для ефективного управління репутацією бренду.

Завдання дослідження включають аналіз існуючих методів та інструментів управління репутацією бренду в соціальних мережах, дослідження методологій аналізу тональності тексту та їх застосування для аналізу контенту в соціальних мережах, оцінку можливостей моніторингу соціальних мереж на прикладі Facebook, Instagram, Twitter (X) та TikTok, розробку та реалізацію прототипу системи для моніторингу соціальних мереж і здійснення сентимент-аналізу, а також проведення експериментальної перевірки та аналізу ефективності розробленої системи.

Дослідження обґрунтовувалося на основі наукових статей та технічних документацій, що надає надійності та перевіреності викладеної інформації.

Наукова новизна цього дослідження полягає у розробці інтегрованої системи для соціального прослуховування та аналізу тональності тексту, яка поєднує сучасні мовні моделі та автоматизовані методи моніторингу соціальних мереж. Розроблений прототип системи дозволяє в реальному часі здійснювати

аналіз настроїв та генерувати аналітичні звіти, що сприяє більш ефективному управлінню репутацією бренду.

Результати цього дослідження можуть бути корисними для маркетологів та спеціалістів з управління репутацією, які прагнуть покращити свої стратегії взаємодії з аудиторією, бізнесів та брендів, які хочуть підвищити свою конкурентоспроможність та ефективно управляти своєю репутацією в інтернеті, а також для дослідників та розробників у галузі аналізу даних та обробки природної мови, які можуть використовувати розроблені методи та інструменти для подальших досліджень. Таким чином, це дослідження спрямоване на створення практичного інструменту, що дозволить компаніям ефективно взаємодіяти зі своєю аудиторією, оперативно реагувати на зміни в настроях та підтримувати позитивний імідж у соціальних мережах.

РОЗДІЛ 1. УПРАВЛІННЯ РЕПУТАЦІЄЮ БРЕНДУ В СОЦІАЛЬНИХ МЕРЕЖАХ

Цей розділ присвячений дослідженню ключових аспектів управління репутацією бренду в соціальних мережах. Перший підрозділ розкриває поняття управління репутацією бренду у соціальних мережах, визначаючи основні методи та інструменти, що використовуються для підтримання позитивного іміджу бренду в онлайн-середовищі. Другий підрозділ аналізує значення репутації бренду у соціальних мережах, підкреслюючи важливість активної та прозорої комунікації з аудиторією. Третій підрозділ зосереджується на стратегіях управління репутацією, охоплюючи такі аспекти, як моніторинг соціальних мереж, оперативна реакція на відгуки споживачів, прозорість у комунікаціях та антикризове управління.

1.1. Поняття управління репутацією бренду у соціальних мережах

Управління репутацією в соціальних мережах є комплексним процесом, який включає систематичний моніторинг та аналіз активності у соціальних мережах з метою підтримання позитивного іміджу бренду в онлайн-середовищі. Цей процес охоплює спостереження за соціальними мережами, а також аналіз відгуків користувачів на різних платформах, таких як Yelp та Google Reviews. Оскільки сприйняття бренду цільовою аудиторією може змінюватися з часом під впливом різних факторів, необхідно забезпечувати постійне та ретельне спостереження за ситуацією для своєчасної реакції на зміни [1].

Процес управління репутацією включає кілька аспектів, таких як створення та публікація релевантного контенту, який відповідає очікуванням та потребам аудиторії. Реагування на відгуки та коментарі користувачів є частиною цього, оскільки своєчасна та адекватна відповідь на негативні відгуки може впливати на імідж компанії в очах споживачів. Крім того, управління репутацією передбачає проведення регулярного аудиту ефективності комунікаційних стратегій для

виявлення сильних та слабких сторін, а також коригування підходів до взаємодії з аудиторією.

Використання даних соціального прослуховування є інструментом у процесі управління репутацією. Завдяки таким даним можна отримати інформацію про ставлення аудиторії до бренду, виявити основні тенденції та настрої, а також вчасно реагувати на потенційні кризи. Таким чином, управління репутацією в соціальних мережах спрямоване на формування та підтримку довготривалих відносин з цільовою аудиторією, що сприяє зміцненню позицій бренду на ринку та підвищенню його конкурентоспроможності.

1.2. Значення репутації бренду в соціальних мережах

«Комунікація за допомогою соціальних медіа принципово відрізняється від комунікації за допомогою традиційних медіа, коли компанія могла направити добре підготовлене і сплановане повідомлення групі ізольованих отримувачів. Соціальні медіа створюють нове середовище комунікацій, де недостатньо лише говорити, тепер потрібно розмовляти; недостатньо просто комунікувати, потрібно слухати; недостатньо лише показувати, потрібно залучатися, а також піклуватися про кожного окремо (замість того, щоб говорити до всіх одразу) і бути перевіреним 24/7» [2, с. 164].

Управління репутацією бренду у соціальних мережах є важливим аспектом корпоративного управління, оскільки соціальні медіа створюють нове середовище для комунікації, де кожен користувач може поширювати свою думку про компанію та її продукцію. Це середовище створює як можливості, так і загрози для репутації компаній. На одному боці, соціальні медіа дозволяють будувати інтерактивні відносини з необмеженою кількістю отримувачів, включаючи клієнтів, співробітників, інвесторів та бізнес-партнерів. З іншого боку, вони можуть призвести до значних загроз для репутації, оскільки негативний коментар або відгук може швидко поширитися, завдавши шкоди іміджу компанії.

Дослідження показують, що польські компанії починають визнавати потенціал соціальних медіа та намагаються застосовувати їх у своїх

маркетингових зусиллях. Проте, їх використання часто обмежується традиційним підходом, де соціальні медіа розглядаються як ще один промоційний інструмент, використовуваний департаментами маркетингу та зв’язків з громадськістю. Цей підхід не є інтегрованим і не дозволяє ефективно будувати та захищати репутацію в інтернет-середовищі [2, с. 161].

Використання соціальних медіа стає необхідністю та можливістю для зростання компаній, дозволяючи встановлювати тісні відносини з різними групами зацікавлених сторін, створюючи прозорість та відкритий діалог, що сприяє підвищенню довіри з боку цих груп. Довіра до інтернет-медіа також зростає. Згідно з дослідженням Edelman Trust Barometer, рівень довіри до пошукових систем зріс з 62% у 2012 році до 64% у 2015 році, а до соціальних медіа – з 45% у 2012 році до 48% у 2015 році. Водночас, рівень довіри до традиційних медіа знижується: з 67% у 2012 році до 62% у 2015 році [2, с. 165].

Соціальні мережі виступають сучасним та дієвим інструментом для взаємодії з різними групами зацікавлених сторін, що допомагає формувати позитивний імідж і зміцнювати репутацію компанії. Вони також надають можливість оперативно реагувати на негативні відгуки та кризові ситуації, що виникають в онлайн-просторі. Для цього необхідна постійна активність у соціальних мережах, їх безперервний моніторинг і впровадження механізмів швидкого реагування. Згідно досліджень, 94% споживачів відмовляються від вибору певного бізнесу через негативний відгук, 53% очікують отримати відповідь на негативний відгук не пізніше ніж за три доби, а 63% опитуваних стверджують про досвід того, що хоча б одна компанія навіть не відповіла на негативний відгук [5]

1.3. Стратегії управління репутацією бренду в соціальних мережах

Стратегії управління репутацією бренду в соціальних мережах є важливими для забезпечення конкурентоспроможності, мінімізації ризиків, підвищення залученості та лояльності клієнтів, формування позитивного іміджу та збору зворотного зв'язку.

Моніторинг соціальних мереж є фундаментальним кроком у управлінні репутацією бренду. Цей процес включає відстеження згадок про бренд, аналіз тональності повідомлень (позитивна, негативна, нейтральна) та визначення основних тем і трендів, що пов'язані з брендом. Автоматизація цього процесу може включати в себе розробку систем, які використовують алгоритми машинного навчання та обробки природної мови для автоматичного збирання та аналізу даних із соціальних мереж. Це забезпечує конкурентоспроможність, оскільки компанія може оперативно реагувати на зміни на ринку та задовольняти потреби споживачів.

Оперативна і професійна реакція на відгуки споживачів є важливим аспектом управління репутацією. Це включає відповідь на коментарі та запити користувачів у соціальних мережах, вирішення скарг та висловлення вдячності за позитивні відгуки. Споживачі цінують увагу до їхніх потреб і проблем, тому ввічливість і професіоналізм у спілкуванні зміцнюють довіру до бренду. Використання автоматизованих систем для відстеження відгуків дозволяє швидко і точно відповідати на коментарі, що підвищує залученість та лояльність клієнтів.

Прозорість у комунікаціях зі споживачами та відповідальність за свої дії щодо важливих подій, змін або помилок, які можуть вплинути на споживачів, сприяє формуванню довіри. Визнання помилок та пропозиція щодо їх вирішення демонструють відповідальність бренду перед своїми клієнтами. Це мінімізує ризики та ефективно управляє кризовими ситуаціями, зменшуючи негативний вплив на репутацію.

Якісний контент є основою позитивного іміджу бренду в соціальних мережах. Це включає публікацію корисних та цікавих матеріалів, які відповідають інтересам цільової аудиторії. Використання різних форм контенту, таких як статті, відео, інфографіки та опитування, залучає і утримує увагу споживачів. Автоматизовані системи можуть допомагати визначати найбільш ефективні типи контенту та оптимізувати стратегії його створення.

Антикризове управління передбачає розробку стратегічних планів і процедур для оперативного реагування на кризові ситуації, що можуть негативно позначитися на репутації бренду. Це включає підготовку прес-релізів, створення внутрішніх інструкцій для персоналу та використання соціальних мереж для комунікації зі споживачами під час кризи. Швидка та прозора реакція зменшує негативні наслідки і відновлює довіру до бренду, що є критично важливим для мінімізації ризиків. Яскравим прикладом успішного виконання антикризового плану є випадок 2018 року, коли мережа ресторанів KFC у Великобританії зіткнулася з серйозною кризою через перебої в постачанні, що призвели до нестачі основних інгредієнтів і тимчасового закриття 900 ресторанів. Така ситуація викликала значний резонанс серед споживачів, які активно висловлювали своє незадоволення в соціальних мережах. KFC обрала креативний підхід до вирішення цієї кризи. Вони запустили рекламну кампанію, в якій використали змінену абревіатуру свого бренду на «FCK», що привернуло увагу і викликало позитивні емоції у споживачів. У результаті, показник сприйняття бренду який впав до -12 під час кризи, зріс до -1, що на один пункт краще, ніж до кризи [6].

Висновки до розділу 1

У цьому розділі розглянуто ключові аспекти управління репутацією бренду в соціальних мережах, що є важливими для сучасних компаній.

Управління репутацією включає систематичний моніторинг та аналіз активності у соціальних мережах, своєчасну реакцію на відгуки користувачів та публікацію релевантного контенту. Соціальні медіа створюють нові можливості для комунікації з аудиторією, але також несуть ризики швидкого поширення негативної інформації.

Стратегії управління репутацією охоплюють моніторинг згадок про бренд, аналіз тональності повідомлень, оперативну реакцію на відгуки, прозорість у комунікаціях та антикризове управління. Використання сучасних технологій дозволяє автоматизувати ці процеси, підвищуючи їх ефективність.

Таким чином, ефективне управління репутацією бренду в соціальних мережах сприяє підтриманню позитивного іміджу, зміцненню довіри споживачів та забезпеченню конкурентоспроможності компаній.

РОЗДІЛ 2. АНАЛІЗ ТОНАЛЬНОСТІ ТЕКСТУ

Цей розділ присвячений аналізу тональності тексту та його значенню в контексті управління репутацією бренду. Перший підрозділ висвітлює основні поняття та значення аналізу тональності тексту, зокрема, його здатність визначати емоційне забарвлення тексту та можливість масштабованої обробки даних. Другий підрозділ охоплює різні методології та підходи до аналізу тональності тексту, включаючи підходи на основі правил, методи машинного навчання та гібридні моделі, підкреслюючи їхні переваги та недоліки. Третій підрозділ розглядає основні виклики аналізу тональності тексту, такі як суб'єктивність, контекстуальна залежність, іронія та обробка емодзі, що впливають на точність аналізу. Четвертий підрозділ зосереджується на застосуванні великих мовних моделей для аналізу тональності тексту, демонструючи їх високу точність та адаптивність до різних форматів тексту, а також їхню здатність покращувати процес прийняття рішень та управління репутацією бренду.

2.1. Поняття та значення аналізу тональності тексту

Аналіз тональності або аналіз думок відноситься до комп'ютерного дослідження людської поведінки, перспектив та почуттів щодо об'єкта, яким може бути людина, подія або тема. Підхід обробки природної мови, відомий як аналіз тональності, намагається вилучати з текстів думки та ставлення. Основною метою SA є визначення, чи відображає конкретний текст позитивне або негативне ставлення; таким чином, це завдання можна розглядати як проблему класифікації тексту [7, с. 1].

Аналіз тональності надає можливість масштабованості обробки даних, оскільки він дозволяє ефективно обробляти великі обсяги неструктурованих даних, що було б неможливо зробити вручну. Також сентимент аналіз ідентифікує проблеми в реальному часі, наприклад, зростання кризової ситуації

в соціальних мережах або ризик втрати клієнта, дозволяючи компаніям негайно вжити заходів. Згідно з дослідженнями, оцінка сентименту від різних людей сходиться лише близько у 60-65% [8]. Це надає аналізу сентименту перевагу в тому, що він є детермінованим та позбавленим суб’єктивності, що створює можливість більш точних даних.

2.2. Методології та підходи до аналізу тональності тексту

Аналіз тональності тексту є складною задачею, яка потребує використання різноманітних методологій та підходів. Існує кілька основних методів, що застосовуються для визначення емоційного забарвлення тексту. Серед них — підходи на основі правил, методи машинного навчання та гібридні моделі. Кожен з цих методів має свої переваги та недоліки, які залежать від конкретного застосування та типу даних. Обрання найефективнішої стратегії для аналізу тональності в різних контекстах потребує ретельного аналізу та порівняння цих підходів.

2.2.1. Підходи на основі правил

Підходи на основі правил зазвичай використовують набір створених людиною правил для ідентифікації суб'єктивності, полярності або теми думки. Ці правила можуть включати різні техніки NLP, такі як стемінг, токенізація, надання тегів для різних частин мови та парсинг. Крім того, застосовуються лексикони (тобто списки слів та виразів). Наприклад, система на основі правил може визначити два списки поляризованих слів (негативні та позитивні), підраховувати кількість поява цих слів у тексті та визначати загальний настрій тексту. Однак, такі системи не враховують, як слова комбінуються в реченні, що може призвести до помилок. Більш складні методи обробки та додавання нових правил можуть поліпшити точність, але також ускладнюють систему і вимагають регулярного налаштування та підтримки.

2.2.2. Автоматизовані підходи

Автоматизовані підходи до аналізу тональності тексту, на відміну від систем, що базуються на правилах, використовують методи машинного навчання.

Завдання аналізу тональності часто моделюється як проблема класифікації, де класифікатор отримує текст та повертає категорію, наприклад, позитивну, негативну або нейтральну.

Процес навчання і прогнозування можна описати наступним чином. Під час процесу навчання модель асоціює певний вхідний текст із відповідним тегом на основі навчальних прикладів. Екстрактор ознак перетворює текст на вектор ознак. Пара векторів ознак і тегів (наприклад, позитивний, негативний або нейтральний) подається в алгоритм машинного навчання для створення моделі.

На етапі прогнозування екстрактор ознак використовується для перетворення нових текстових даних на вектори ознак. Ці вектори подаються в модель, яка генерує прогнозовані теги.

Першим кроком у машинному навчанні для текстової класифікації є витягання ознак з тексту. Класичним підходом є використання «мішку слів» або «мішку n-грамів» з їх частотою. Останнім часом застосовуються нові техніки екстракції ознак, базовані на векторних уявленнях слів. Ці уявлення дозволяють словам з подібним значенням мати схожі векторні представлення, що може покращити продуктивність класифікаторів.

На етапі класифікації зазвичай використовуються статистичні моделі, такі як наївний Баєс, логістична регресія, метод опорних векторів та нейронні мережі. Наївний Баєс використовує теорему Баєса для прогнозування категорії тексту. Логістична регресія прогнозує значення на основі набору ознак. Метод опорних векторів використовує представлення текстових прикладів як точок у багатовимірному просторі, де нові тексти класифікуються на основі подібності з існуючими текстами. Глибоке навчання використовує штучні нейронні мережі для обробки даних, намагаючись імітувати роботу людського мозку.

2.2.3. Гібридні підходи

Гібридні підходи поєднують у собі бажані елементи методів, що базуються на правилах, та автоматичних технік, утворюючи одну систему. Це дозволяє використовувати сильні сторони обох підходів: точність та конкретність правил разом з гнучкістю та адаптивністю машинного навчання. Основною перевагою

таких систем є підвищена точність результатів. Гібридні системи можуть адаптуватися до нових даних, зберігаючи при цьому точність на основі попередньо встановлених правил. Це забезпечує більш ефективне та надійне рішення для аналізу тональності тексту, що враховує як загальні тенденції, так і специфічні особливості аналізованих даних.

Гібридні підходи застосовуються в таких випадках, коли потрібно забезпечити високу точність аналізу та врахувати контекстуальні фактори, які можуть впливати на результати. Вони можуть використовуватися для аналізу великих обсягів даних у реальному часі, забезпечуючи при цьому стабільність і точність прогнозів. Завдяки інтеграції методів на основі правил і машинного навчання, гібридні системи є більш гнучкими і можуть краще справлятися зі складними завданнями аналізу тональності.

2.3. Виклики аналізу тональності тексту

Аналіз тональності є однією з найскладніших задач у галузі обробки природної мови, оскільки навіть людям важко точно визначити емоційне забарвлення тексту. Незважаючи на покращення точності класифікаторів, залишається багато невирішених питань.

Однією з ключових проблем є суб'єктивність текстів. Суб'єктивні тексти містять явні емоції, тоді як об'єктивні їх не мають. Наприклад, фраза «цей телефон чудовий» має позитивну тональність, тоді як «цей телефон чорний» — нейтральну.

Контекст і полярність також є важливими аспектами. Тексти можуть змінювати свою тональність залежно від контексту, в якому вони були написані. Машини не завжди можуть зрозуміти контекст, якщо він не зазначений явно, що ускладнює точний аналіз.

Іронія та сарказм створюють значні труднощі, оскільки люди часто використовують позитивні слова для вираження негативних емоцій. Без розуміння контексту такі вирази важко розпізнати.

Порівняння можуть ускладнювати аналіз тональності. Наприклад, фрази «цей продукт кращий за інші» або «це краще, ніж нічого» потребують додаткового контексту для точної класифікації.

Емодзі відіграють важливу роль у вираженні емоцій, особливо у твітах. Вони можуть бути західними (:D) або східними (¯_(ツ)_/¯). Правильна обробка емодзі на рівні символів і слів є важливим аспектом попередньої обробки даних для аналізу тональності.

Визначення нейтральної тональності також є важливим завданням. Об'єктивні тексти та неактуальна інформація зазвичай відносяться до нейтральних, але важливо ретельно визначати категорії, щоб не додати шуму до класифікатора.

Точність людських анотацій також викликає труднощі, оскільки навіть люди не завжди погоджуються щодо емоційної забарвленості тексту. Машини навчаються на даних, а якщо ці дані позначаються непослідовно, точність класифікаторів знижується. Проте аналіз тональності залишається цінним інструментом, оскільки навіть з певними похибками він дозволяє значно економити час та ресурси при аналізі великих обсягів текстових даних.

2.4. Застосування великих мовних моделей для аналізу тональності тексту

Великі мовні моделі показали виняткові можливості в аналізі тональності з високою точністю, навіть без спеціального навчання для конкретних завдань, завдяки методу навчання з нуля. Наприклад, GPT-4 досяг середньої точності 93% у завданнях бінарної класифікації, перевершивши багато традиційних моделей трансферного навчання [9, с. 3]. Цей рівень продуктивності свідчить про те, що компанії можуть використовувати LLM для аналізу тональності з мінімальною настройкою, заощаджуючи час і ресурси, зазвичай витрачені на навчання спеціальних моделей.

Однією з важливих переваг LLM є їх адаптивність до різних форматів тексту та складності. Дослідження показало, що документи з більш довгими

словами та декількома реченнями значно покращують продуктивність класифікації у всіх протестованих LLM [9, с. 15]. Це підкреслює те, що LLM особливо підходять для аналізу детальних оглядів продуктів, відгуків клієнтів та інших контекстуально багатих текстів. Проте, їх продуктивність може знижуватися при аналізі коротких, неструктурованих текстів, таких як твіти, які часто містять розмовну та двозначну мову [9, с. 15].

Дослідження підкреслює важливість інженерії підказок для покращення продуктивності LLM. Контекстуальні підказки, які надають більш детальну фонову інформацію, та навчання з кількома прикладам, що включає невелику кількість прикладів у підказку, показали покращення точності класифікацій тональності [9, с. 4]. Це свідчить про те, що компаніям варто інвестувати в розробку ефективних стратегій підказок, щоб максимізувати точність аналізу тональності при використанні LLM.

Пояснюваність є критичним фактором для впровадження штучного інтелекту в реальних застосуваннях. LLM показали здатність надавати детальні та зрозумілі пояснення своїх класифікацій тональності, що підвищує прозорість і довіру [9, с. 14]. Ця здатність є важливою для застосувань у чутливих сферах, таких як охорона здоров'я, фінанси та обслуговування клієнтів, де розуміння причин, що лежать в основі рішень, прийнятих штучним інтелектом, є необхідним для побудови довіри користувачів.

Універсальність та висока продуктивність LLM роблять їх придатними для широкого спектру практичних застосувань. Компанії можуть використовувати LLM для аналізу великих обсягів відгуків клієнтів, визначення ключових настроїв і трендів для інформування про розробку продуктів та маркетингові стратегії. Також вони можуть бути використані для моніторингу настроїв бренду на платформах соціальних мереж, надаючи реальні дані щодо сприйняття публікою та дозволяючи проактивне управління репутацією. Аналізуючи думки споживачів, висловлені в різних онлайн-форумах та опитуваннях, LLM можуть допомогти компаніям отримати глибше розуміння про ринкові тенденції та переваги клієнтів.

Висновки до розділу 2

Аналіз тональності тексту є ключовим інструментом у сучасному світі, де обсяг інформації постійно зростає. Він дозволяє автоматизувати обробку великих обсягів неструктурованих даних, надаючи можливість компаніям своєчасно реагувати на зміни настроїв та кризові ситуації. Використання різних методологій, таких як підходи на основі правил, методи машинного навчання та гібридні моделі, дозволяє вибирати найбільш відповідний підхід для конкретного застосування, що підвищує точність і ефективність аналізу. Незважаючи на численні виклики, включаючи суб'єктивність текстів, контекстуальну залежність, іронію, сарказм та обробку емодзі, аналіз тональності залишається важливим інструментом для бізнесу та наукових досліджень.

Сучасні великі мовні моделі, такі як GPT-4, демонструють значні переваги у порівнянні з традиційними методами. Вони показують високу точність, адаптивність до різних текстових форматів і здатність надавати зрозумілі пояснення своїх класифікацій, що підвищує прозорість та довіру до результатів. Ці моделі особливо корисні для аналізу великих обсягів даних, таких як детальні огляди продуктів, відгуки клієнтів та інші контекстуально багаті тексти. Вони допомагають компаніям отримати глибше розуміння ринкових тенденцій, переваг клієнтів та ефективно керувати репутацією бренду на платформах соціальних мереж.

РОЗДІЛ 3. АНАЛІЗ МОЖЛИВОСТЕЙ МОНІТОРИНГУ СОЦІАЛЬНИХ МЕРЕЖ

Цей розділ присвячений аналізу можливостей моніторингу соціальних мереж, що є важливим елементом управління репутацією бренду. У цьому розділі розглядаються концепції, значення та інструменти, які використовуються для ефективного відстеження та аналізу активності в соціальних медіа.

У першому підрозділі розкривається поняття та значення моніторингу соціальних мереж, а також обговорюються його ключові переваги та стратегічні можливості для компаній.

Другий підрозділ містить огляд можливостей API популярних соціальних мереж для моніторингу:

1. Детально розглянуто Meta Graph API, що використовується для доступу до даних Facebook та Instagram, з описом основних сутностей та функціональних можливостей.
2. Аналізується Twitter/X API, включаючи структуру, методи аутентифікації та обмеження безкоштовного доступу.
3. Досліджуються можливості TikTok API, з акцентом на обмеженнях, пов'язаних з отриманням коментарів та органічного контенту.

3.1. Поняття та значення моніторингу соціальних мереж

Моніторинг соціальних мереж визначається як процес систематичного відстеження, збору та аналізу інформації з соціальних медіа платформ для виявлення та оцінки згадок, настроїв, тенденцій і взаємодій, що стосуються певного бренду, компанії, продукту чи теми. Цей процес включає використання спеціалізованих програмних засобів та методик для автоматизації збору даних, їх подальшої обробки та аналізу з метою отримання актуальної інформації, необхідної для прийняття управлінських рішень та розробки стратегії комунікації.

Моніторинг соціальних мереж є важливим стратегічним інструментом для оцінки та покращення ефективності комунікацій. Він дозволяє відслідковувати, наскільки медіа-активності привертають увагу, оцінювати охоплення цільової аудиторії та аналізувати їхню реакцію на продукти чи послуги компанії. Дані, отримані через моніторинг соціальних мереж, слугують надійним джерелом для прийняття управлінських рішень. Зручність інструментів моніторингу полягає у можливості автоматичного збору та аналізу даних, складання звітів та надсилання їх на електронну пошту у визначені терміни. Це економить час та дозволяє зосередитися на розробці стратегії і плануванні майбутніх дій.

Соціальні медіа інструменти, підтримані моніторингом, допомагають удосконалювати поточні маркетингові стратегії. Вони дозволяють визначати потреби цільової аудиторії та задовольняти їх, отримуючи зворотній зв'язок без необхідності прямого опитування. Використовуючи ці інструменти, можна отримати цінні відгуки від аудиторії, персоналізувати комунікацію, зменшити небажаний шум у соціальних мережах і підвищити ефективність маркетингових зусиль.

Моніторинг соціальних мереж також допомагає знаходити впливових осіб, які можуть розширити охоплення аудиторії бренду. Це дозволяє налагодити з ними співпрацю, підвищуючи рівень довіри та конверсії, оскільки їхня аудиторія схильна довіряти рекомендаціям авторитетних особистостей.

Моніторинг соціальних мереж сприяє генерації нових ідей для контенту, дозволяючи відстежувати актуальні теми та тренди у галузі. Це допомагає знаходити інноваційні ідеї для контенту, використовуючи досвід конкурентів як джерело натхнення або як приклад того, що не працює на ринку.

Інструменти моніторингу соціальних мереж відіграють важливу роль у управлінні репутацією бренду та ефективному реагуванні на кризові ситуації. Вони дозволяють відстежувати і миттєво реагувати на згадки, забезпечуючи швидку відповідь на відгуки клієнтів і показуючи, що компанія дбає про їхні потреби та досвід.

3.2. Огляд можливостей API популярних соціальних мереж для моніторингу

API популярних соціальних мереж надають засоби для інтеграції та взаємодії з платформами, дозволяючи автоматично збирати та аналізувати дані. Використання офіційних API є рекомендованим підходом, оскільки він забезпечує стабільність, безпеку та дотримання політик платформ. Застосування веб-скрепінгу, навпаки, може порушувати правила користування платформами, призводити до блокування акаунтів та не гарантує точності й повноти зібраних даних.

3.2.1. Meta Graph API

Graph API є основним механізмом для отримання та передачі даних у межах платформи Facebook. Це API, засноване на протоколі HTTP, дозволяє додаткам програмно запитувати дані, публікувати новий контент, керувати рекламними оголошеннями, завантажувати зображення та виконувати безліч інших операцій.

Назва Graph API походить від концепції «соціального графа», яка представляє структуровану інформацію на Facebook. Воно складається з вузлів, ребер і полів. Вузли використовуються для отримання даних про конкретні об'єкти; ребра полегшують отримання колекцій об'єктів, пов'язаних з однією сутністю; а поля надають детальну інформацію про окремі об'єкти або елементи у колекції. Усі обміни даними відповідають стандарту HTTP/1.1 і вимагають використання HTTPS [10]. Сутність Graph API, заснована на HTTP, забезпечує його сумісність з будь-якою мовою програмування, яка має бібліотеку HTTP, такою як cURL або urllib, що дозволяє безпосередньо взаємодіяти з API через веб-браузери.

Токени доступу є важливими для доступу до Graph API. Більшість кінцевих точок вимагають певної форми токена доступу для аутентифікації запитів. Ці токени виконують дві основні функції:

1. Вони дозволяють додатку отримувати доступ до інформації користувача без необхідності введення пароля користувача. Наприклад, якщо додатку потрібна електронна адреса користувача, користувач може авторизувати додаток для отримання електронної адреси з Facebook без введення пароля Facebook [10].
2. Вони дозволяють ідентифікувати додаток, користувача, що використовує додаток, і конкретні дані, до яких користувач надав додатку доступ [10].

Вузол представляє окремий об'єкт із унікальним ідентифікатором, таким як користувачі, сторінки, групи, пости, фотографії та коментарі, кожен із яких є окремим вузлом у соціальному графі Facebook.

Ребро позначає взаємозв'язок між двома вузлами. Наприклад, вузол користувача може бути пов'язаний із кількома фотографіями, а вузол фотографії може мати кілька коментарів, пов'язаних із ним, що ілюструє взаємопов'язану структуру соціального графа.

Facebook Graph API надає широкий набір сутностей для доступу до даних і аналізу активності в соціальних мережах. Кожна сутність відображає різні аспекти взаємодії користувачів, контенту та залученості на платформі.

1. Сутність User представляє користувача Facebook та забезпечує доступ до таких даних, як інформація профілю, публікації, друзі та вподобання (сторінки, публікації та інший контент, що сподобався користувачу).
2. Сутність Page представляє сторінку на Facebook, яку використовують бізнеси, бренди, знаменитості та організації. Вона включає базову інформацію (назва сторінки, категорія, опис, фото профілю та обкладинки), публікації (текстовий контент, зображення, відео, посилання, а також метрики залученості, такі як вподобання, коментарі та поширення), аналітику (детальні показники ефективності сторінки, такі як перегляди сторінки, вподобання, охоплення публікацій, залученість) та інформацію про підписників (демографічні дані аудиторії сторінки).

3. Сутність Post представляє окрему публікацію на Facebook, що включає контент (текст, зображення, відео, посилання), коментарі (коментарі користувачів до публікації, що надають багате джерело користувацького контенту та настроїв), реакції (різні типи реакцій, що вказують на настрої користувачів та їхню залученість) та поширення (кількість поширень публікації, що свідчить про її охоплення та вірусний потенціал).
4. Сутність Comment представляє коментар користувача до публікації, включаючи текст коментаря, відповіді (відповіді на коментар, що надають додатковий контекст і настрої) та вподобання (кількість вподобань коментаря, що вказує на його популярність і вплив).
5. Сутність Reaction представляє реакцію користувача на публікацію або коментар, включаючи тип реакції та кількість кожного типу реакції, що надає кількісні дані про залученість користувачів.
6. Сутність Insights надає детальну аналітику та метрики для сторінок, публікацій і відео. Вона включає показники сторінки (перегляди сторінки, вподобання сторінки, охоплення публікацій, залученість, демографічні дані аудиторії), показники публікацій (охоплення, покази, залученість, взаємодії) та показники відео (перегляди, середній час перегляду, утримання аудиторії).
7. Сутність Event представляє подію, створену на Facebook, включаючи інформацію про подію (назва, опис, місце проведення, час початку та завершення) та учасників (список користувачів, які зацікавлені або відвідують подію, що надає уявлення про популярність та охоплення події).
8. Сутність Group представляє групу на Facebook, включаючи базову інформацію, учасників та публікації (які можна аналізувати для оцінки залученості та настроїв).
9. Сутність Ad представляє рекламне оголошення на Facebook, включаючи інформацію про рекламу (креативний контент, критерії таргетингу,

бюджет) та показники ефективності (дані про покази, кліки, конверсії та інші ключові показники ефективності).

10. Сутність Hashtag представляє хештег, що використовується в публікаціях на Facebook, дозволяючи відстежувати публікації, що містять певний хештег, для моніторингу специфічних тем і трендів.

Instagram Graph API є підмножиною Facebook Graph API, використовує аналогічний інтерфейс для доступу до даних, має певні однакові кінцеві точки для певних вузлів, як і на Facebook Graph API, але також має свої унікальні додаткові вузли.

3.2.2. Twitter/X API

API Twitter/X є основним інструментом для програмної взаємодії з платформою Twitter/X. Він дозволяє розробникам отримувати доступ, керувати та публікувати дані на Twitter/X, полегшуючи виконання широкого спектру операцій, від запитів твітів та інформації про користувачів до керування рекламою на Twitter/X та автоматизації дій акаунта.

API структуроване відповідно до принципів RESTful і підтримує HTTP-методи, такі як GET, POST і DELETE. Ця структура забезпечує безперешкодну інтеграцію з різними мовами програмування, які підтримують бібліотеки HTTP, такими як cURL та requests. Кожна кінцева точка в API Twitter/X відповідає певному набору операцій, і вся комунікація вимагає використання HTTPS для забезпечення безпечної передачі даних.

API Twitter/X побудоване за ієрархічною структурою ресурсів і підресурсів. Основні ресурси включають твіти, користувачів, медіа та прямі повідомлення, кожен з яких має свій набір атрибутів і зв'язків з іншими ресурсами. Наприклад, ресурс твіт може включати дані про автора, зміст, метрики залучення (такі як лайки та ретвіти) і зв'язки з пов'язаними твітами через відповіді або треди.

Аутентифікація є критичним компонентом API Twitter/X і підтримує кілька методів аутентифікації, включаючи OAuth 1.0a User Context, OAuth 2.0 Bearer Token та OAuth 2.0 User Context. Ці методи забезпечують безпечний доступ до

API, дозволяючи додаткам виконувати дії від імені користувачів без необхідності введення їх паролів. Токени, згенеровані за допомогою цих методів аутентифікації, надають необхідні дозволи для взаємодії з конкретними кінцевими точками, забезпечуючи роботу додатку в межах наданого користувачем доступу [11].

API також накладає обмеження на частоту запитів для підтримання стабільності та продуктивності платформи Twitter. Ці обмеження визначають кількість запитів, які можна виконати до конкретних кінцевих точок протягом певного часу, заохочуючи ефективне та відповідальне використання API. Розробники можуть відслідковувати свої обмеження на частоту запитів та відповідно коригувати поведінку своїх додатків, щоб залишатися в межах цих обмежень.

Кінцеві точки API Twitter/X організовані в різні категорії, такі як Твіти, Користувачі, Простори та Медіа. Кожна категорія охоплює ряд кінцевих точок, призначених для обробки конкретних функціональних можливостей. Наприклад, категорія Твіти включає кінцеві точки для публікації твітів, отримання деталей твітів та пошуку твітів на основі таких критеріїв, як хештеги або згадки користувачів.

API Twitter/X пропонує декілька рівнів доступу, кожен з яких має свої унікальні можливості та обмеження. Безкоштовний рівень, хоча і надає початкову можливість для базових взаємодій з платформою, має суттєві обмеження, які унеможливлюють повноцінну реалізацію систем моніторингу соціальних мереж.

Безкоштовний рівень накладає обмеження у 1 500 одиниць запитів для створення публікацій, і, що важливо, не підтримує функції отримання даних [11]. Відсутність можливості отримання контенту є критичним обмеженням, оскільки соціальний моніторинг значною мірою залежить від можливості отримувати, аналізувати та відповідати на великий обсяг публікацій у соціальних мережах.

Вищі рівні доступу надають значно більші можливості. Основний рівень дозволяє створювати 3 000 публікацій і отримувати 10 000 одиниць даних, а рівні Pro та Enterprise пропонують експоненціально більші ліміти, що дозволяють

здійснювати широкомасштабну обробку та аналіз даних. Однак, плани Basic і Pro коштують відповідно $100 і $5000 на місяць [11].

3.2.3. TikTok API

API TikTok дозволяє отримувати доступ, керувати та публікувати контент на TikTok, забезпечуючи широкий спектр функціональних можливостей, від запитів даних про відео та інформації про користувачів до керування взаємодіями користувачів та автоматизації завантаження контенту.

API TikTok структуроване відповідно до принципів RESTful та підтримує стандартні HTTP-методи, такі як GET, POST і DELETE, що забезпечує сумісність з різними мовами програмування, які підтримують бібліотеки HTTP, такими як cURL та requests. Кожна кінцева точка в API TikTok призначена для виконання певних завдань, і вся комунікація захищена за допомогою HTTPS для збереження цілісності та конфіденційності даних.

Ключові ресурси API TikTok включають відео, користувачів, коментарі та хештеги, кожен з яких має свої атрибути та взаємозв'язки. Наприклад, ресурс відео може містити дані про його творця, опис, метрики залучення (такі як лайки, коментарі та поширення) і посилання на пов'язаний контент через хештеги або згадки користувачів.

Аутентифікація є важливим аспектом API TikTok, який використовує OAuth 2.0 як основний метод. Цей протокол дозволяє додаткам діяти від імені користувачів без необхідності введення їх паролів, тим самим забезпечуючи безпечний доступ до API. Токени доступу, згенеровані за допомогою OAuth 2.0, надають необхідні дозволи для взаємодії з конкретними кінцевими точками, забезпечуючи роботу додатків у межах наданих користувачем дозволів.

API накладає обмеження на частоту запитів для підтримання стабільності та продуктивності платформи. Ці обмеження визначають кількість запитів, які можна виконати до різних кінцевих точок протягом певного періоду.

Кінцеві точки API TikTok організовані в кілька категорій, таких як Відео, Користувач, Коментар і Хештег. Кожна категорія охоплює кілька кінцевих точок, призначених для обробки конкретних функціональних можливостей. Наприклад,

категорія Відео включає кінцеві точки для завантаження відео, отримання деталей відео та пошуку відео на основі таких критеріїв, як хештеги або ідентифікатори користувачів.

Однією із основних проблем у роботі з TikTok API є відсутність можливості витягання коментарів та їх контенту з відео. TikTok надає різні категорії API, зокрема Research API і TikTok for Business API.

Research API надає широкі можливості для аналізу та моніторингу соціальної мережі, однак доступ до нього може бути отриманим лише через офіційне підтвердження від компанії, а також вони забороняють використання цього API у комерційних цілях. Оскільки система-прототип розроблялася з наміром її використання у комерційних цілях, доступ до Research API не було отримано [12].

Реалізація повноцінного моніторингу соціальних мереж для TikTok за допомогою TikTok for Business API представляє значні виклики через кілька обмежень, зокрема відсутність кінцевих точок для отримання коментарів. TikTok for Business API не надає прямого методу для отримання коментарів до відео. Коментарі є важливим аспектом моніторингу соціальних мереж, оскільки вони надають інсайти щодо настроїв аудиторії, зворотного зв'язку та залучення. Без доступу до цих даних стає складно зрозуміти, як користувачі взаємодіють з контентом, що вони говорять і як вони ставляться до бренду або кампанії.

Фокус TikTok for Business API головним чином зосереджений на даних реклами та маркетингу. Це означає, що органічний (не рекламний) контент і взаємодії користувачів з таким контентом не повністю доступні. Моніторинг органічного контенту є життєво важливим для всебічного розуміння присутності бренду та поведінки аудиторії на платформі.

Висновки до розділу 3

Аналіз можливостей моніторингу соціальних мереж показав, що сучасні інструменти та API надають широкі можливості для відстеження та аналізу контенту у соціальних мережах.

Facebook та Instagram через Meta Graph API дозволяють отримувати детальну інформацію про користувачів, сторінки, публікації та взаємодії, що забезпечує комплексний аналіз активності на цих платформах.

Twitter/X API також пропонує багатий функціонал для отримання та публікації даних, хоча безкоштовний рівень робить неможливим виконання повноцінного моніторингу.

TikTok API, незважаючи на свої можливості, має обмежений доступ до важливих даних, таких як коментарі, що створює виклики для моніторингу соціальних мереж на цій платформі. Доступні інструменти TikTok зосереджені переважно на рекламних даних, що обмежує можливості всебічного аналізу органічного контенту та взаємодій.

Загалом, інтеграція різних API та використання спеціалізованих програмних засобів дозволяє автоматизувати процеси збору, обробки та аналізу даних, забезпечуючи актуальну інформацію для прийняття стратегічних рішень, однак для ефективного використання необхідно враховувати обмеження та специфіку кожної платформи.

РОЗДІЛ 4. ОПИС РОЗРОБЛЕНОЇ СИСТЕМИ-ПРОТОТИПУ

У цьому розділі буде детально розглянуто розроблену систему-прототип для аналізу та управління соціальними медіа. Кожен підрозділ розкриває ключові аспекти системи, її архітектуру, функціональні компоненти та механізми взаємодії. Перший підрозділ охоплює структуру даних, що використовується в системі, пояснюючи, як дані організовані та зберігаються. Другий підрозділ надає огляд архітектурного дизайну системи, описуючи основні компоненти та їх взаємодію. Третій підрозділ присвячений платформі Hasura, яка використовується для управління базою даних та забезпечення швидкого доступу до даних через GraphQL. Четвертий підрозділ розглядає механізми інтеграції з соціальними медіа платформами Facebook та Instagram. П’ятий підрозділ описує сервіси-аналізатори GPT, Claude та Gemini, які відповідають за аналіз настроїв та надання інсайтів на основі контенту соціальних медіа. Шостий підрозділ розглядає використання черг RabbitMQ для обробки повідомлень. Сьомий підрозділ надає огляд інтерфейсу користувача та функціональних можливостей користувацького порталу.

4.1. Схема даних

Схема даних передбачає керування інтеграціями користувачів із зовнішніми платформами соціальних мереж, обробки різних типів контенту, що генерується користувачами, та аналізу цього контенту за допомогою різних моделей.

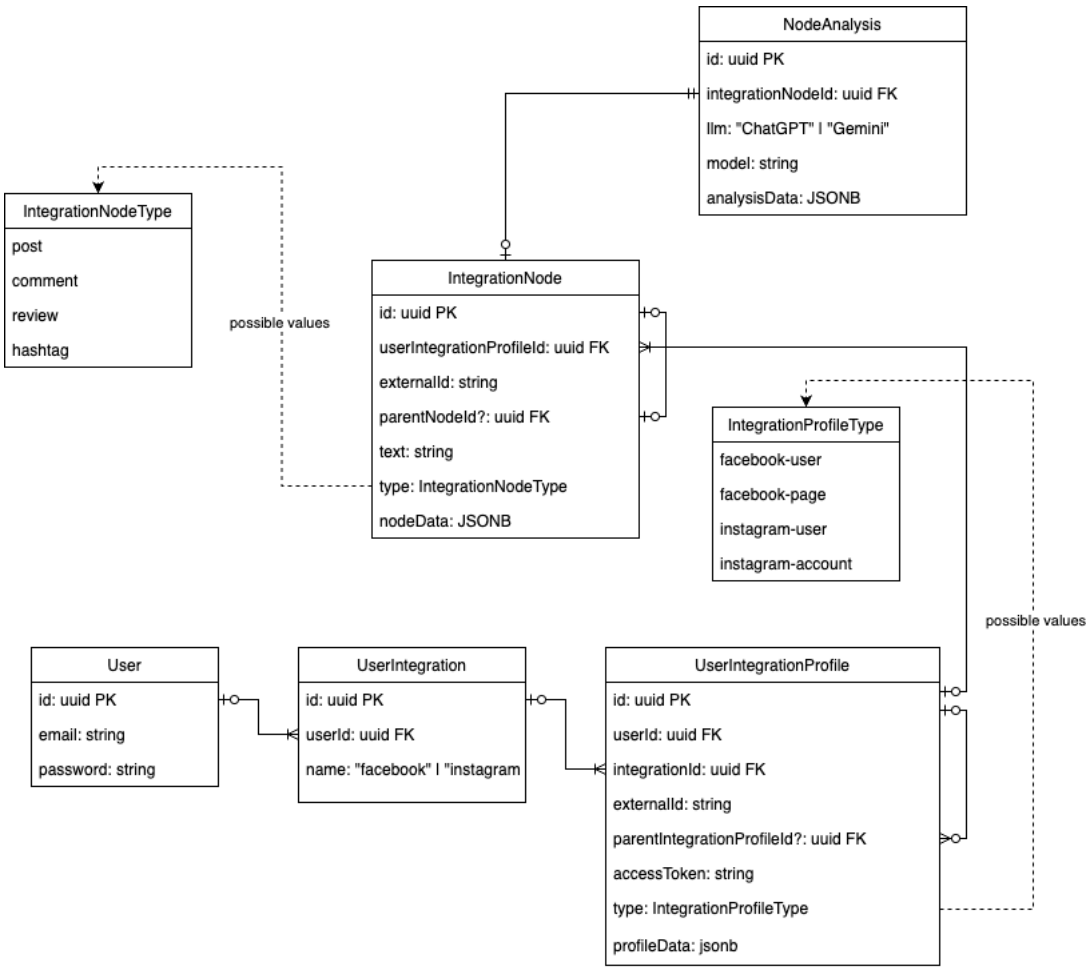


Рисунок 4.1. Схема даних системи-прототипу

- **Управління Користувачами**

Схема починається з таблиці User, яка містить основну інформацію про користувачів. Кожен користувач може підключатися до кількох зовнішніх платформ, таких як Facebook та Instagram, через таблицю UserIntegration.

- **Профілі Інтеграції**

Кожна інтеграція користувача може мати кілька профілів, представлених у таблиці UserIntegrationProfile. Наприклад, інтеграція з Facebook може включати кілька сторінок або акаунтів. Кожен профіль пов'язаний з конкретним користувачем та інтеграцією, і може включати додаткові ієрархічні відносини через необов'язкові батьківські профілі. Така структура підтримує складні сценарії, де профілі можуть вкладатися один в один. Наприклад, на платформі Facebook кожна сутність сторінки відноситься до якогось користувача, тому ця схема даних підтримує можливість для внутрішнього користувача мати декілька

аккаунтів користувача з платформи разом із усіма відповідними дочірніми обліковими записами.

- **Вузли Контенту**

Контент, створений користувачами на цих платформах, такий як пости, коментарі, відгуки чи хештеги, зберігається в таблиці IntegrationNode. Кожен вузол пов'язаний з певним профілем інтеграції користувача і може також посилатися на батьківський вузол, що дозволяє представляти вкладений контент, такий як коментарі-відповіді на коментар або пов'язані пости. Вузли можуть містити різні типи контенту, визначені заздалегідь, як пост, коментар, відгук чи хештег.

- **Аналіз Контенту**

Для надання аналітичних даних про вузли контенту використовується таблиця NodeAnalysis. Кожен вузол контенту може мати відповідний запис аналізу. Ця таблиця зберігає дані про виконаний аналіз, такі як модель мови, що використовувалася (наприклад, ChatGPT або Gemini), та результати аналізу. Відношення між IntegrationNode та NodeAnalysis є один до одного, що забезпечує унікальність аналізу кожного фрагменту контенту.

- **Переліки**

Схема також включає переліки для визначення типів вузлів і профілів. Переліки IntegrationNodeType та IntegrationProfileType містять можливі значення для типів вузлів контенту та профілів інтеграції відповідно. Такий підхід забезпечує узгодженість і цілісність даних, обмежуючи введення до заздалегідь визначених типів.

Загалом, наведена схема підтримує управління взаємодіями користувачів із соціальними мережами, обробку складних відносин між контентом та виконання детального аналізу контенту. Вона спроектована так, щоб бути гнучкою, масштабованою та зручною в обслуговуванні, з надійною підтримкою ієрархічних даних та розширених типів даних.

4.2. Архітектура системи

Система призначена для синхронізації, обробки та аналізу даних із соціальних мереж Facebook та Instagram. Вона побудована на основі бази даних PostgreSQL із використанням сервісу Hasura, який забезпечує рівень доступу до даних. Центральний сервер координує взаємодію між користувацьким інтерфейсом та різними бекенд-сервісами, включаючи інтеграційні сервіси для платформ соціальних мереж та аналітичні сервіси, що використовують великі мовні моделі (LLM). Для забезпечення ефективної та масштабованої обробки даних система використовує RabbitMQ для організації черг повідомлень.

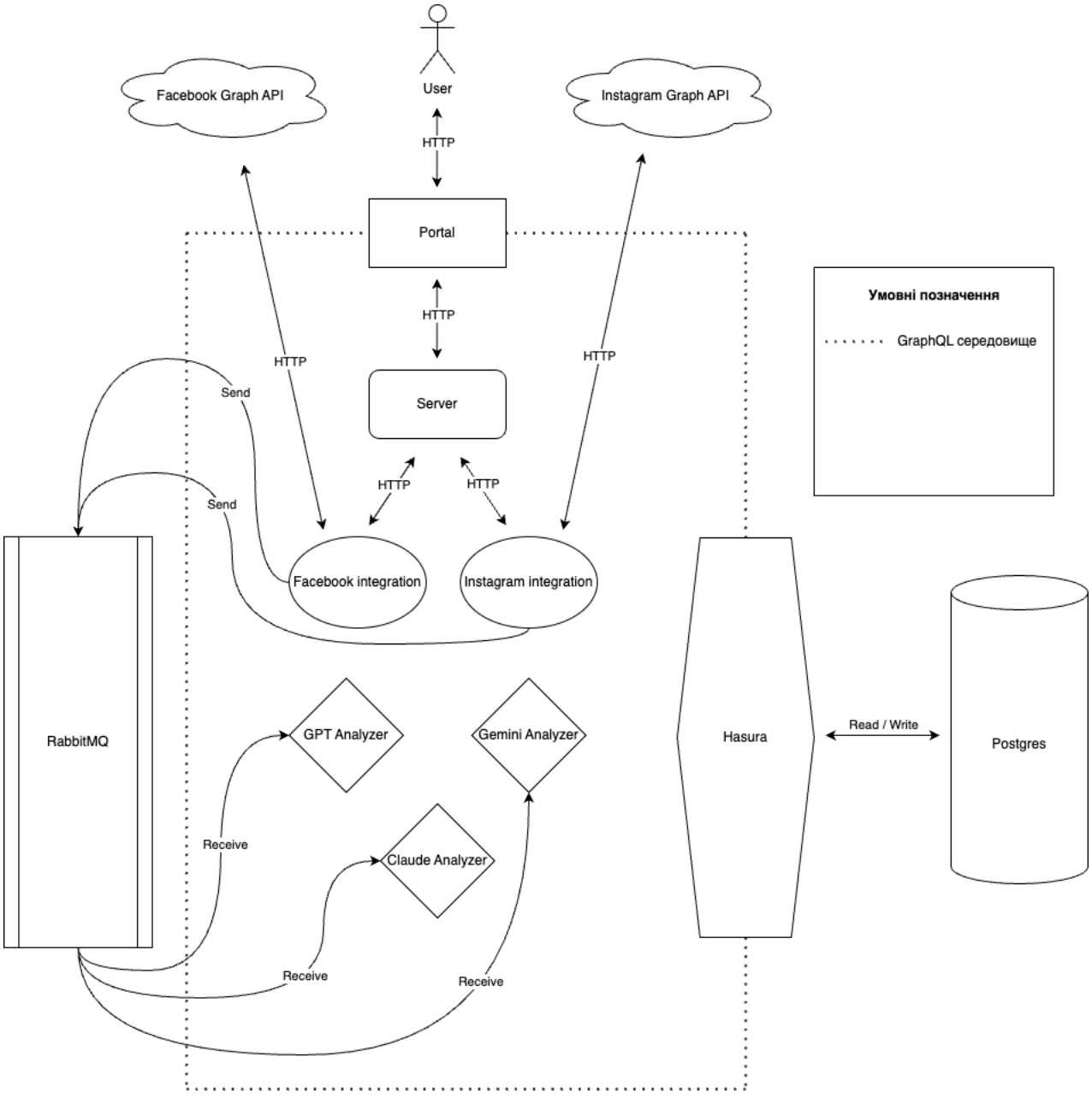


Рисунок 4.2. Діаграма архітектури розробленої системи-прототипу

Короткий опис кожного сервісу-компоненти:

- **Hasura**

Hasura використовується як рівень доступу до даних, надаючи інтерфейс GraphQL для бази даних. Цей рівень абстрагує складності взаємодії з базою даних, дозволяючи безперешкодно виконувати операції читання та запису, а також забезпечує, що всі операції з базою даних виконуються ефективно та безпечно.

- **Центральний сервер**

Центральний сервер є основним координатором системи. Він обробляє HTTP-запити від користувацького порталу та керує всіма бекенд-процесами. Сервер виконує автентифікацію користувачів, запускає синхронізацію даних з платформ соціальних мереж та координує взаємодію між інтеграційними сервісами та чергою повідомлень.

- **Користувацький портал**

Користувацький портал є веб-інтерфейсом, через який користувачі взаємодіють із системою. Він надає функціональні можливості, такі як реєстрація користувачів, вхід у систему та процес налаштування інтеграцій з Facebook та Instagram, а також інформаційну панель про поточний рівень сентименту, статистичні дані аналізу та інші актуальні для користувачів дані. Портал взаємодіє з центральним сервером через HTTP-запити, а також застосовує GraphQL для звернення до сервісу Hasura.

- **Інтеграційні сервіси (Facebook, Instagram)**

Ці сервіси відповідають за взаємодію з платформами Facebook та Instagram. Вони обробляють автентифікацію через OAuth2, отримують дані користувачів та витягають необхідний контент із платформ. Після успішного отримання даних сервіси порівнюють нові дані з існуючими записами в базі даних для виявлення нових або оновлених вузлів. Потім ці сервіси трансформують дані та зберігають їх у базі даних PostgreSQL через Hasura.

- **RabbitMQ**

RabbitMQ використовується як система організації черг повідомлень. Він забезпечує асинхронну комунікацію між центральним сервером, інтеграційними та аналітичними сервісами. Після збереження нових або оновлених даних у базі даних, вони надсилаються в RabbitMQ. Це розділяє процеси отримання та обробки даних для підвищує масштабованості і продуктивності системи.

- **Аналітичні сервіси (GPT Analyzer, Gemini Analyzer, Claude Analyzer)**

Аналітичні сервіси відповідають за обробку та аналіз даних, отриманих із платформ соціальних мереж. Вони отримують повідомлення з RabbitMQ у круговому порядку (round-robin), забезпечуючи балансування навантаження. Кожен сервіс трансформує отримані дані у відповідний формат та надсилає їх до відповідної великої мовної моделі для аналізу сентименту. Результати аналізу потім зберігаються в базі даних PostgreSQL як сутності аналізу вузлів, кожна з яких пов'язана з оригінальними вузлами даних.

4.3. Hasura

Hasura є потужним інструментом з відкритого джерела, який забезпечує миттєве створення реальних GraphQL та REST API на нових і існуючих базах даних PostgreSQL. Архітектурно Hasura складається з кількох компонентів. Hasura GraphQL Engine взаємодіє з базою даних PostgreSQL для автоматичного створення GraphQL та REST API. Він відповідає за обробку запитів, забезпечення дозволів та синхронізацію даних в реальному часі. Hasura Console надає графічний інтерфейс для управління схемою бази даних, налаштування дозволів та конфігурації API кінцевих точок, а також для моніторингу та налагодження API запитів.

Основна функція Hasura полягає в автоматичному генеруванні GraphQL та REST API на основі існуючої схеми бази даних. Це усуває потребу в написанні великої кількості шаблонного коду, що значно скорочує час розробки. Інструмент також підтримує реальні запити та підписки, дозволяючи отримувати оновлення з бази даних в режимі реального часу без необхідності у додатковій інфраструктурі або складному кодуванні.

Ключовий аспект Hasura – її потужні механізми авторизації та контролю доступу. Платформа дозволяє встановлювати детальні, рольові дозволи як на рівні рядків, так і на рівні колонок, що забезпечує безпечний та відповідний доступ до даних. Крім того, Hasura можна розширювати за допомогою користувацької бізнес-логіки через віддалені схеми та тригери подій. Віддалені схеми дозволяють інтеграцію зовнішніх GraphQL схем, тоді як тригери подій можуть викликати безсерверні функції або вебхуки у відповідь на події бази даних, що сприяє реалізації складної бізнес-логіки.

Hasura надає можливість інтеграції з різними провайдерами аутентифікації, включаючи JWT, Auth0 та Firebase Auth, для ефективного управління доступами та ролями користувачів. Ця вбудована підтримка аутентифікації спрощує процес захисту API кінцевих точок та управління сеансами користувачів.

Ще однією важливою особливістю Hasura є можливість об'єднання даних з декількох джерел. Це дозволяє розробникам об'єднувати дані з різних баз даних або зовнішніх API в єдиний уніфікований GraphQL API, забезпечуючи єдиний та комплексний шар доступу до даних.

Платформа також включає інструменти для управління міграціями бази даних та метаданими, які є необхідними для підтримання послідовності та відстежуваності в розробці додатків. Метадані в Hasura включають інформацію про GraphQL схему, дозволи, зв'язки та користувацьку логіку, тоді як міграції забезпечують контроль версій схеми бази даних та метаданих, забезпечуючи узгоджені розгортання у різних середовищах.

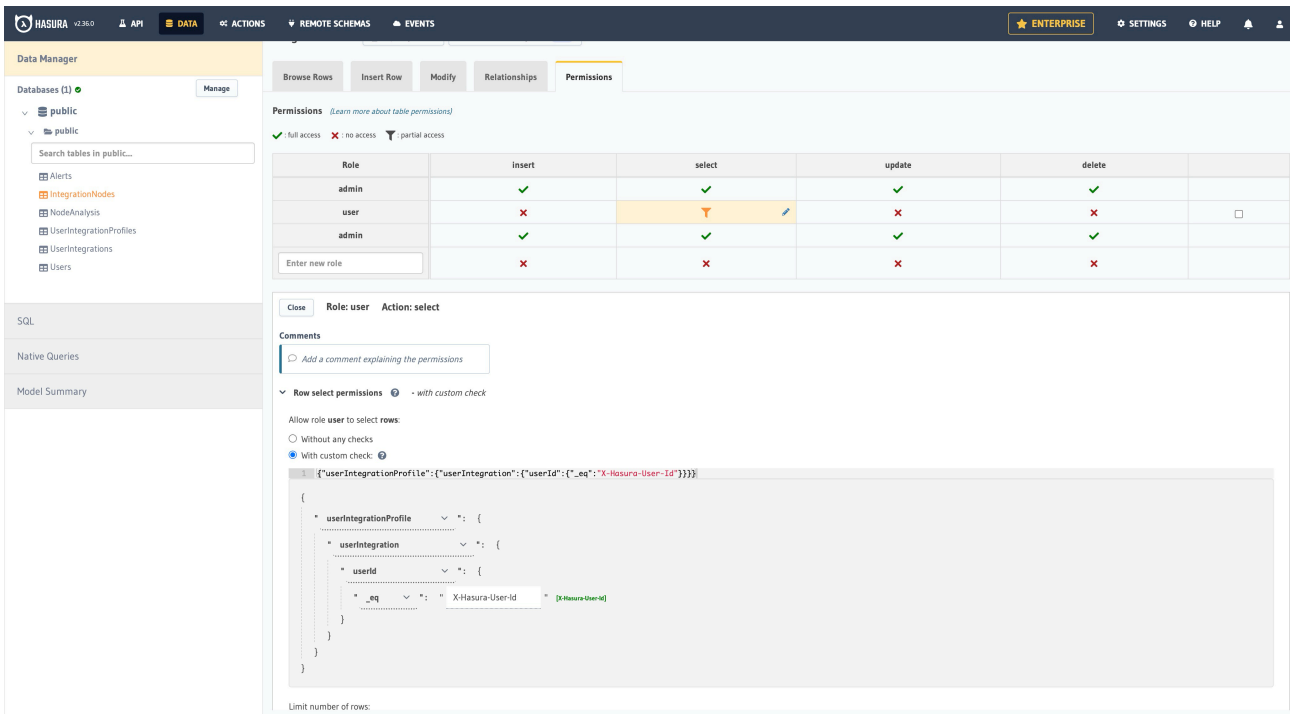


Рисунок 4.3. Приклад налаштування дозволів для ролі користувача у графічному інтерфейсі інструменту Hasura

4.4. Сервіси-інтеграції Facebook та Instagram

Інтеграційні сервіси для Facebook та Instagram розроблені для забезпечення безперебійної взаємодії між платформами та застосунком. Основне призначення інтеграційних сервісів для Facebook та Instagram – підключення системи до відповідних соціальних медіа платформ, що дозволяє витягувати та синхронізувати дані користувачів та контент. Ці сервіси забезпечують процеси автентифікації, отримання даних та їх перетворення у формат, придатний для зберігання та аналізу.

Інтеграційні сервіси використовують протоколи OAuth2 для автентифікації користувачів у Facebook та Instagram. Цей процес включає перенаправлення користувачів на сторінку входу соціальної медіа платформи, де вони надають дозволи для додатка на доступ до їхніх даних. Після успішної автентифікації сервіс отримує токени доступу, які дозволяють виконувати API виклики від імені користувача.

Після автентифікації інтеграційні сервіси отримують дані користувачів та відповідний контент з Facebook та Instagram за допомогою Graph API. Це включає

пости, коментарі, лайки, поширення та будь-які інші відповідні дані, доступні через API. Сервіси забезпечують, щоб процес отримання даних був ефективним і відповідав обмеженням швидкості та політикам, встановленим соціальними медіа платформами.

Отримані дані порівнюються з існуючими записами в базі даних PostgreSQL системи. Цей етап включає ідентифікацію нового або оновленого контенту, який потрібно синхронізувати. Інтеграційні сервіси забезпечують логіку для об'єднання нових даних з існуючими записами, гарантуючи, що база даних завжди актуальна з останньою інформацією з соціальних медіа платформ.

Після ідентифікації нових або оновлених даних інтеграційні сервіси перетворюють дані у формат, сумісний зі схемою бази даних. Це включає парсинг JSON відповідей від API, нормалізацію даних та забезпечення їх відповідності реляційній структурі бази даних PostgreSQL. Перетворені дані зберігаються в базі даних через інтерфейс Hasura GraphQL.

4.4.1. Правила синхронізації даних платформ з системою

Синхронізація контенту відбувається за наступними правилами:

- Через невелику кількість сторінок Facebook та під’єднаних облікових записів Instagram, пов'язаних з бізнесом (у порівнянні з можливою кількістю постів і коментарів), синхронізація сторінок буде виконуватись за допомогою запланованого завдання, яке періодично запускатиме синхронізацію сторінок. Це завдання також запускатиме синхронізацію постів на сторінках кожного разу, навіть для тих сторінок, які вже присутні в базі даних.
- Кожен пост, отриманий з Facebook чи Instagram, має поле updated_time, яке вказує на останню зміну, зроблену з цим постом, включаючи випадки, коли користувач опублікував коментар стосовно цього посту. Збереження updated_time у нашій базі даних і порівняння його з майбутніми отриманими постами допоможе нам зрозуміти, чи були зміни в пості або чи було додано коментар.

- Коментарі синхронізуються тоді, коли у пов'язаного поста буде оновлюється значення поля updated_time

4.4.2. Механізм роботи з хештегами платформи Instagram

Для інтеграційного сервісу Instagram був побудований складний механізм роботи з хештегами, оскільки Instagram має обмеження щодо кількості запитів кожного унікального хештегу, яке не повинно перевищувати 30 запитів за 7 днів для кожного користувача.

Користувачі мають можливість створювати списки хештегів, які будуть періодично зчитуватися для отримання та синхронізації пов'язаних медіафайлів. Хештеги представлені в базі даних як IntegrationNode:

- Поле **text** – представляє назву хештегу
- Поле **externalId** – ID хештегу в Instagram
- Поле **updatedAt** – це час останнього запиту цього хештегу
- Поле **nodeData.isActive** – вказує, чи повинен цей хештег зчитуватися
- Поле **nodeData.isNew** – вказує, чи хештег є новим (це запускає синхронізацію і після першої обробки встановлюється в значення **false**)
- Поле **nodeData.initallyQueriedAt** – час початкового запиту хештегу для відстеження 7-денного періоду.

Процес синхронізації хештегів відбувається наступним чином:

1. Отримати всі хештеги користувача, для яких поле **initiallyQueriedAt** не є порожнім або **isNew** істинне.
2. Відфільтрувати застарілі хештеги, для яких значення **initiallyQueriedAt** + 7 днів більше або дорівнює поточній даті:
    a. Якщо застарілий хештег має **isActive** хибне, встановити для нього **initiallyQueriedAt** порожнє значення.
    b. Якщо застарілий хештег має **isActive** істинне, встановити для нього **initiallyQueriedAt** порожнє значення та **isNew** істинне.
3. Обробити всі хештеги, які мають істинне значення **isNew**:
    a. Встановити для них значення **updatedAt** поточної дати та часу.

b. Встановити для них значення **initiallyQueriedAt** поточної дати та часу.

4.5. Сервіси-аналізатори GPT, Claude та Gemini

Сервіси-аналізатори для GPT, Claude та Gemini розроблені для обробки та аналізу даних, отриманих із соціальних медіа платформ, надаючи інсайти та аналіз настроїв за допомогою передових великих мовних моделей. Кожен аналізатор виконує однакові основні завдання, але архітектура передбачає майбутню можливість для користувачів надавати пріоритет результатам від певних LLM залежно від їхніх уподобань та майбутніх потреб.

Основні компоненти та процеси аналізаторів:

1. **Отримання даних та їх черговість**

Сервіси-аналізатори отримують дані з черги повідомлень RabbitMQ. Це розділяє отримання даних від їх обробки, забезпечуючи масштабованість і ефективність системи. Черга повідомлень розподіляє дані між аналізаторами за круговою схемою, забезпечуючи балансування навантаження та ефективну обробку.

2. **Трансформація даних**

Кожен аналізатор перетворює отримані дані у формат, придатний для аналізу відповідною LLM. Це включає парсинг контенту, вилучення релевантних характеристик та структурування даних відповідно до вимог моделей.

3. **Аналіз настроїв та обробка**

Використовуючи передові LLM, кожен аналізатор обробляє трансформовані дані для визначення настроїв та інших релевантних метрик. Специфічні моделі, які використовуються кожним аналізатором, такі:

- **GPT Аналізатор**: використовує Generative Pre-trained Transformer (GPT) від OpenAI для проведення всебічного аналізу настроїв і генерації інсайтів на основі контенту. GPT відзначається розумінням контексту та створенням детальних оцінок настроїв.

- **Claude Аналізатор**: використовує модель Claude від Anthropic, яка забезпечує надійний аналіз настроїв та контекстуальне розуміння, з акцентом на етичні міркування та інтерпретацію намірів користувачів.
- **Gemini Аналізатор**: використовує модель Gemini від Google, оптимізовану для передового аналізу настроїв у різних типах даних, включаючи текст, зображення та відео.

4. **Зберігання результатів аналізу**

Після обробки результати аналізу зберігаються у базі даних PostgreSQL. Кожен результат пов'язується з оригінальними вузлами даних, забезпечуючи відстежуваність і узгодженість. Це передбачає створення нових вузлів у базі даних, що представляють сутності аналізу настроїв, які асоціюються з відповідним контентом соціальних медіа.

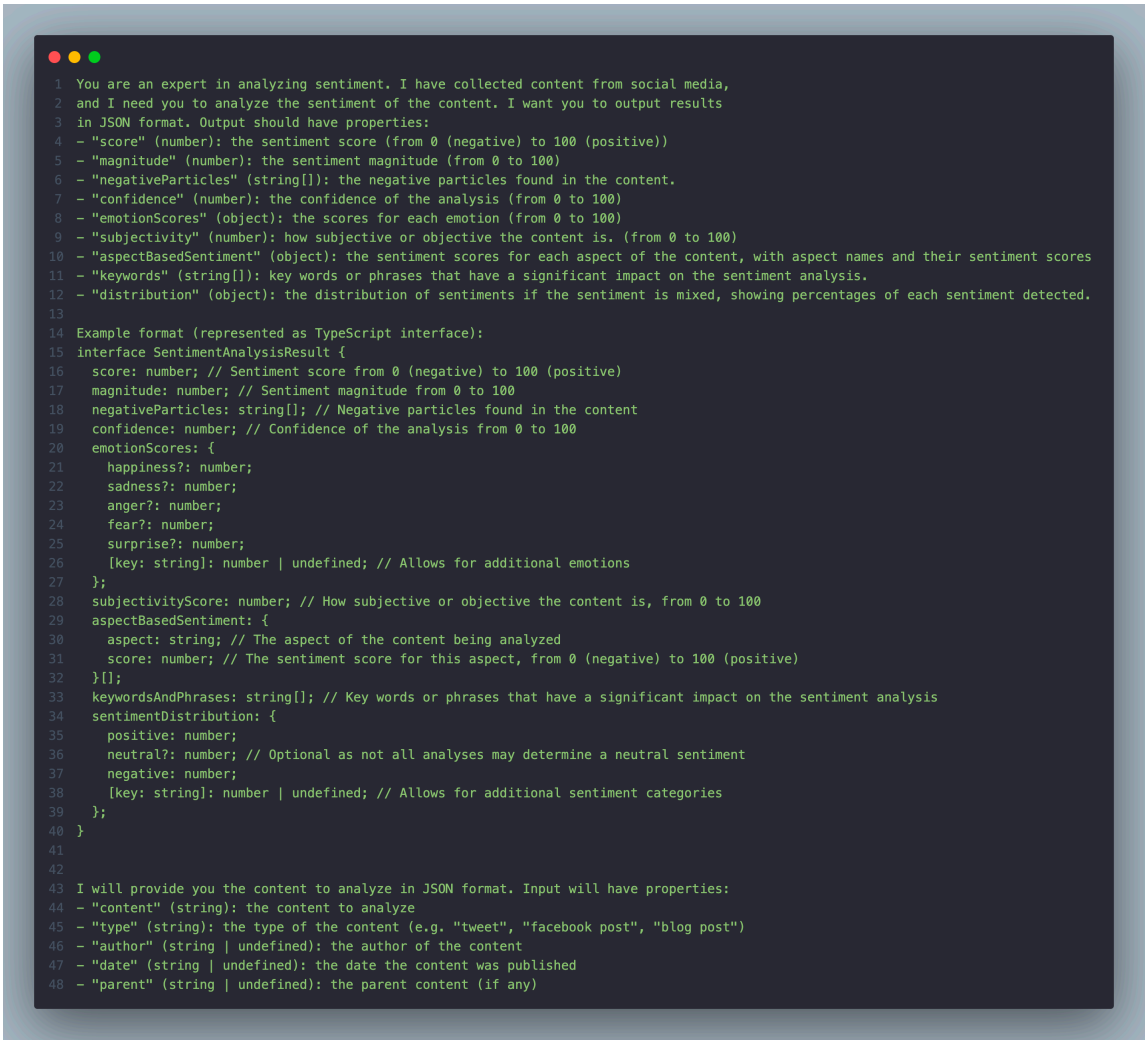
5. **Обробка помилок та логування**

Кожен аналізатор включає надійні механізми обробки помилок для управління будь-якими проблемами, що виникають під час обробки даних. Помилки логуються для налагодження та аналізу, забезпечуючи надійність і стабільність аналізаторів.

Однією з важливих переваг цієї архітектури є гнучкість, яку вона надає користувачам. У майбутньому користувачі можуть обирати пріоритетність результатів аналізу від певної LLM залежно від їхніх потреб та уподобань. Ця можливість дозволятиме отримувати інсайти та більш релевантний аналіз, підвищуючи корисність системи для різних сценаріїв використання.

Кожен з аналізаторів використовує аналогічні системні підказки для генерації результатів аналізу. У майбутньому планується оновлювати підказки для кожної з інтеграцій залежно від їх ефективності та точності, базовано на відгуках користувачів та аналітиці їхніх пріоритезованих великих мовних моделей. Ідея такої структури полягає в отриманні максимальної кількості корисної інформації від великих мовних моделей. Це включає не лише задачу класифікації тексту в межах трьох категорій – позитивної, негативної чи

нейтральної – а й витяганні допоміжної інформації, яка буде корисна користувачам. Структура результату аналізу наведена на рисунку 4.4.



```
You are an expert in analyzing sentiment. I have collected content from social media,
and I need you to analyze the sentiment of the content. I want you to output results
in JSON format. Output should have properties:
- "score" (number): the sentiment score (from 0 (negative) to 100 (positive))
- "magnitude" (number): the sentiment magnitude (from 0 to 100)
- "negativeParticles" (string[]): the negative particles found in the content.
- "confidence" (number): the confidence of the analysis (from 0 to 100)
- "emotionScores" (object): the scores for each emotion (from 0 to 100)
- "subjectivity" (number): how subjective or objective the content is. (from 0 to 100)
- "aspectBasedSentiment" (object): the sentiment scores for each aspect of the content, with aspect names and their sentiment scores
- "keywords" (string[]): key words or phrases that have a significant impact on the sentiment analysis.
- "distribution" (object): the distribution of sentiments if the sentiment is mixed, showing percentages of each sentiment detected.

Example format (represented as TypeScript interface):
interface SentimentAnalysisResult {
  score: number; // Sentiment score from 0 (negative) to 100 (positive)
  magnitude: number; // Sentiment magnitude from 0 to 100
  negativeParticles: string[]; // Negative particles found in the content
  confidence: number; // Confidence of the analysis from 0 to 100
  emotionScores: {
    happiness?: number;
    sadness?: number;
    anger?: number;
    fear?: number;
    surprise?: number;
    [key: string]: number | undefined; // Allows for additional emotions
  };
  subjectivityScore: number; // How subjective or objective the content is, from 0 to 100
  aspectBasedSentiment: {
    aspect: string; // The aspect of the content being analyzed
    score: number; // The sentiment score for this aspect, from 0 (negative) to 100 (positive)
  }[];
  keywordsAndPhrases: string[]; // Key words or phrases that have a significant impact on the sentiment analysis
  sentimentDistribution: {
    positive: number;
    neutral?: number; // Optional as not all analyses may determine a neutral sentiment
    negative: number;
    [key: string]: number | undefined; // Allows for additional sentiment categories
  };
}


I will provide you the content to analyze in JSON format. Input will have properties:
- "content" (string): the content to analyze
- "type" (string): the type of the content (e.g. "tweet", "facebook post", "blog post")
- "author" (string | undefined): the author of the content
- "date" (string | undefined): the date the content was published
- "parent" (string | undefined): the parent content (if any)
```

Рисунок 4.4. Приклад системної підказки у сервісі-аналізаторі для великої мовної моделі

Така структура бажаного результату аналізу даних, отриманих з соціальних мереж, має такі ключові аспекти:

- Запит чітко визначає бажаний формат вихідних даних, зменшуючи неоднозначність та забезпечуючи узгоджені, структуровані результати аналізу настроїв. Це сприяє подальшому парсингу даних та їх використанню.
- Запитуючи широкий спектр метрик настроїв, включаючи загальний показник настрою, його силу, впевненість, суб'єктивність та детальні оцінки емоцій, запит дозволяє глибоко зрозуміти настрої, захоплюючи як загальний тон, так і специфічні емоційні нюанси.

- Включення аналізу настроїв на основі аспектів дозволяє детально досліджувати настрої, пов'язані з різними елементами контенту. Це особливо цінно для складних постів у соціальних мережах, де різні аспекти можуть мати різні настрої.
- Витягування негативних часток і значущих ключових слів допомагає зрозуміти конкретні компоненти, які впливають на настрій. Це знання корисне для глибокого аналізу та цілеспрямованих відповідей або втручань на основі контенту.
- Включення показника впевненості допомагає оцінити надійність аналізу, що важливо для прийняття обґрунтованих рішень. Показник суб'єктивності дає уявлення про ступінь, до якого контент є думковим або фактичним, впливаючи на інтерпретацію настроїв.
- Аналіз розподілу настроїв надає розбивку змішаних настроїв у контенті, пропонуючи всебічний огляд загального настрою, особливо в контенті з як позитивними, так і негативними елементами.
- Структурований вихідний формат з опціональними та додатковими полями (наприклад, опціональний нейтральний настрій, додаткові емоції та категорії настроїв) забезпечує гнучкість і розширюваність, що дозволяє вносити майбутні зміни або додавати нові категорії настроїв з мінімальними структурними змінами.
- Зазначений формат введення, що включає тип контенту, автора, дату та батьківський контент, надає контекстну інформацію, що покращує аналіз настроїв. Наприклад, тип контенту (наприклад, твіт, пост у Facebook) дозволяє налаштовувати аналіз відповідно до характеристик різних соціальних медіа платформ.
- Використання JSON як для введення, так і для виведення забезпечує сумісність з різними системами та мовами програмування.

Робота сервісів-аналізаторів напряму залежить від реалізації черги повідомлень, яка буде описана у наступному підрозділі.

4.6. Механізм роботи черги повідомлень

Логіка обробки з'єднань RabbitMQ включає використання трьох конкретних типів черг: черги аналізу, черги затримки та черги непридатних повідомлень. Кожна черга виконує окрему функцію для забезпечення надійної передачі та обробки повідомлень у системі.

4.6.1. Черга аналізу

Черга аналізу є основною чергою, куди спочатку надсилаються повідомлення для обробки. Ці повідомлення зазвичай містять інтеграційні вузли, отримані від інтеграційних сервісів, які потрібно проаналізувати аналітичними сервісами.

Логіка черги:

- **Отримання повідомлень**

  Коли повідомлення надсилається в систему, воно поміщається в чергу аналізу.

- **Обробка**

  Споживач прослуховує повідомлення в черзі аналізу. Коли повідомлення отримується, воно обробляється за допомогою наданої функції обробки, яка виконує аналіз настроїв або інші необхідні завдання.

- **Підтвердження**

  Якщо повідомлення успішно оброблено, воно підтверджується, і RabbitMQ видаляє його з черги.

- **Відхилення**

  Якщо під час обробки виникає помилка, повідомлення не підтверджується. Замість цього воно відхиляється і направляється в чергу затримки для повторної спроби обробки через певний час.

4.6.2. Черга затримки

Черга затримки використовується для тимчасового зберігання повідомлень, які не вдалося обробити. Вона реалізує механізм повторних спроб,

утримуючи повідомлення протягом заданого часу існування (TTL), перш ніж вони знову потрапляють у чергу аналізу для повторної спроби обробки.

Логіка черги:

- **Маршрутизація повідомлень**

  Коли повідомлення не вдається обробити з черги аналізу, воно направляється в чергу затримки.

- **TTL**

  Повідомлення знаходиться в черзі затримки протягом заданого часу затримки.

- **Черга непридатних повідомлень**

  Після закінчення TTL повідомлення автоматично направляється назад у чергу аналізу через механізм dead-letter exchange.

- **Повторні спроби**

  Цей процес повторюється, дозволяючи повідомленню спробувати обробку кілька разів. Кількість повторних спроб відстежується за допомогою заголовка x-death, який підраховує кількість разів, коли повідомлення проходило через чергу затримки.

4.6.3. Черга непридатних повідомлень

Черга непридатних повідомлень використовується для захоплення та зберігання повідомлень, які перевищили максимальну кількість повторних спроб. Ця черга слугує запобіжним заходом для запобігання безкінечним циклам повторних спроб і дозволяє ручне втручання або подальший аналіз проблемних повідомлень.

Логіка черги:

- **Обмеження повторних спроб**

  Коли повідомлення відхиляється з черги аналізу і було повторно оброблено більше разів, ніж дозволено, воно більше не направляється в чергу затримки.

- **Маршрутизація**

Замість цього повідомлення направляється в чергу непридатних повідомлень, що вказує на те, що неодноразово не вдалося обробити повідомлення і потребує уваги.

- **Ручне втручання**

  Повідомлення в цій черзі можуть бути переглянуті та оброблені вручну системними адміністраторами або розробниками. Це може включати налагодження проблеми, зміну повідомлення або вживання інших коригувальних заходів.

- **Логування та сповіщення**

  Система реєструє детальну інформацію про повідомлення, направлені в чергу непридатних повідомлень, включаючи причину відмови та кількість повторних спроб, щоб допомогти в діагностиці та вирішенні проблем.

4.7. Користувацький портал

Фронтенд-застосунок користувацького порталу призначений для забезпечення користувачів інтуїтивно зрозумілим та зручним інтерфейсом для керування інтеграціями з соціальними медіа платформами, такими як Facebook та Instagram. Портал слугує основною точкою взаємодії для користувачів, дозволяючи їм переглядати статистичні дані, детальну інформацію про здійснений сентимент аналіз, реагувати на випадки негативних відгуків та швидко на них реагувати.

Інтерфейс розроблений таким чином, щоб бути простим та легким у навігації, з мінімальною кількістю зайвих елементів та чіткими закликами до дії. Дизайн використовує чисту та сучасну естетику, з акцентом на читабельність та комфорт користувача. Кольорова схема проста, але ефективна, з використанням брендових кольорів для кнопок та іконок, що забезпечує візуальну узгодженість. Функції, такі як перемикач теми, підвищують доступність, дозволяючи користувачам переключатися між світлим та темним режимами, адаптуючись до різних умов освітлення та особистих вподобань.

4.7.1. Сторінка процесу адаптації

На цю сторінку перенаправляються усі нові користувачі, де вони можуть активувати необхідні інтеграція для початку роботи з системою. На сторінці присутні на поточний момент 2 кнопки:

- **Кнопка «Підключити Facebook»**

  Синя кнопка з написом «Підключити Facebook» та логотипом Facebook дозволяє користувачам розпочати процес інтеграції зі своїм обліковим записом Facebook. Ця кнопка запускає процес автентифікації за допомогою OAuth2 для безпечного підключення даних користувача з Facebook до застосунку.

- **Кнопка «Підключити Instagram»**

  Рожева кнопка з написом «Підключити Instagram» та логотипом Instagram надає аналогічну функцію для Instagram. Користувачі можуть натиснути цю кнопку, щоб розпочати процес автентифікації та інтеграції зі своїм обліковим записом Instagram

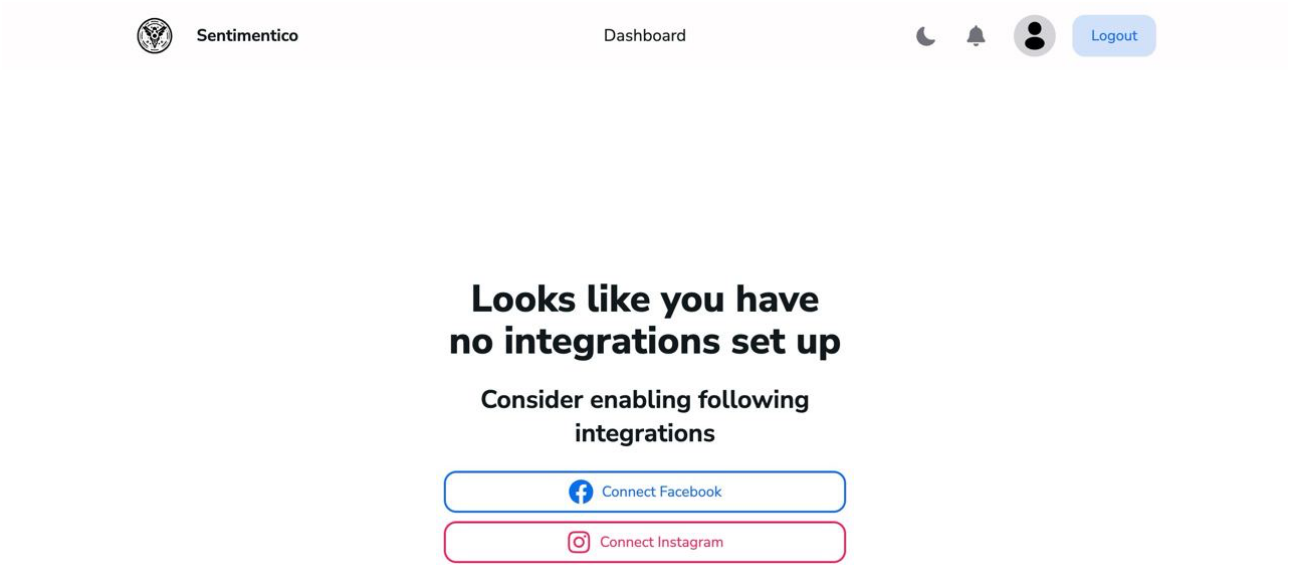


Рисунок 4.5. Сторінка процесу адаптації користувача порталу

Після успішного встановлення інтеграцій – користувач має можливість перейти на сторінку інформаційної панелі.

4.7.2. Вкладка підсумку

Сторінка інформаційної панелі має шаблонний макет, який містить випадаюче меню вибору поточного профілю для перегляду його інформації. Під випадаючим меню відображається список доступних профілів, розподілених за соціальними медіа платформами (наприклад, Facebook, Instagram). Кожен запис профілю включає значок, що представляє платформу, і ім'я профілю, що полегшує користувачам перемикання між різними обліковими записами. За замовчуванням обрано профіль організації, який показуй загальні дані серед усіх синхронізованих профілів користувача.

Перша вкладка інформаційної панелі є вкладка підсумку, яка містить 3 ключові елементи:

- **Загальний показник настрою**

  Круговий індикатор прогресу відображає загальний показник настрою у вигляді відсотка. Цей візуальний елемент дає користувачам швидке уявлення про загальний настрій, отриманий з усіх синхронізованих вузлів.

- **Розподіл настроїв**

  Кільцева діаграма показує розподіл аналізованого контенту за показниками настрою, розподіленими на п'ять рівнів: Поганий, Негативний, Нейтральний, Позитивний та Відмінний. Ця діаграма допомагає користувачам зрозуміти склад настроїв у їхньому контенті з першого погляду.

- **Діаграма площ**

  Діаграма площ візуалізує відсотки позитивних і негативних настроїв за певні часові діапазони. Ця діаграма включає:

  - Позитивний настрій (зелений): представлений зеленими областями, що показують частку позитивного настрою.

- Негативний настрій (червоний): представлений червоними областями, що вказують на частку негативного настрою.
- Лінія загального настрою: помаранчева лінія, що відображає загальну тенденцію настроїв з часом.
- Інтерактивні точки даних: діаграма включає інтерактивні точки даних, на які користувачі можуть навести курсор, щоб побачити детальну статистику за конкретні дати, включаючи відсотки позитивних і негативних настроїв та їхні середні значення.

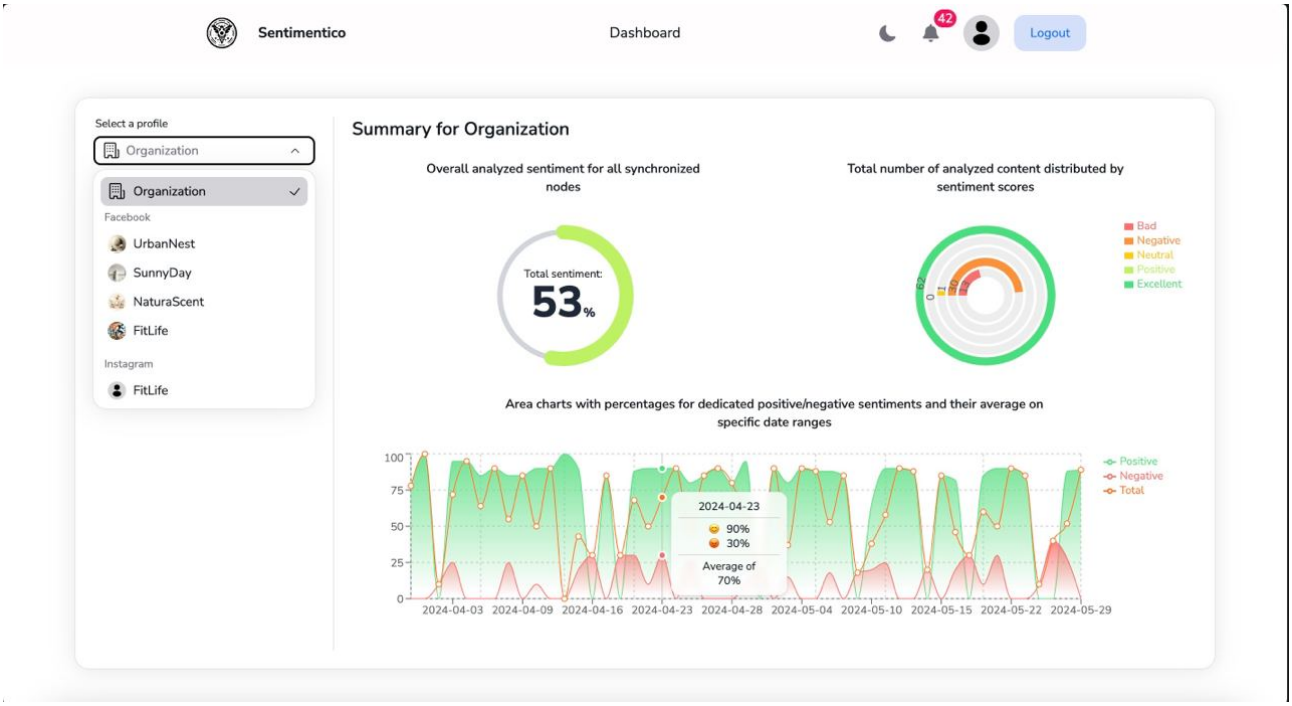


Рисунок 4.6. Сторінка інформаційної панелі з активною вкладкою підсумку

Вкладка підсумку сторінки панелі інструментів надає чіткий і зрозумілий огляд загального настрою, що дозволяє користувачам легко зрозуміти загальний настрій їхньої аудиторії з першого погляду. Цей огляд є важливим для швидкої оцінки ефективності стратегій у соціальних медіа.

Завдяки наданню детального огляду розподілу настроїв за категоріями, такими як Поганий, Негативний, Нейтральний, Позитивний та Відмінний, панель інструментів дозволяє користувачам виявити конкретні області занепокоєння або успіху. Ця розбивка допомагає визначити конкретний контент або теми, які резонують або не резонують з аудиторією.

Діаграма площ, що відображає тенденції настроїв з часом, надає цінний історичний контекст. Користувачі можуть відстежувати, як змінюється настрій у відповідь на різні кампанії, події або типи контенту, отримуючи інсайти щодо того, що викликає позитивні або негативні реакції.

Можливість перемикання між різними профілями соціальних медіа забезпечує користувачам можливість налаштовувати свій аналіз для конкретних облікових записів. Ця функція є особливо цінною для користувачів, які керують кількома брендами або сегментами, оскільки вона дозволяє здійснювати цільовий аналіз та оптимізацію стратегій.

4.7.3. Вкладка інформації про пости

Вкладка інформації про пости розбита на 2 секції:

- **Синхронізовані та проаналізовані пости**

  Центральна панель відображає список синхронізованих та проаналізованих постів з обраного профілю.

- **Огляд вибраного поста**

  Права панель надає детальну інформацію про вибраний пост, що включає ім'я профілю та значок, а також посилання для перегляду поста на платформі, числовий показник настрою, представлений у відсотках, що вказує на загальний настрій поста, кількість проаналізованих елементів контенту для вибраного поста, кількість негативних коментарів, пов'язаних з постом, а також основний текст поста, що дозволяє користувачам прочитати повний текст

- **Секція коментарів**

  Нижче основного контенту поста знаходиться секція для перегляду та управління коментарями, яка містить вміст коментарів з часовими мітками та показниками настрою для кожного коментаря, показники вподобань, відповідей та інших взаємодій з коментарями, посилання для перегляду конкретного коментаря на платформі, а також можливість переглянути коментарі-відповіді у вигляді ієрархічної структури

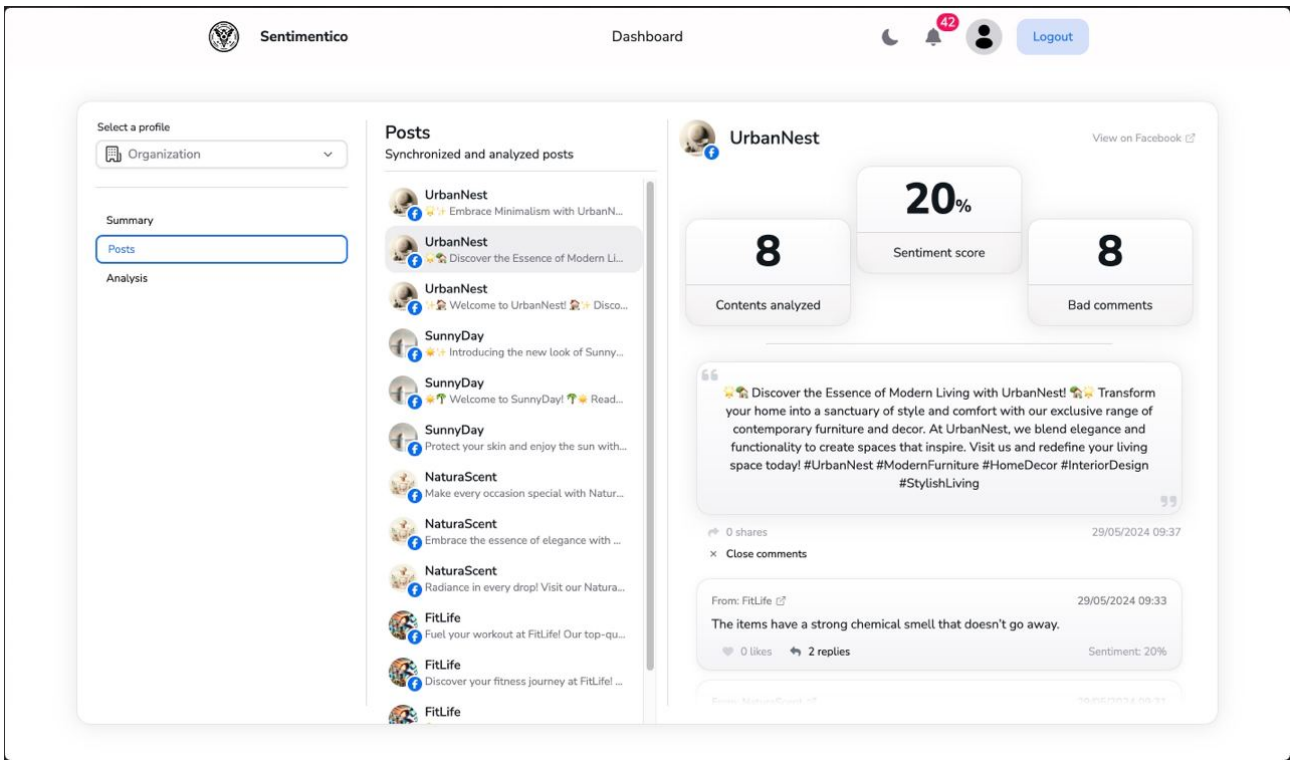


Рисунок 4.7. Сторінка інформаційної панелі з активною вкладкою постів

Ця вкладка надає конкретні оцінки настроїв для кожного поста, що дозволяє користувачам чітко зрозуміти, як їхній контент сприймається аудиторією. Це допомагає оцінити ефективність різних типів постів та контент-стратегій.

Відображаючи кількість проаналізованого контенту та кількість негативних коментарів, користувачі можуть швидко визначити пости, які потребують уваги, або ті, які демонструють хороші результати. Це дозволяє вживати цілеспрямовані дії для зменшення негативних відгуків або використання позитивних реакцій.

Включення секції коментарів з оцінками настроїв, вподобаннями та відповідями дозволяє користувачам безпосередньо взаємодіяти з відгуками своєї аудиторії. Ця взаємодія в режимі реального часу допомагає створювати спільноту та оперативно реагувати на занепокоєння або похвали.

Сторінка консолідує дані з різних платформ соціальних медіа в єдиний інтерфейс, що спрощує процес моніторингу та управління кількома обліковими записами. Цей централізований вигляд економить час і зусилля, роблячи управління соціальними медіа більш ефективним.

Визначаючи негативні коментарі та пости з низькими оцінками настроїв, користувачі можуть проактивно вирішувати потенційні проблеми до того, як вони переростуть у більші. Ця здатність є важливою для підтримання позитивного іміджу бренду та ефективного управління кризовими ситуаціями.

4.7.4. Вкладка аналізу

Вкладка аналізу надає користувачеві вид таблиці на список усіх здійснених аналізів. Користувачі можуть шукати конкретний контент і фільтрувати результати за типом інтеграції, типом контенту та оцінкою настрою. Кожен рядок таблиці має такі поля:

- **Профіль**: профіль соціальних медіа, пов'язаний з контентом.
- **Тип**: тип контенту (наприклад, коментар).
- **Текст**: уривок проаналізованого контенту.
- **LLM**: вказує на мовну модель, використану для аналізу.
- **Оцінка**: оцінка настрою за категоріями (наприклад, Поганий, Негативний, Відмінний).
- **Створено**: час створення контенту.

Задля отримання детальної інформації про повні результати аналізу, здійсненого над конкретним проаналізованим вузлом, користувач може натиснути на рядок таблиці, що в свою чергу відкриє вікно з детальною інформацією, яка надає комплексні метрики, включаючи:

- **Проаналізований текст**: конкретний текст, який був проаналізований.
- **LLM та модель**: деталі мовної моделі, використаної для аналізу.
- **Оцінка настрою**: загальна оцінка настрою у відсотках.
- **Впевненість**: рівень впевненості в аналізі.
- **Ключові слова**: основні терміни, виявлені в аналізі.
- **Сила**: інтенсивність настрою.
- **Суб'єктивність**: ступінь суб'єктивності контенту.
- **Негативні частки**: конкретні негативні терміни або фрази.
- **Розподіл**: розподіл настрою (Нейтральний, Негативний).

- **Оцінки емоцій**: оцінки конкретних емоцій, таких як гнів, сум, радість тощо.

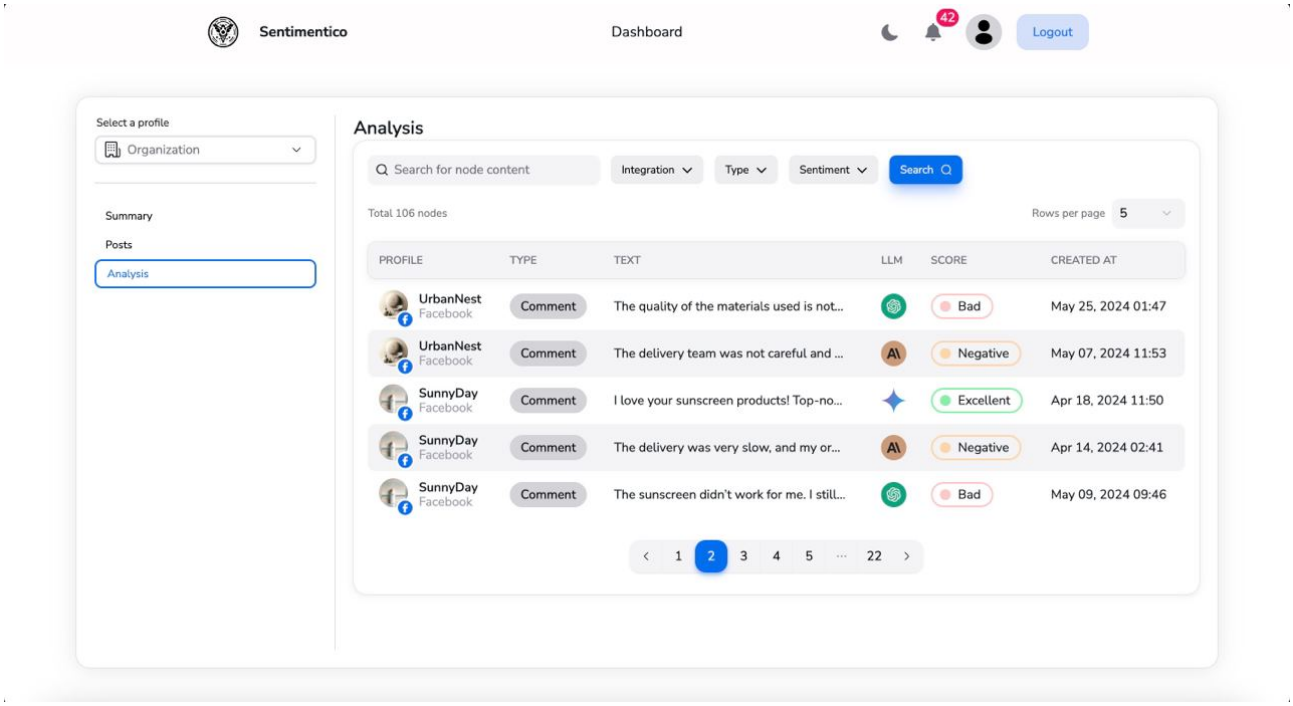


Рисунок 4.8. Сторінка інформаційної панелі з активною вкладкою аналізу

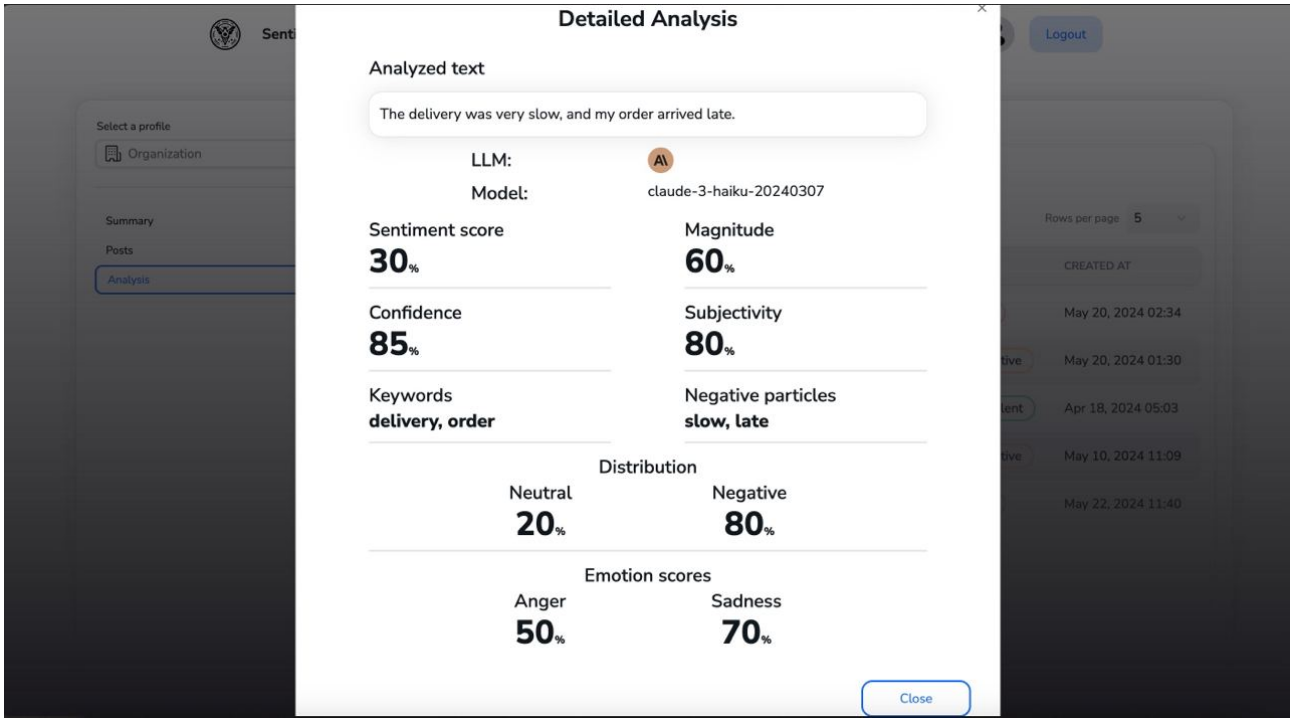


Рисунок 4.9. Сторінка інформаційної панелі з активною вкладкою аналізу та відкритим вікном детальної інформації аналізу

Сторінка аналізу надає детальні оцінки настроїв та рівні впевненості, що дозволяє користувачам чітко розуміти, як їхній контент сприймається. Ця деталізація допомагає виявити конкретні області для покращення або успіху.

Детальні метрики аналізу, включаючи ключові слова, суб'єктивність та оцінки емоцій, надають глибокі інсайти в нюанси відгуків аудиторії. Це допомагає користувачам зрозуміти не лише загальний настрій, але й підхід до емоцій та тем у контенті.

Можливість шукати та фільтрувати контент дозволяє користувачам швидко знаходити та аналізувати конкретні пости або коментарі.

Визначаючи негативні частки та надаючи оцінки емоцій, аналіз допомагає користувачам точно визначити проблемні місця або можливості. Ця інформація може бути використана для створення цілеспрямованих відповідей або коригування контент-стратегій.

Інсайти, отримані з детального аналізу, можуть інформувати стратегічні рішення щодо створення контенту, взаємодії з клієнтами та управління репутацією. Розуміння настроїв та відгуків аудиторії на глибокому рівні дозволяє користувачам приймати обґрунтовані рішення щодо соціальних медіа стратегій.

4.7.5. Оповіщення про негативний сентимент

На навігаційній панелі вгорі знаходиться кнопка виклику, яка забезпечує візуальну підказку з червоною позначкою, що вказує на кількість непрочитаних сповіщень, гарантуючи, що користувачі знають про наявність нових сповіщень. Натискання на цю кнопку здійснює виклик випадаючого меню, в якому відображаються останні сповіщення. Це меню легко доступне з будь-якої сторінки порталу.

Кожне сповіщення містить часову мітку, що вказує, коли було отримано відгук або подію, допомагаючи користувачам зрозуміти актуальність інформації. Текст інформує користувачів про характер відгуку, дозволяючи швидко визначити критичні питання. Також на панелі сповіщення знаходиться аналізована оцінка настрою у відсотках, надаючи миттєве уявлення про серйозність відгуку. Кожне сповіщення містить посилання, який користувачі

можуть натиснути, щоб переглянути перейти на цей конкретний контент на платформі, а також кнопку, яка дозволяє користувачам відзначати сповіщення як прочитані, допомагаючи їм керувати та організовувати свої сповіщення.

Функція сповіщень забезпечує отримання користувачами оновлень у реальному часі про аналіз настроїв та відгуки. Ця своєчасність є вирішальною для оперативного втручання та підтримання позитивної присутності в соціальних медіа.

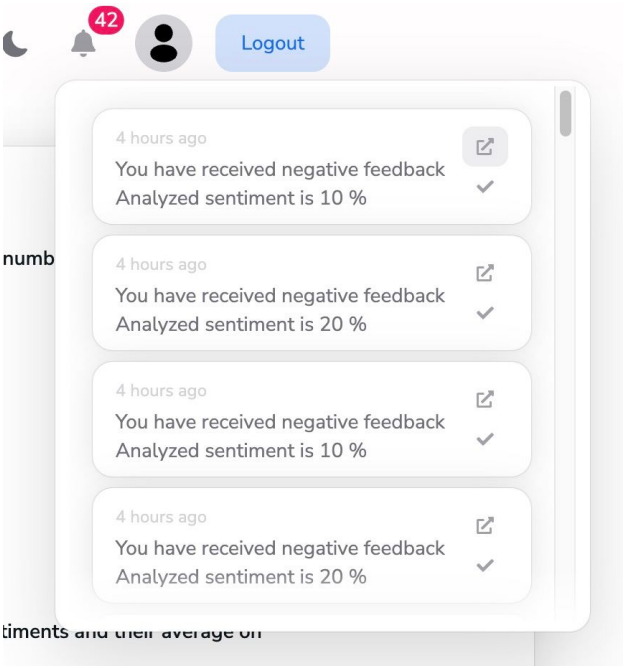


Рисунок 4.10. Випадаюче меню зі списком сповіщень про негативний відгук

Функція сповіщень додає значну цінність до користувацького порталу, надаючи своєчасні та релевантні оновлення, що допомагають користувачам ефективніше керувати своєю присутністю в соціальних медіа.

Отримуючи миттєві сповіщення про негативні відгуки та оцінки настроїв, користувачі можуть проактивно вирішувати питання до того, як вони стануть серйозними. Ця здатність є важливою для підтримання позитивного іміджу бренду та конструктивної взаємодії з аудиторією. Результати аналізу настроїв у реальному часі надають користувачам дієві інсайти, що дозволяють їм приймати обґрунтовані рішення щодо своїх стратегій у соціальних медіа

4.8. Висновки до розділу 4

У цьому розділі було детально описано розроблену систему-прототип для аналізу та управління соціальними медіа. Було розглянуто основні компоненти системи, включаючи структуру даних, архітектуру, використання платформи Hasura, інтеграційні сервіси для Facebook та Instagram, сервіси-аналізатори GPT, Claude та Gemini, а також механізми роботи черги повідомлень RabbitMQ. Окрему увагу було приділено користувацькому порталу, його інтерфейсу та функціональним можливостям. Беручи до уваги те, що на поточний момент система є прототипом і буде поступово покращуватися в майбутньому, вона демонструє високу ефективність у зборі, обробці та аналізі даних, надаючи користувачам потужні інструменти для покращення їхньої присутності в соціальних мережах та управління взаємодією з аудиторією.

ВИСНОВКИ

Ця магістерська робота досліджує розробку та впровадження прототипу програмного рішення для соціального прослуховування, що допомагає компаніям ефективно відстежувати та керувати репутацією свого бренду в інтернеті за допомогою аналізу соціальних мереж і відстеження настроїв. У результаті дослідження було досягнуто кілька важливих висновків, які підтверджують ефективність запропонованого підходу та значущість розробленого прототипу.

Основна мета дослідження полягала у створенні системи-прототипу для моніторингу соціальних мереж Facebook та Instagram, яка б включала засоби для здійснення сентимент-аналізу контенту та надання аналітичної інформації для управління репутацією бренду. В результаті роботи було успішно розроблено та протестовано таку систему, яка показала високу ефективність у реальному часі.

Було досліджено сучасні методи та інструменти управління репутацією бренду в соціальних мережах, розглянуто різні підходи до аналізу тональності тексту, включаючи методи на основі правил, методи машинного навчання та гібридні підходи, а також проведено детальний аналіз можливостей моніторингу соціальних мереж на прикладі Facebook, Instagram, Twitter (X) та TikTok.

Було створено прототип системи для моніторингу соціальних мереж та здійснення сентимент-аналізу контенту. Прототип включає різні компоненти, такі як модулі інтеграції з соціальними мережами, інструменти для збору та обробки даних, а також аналітичні модулі для аналізу тональності тексту. Також було проведено тестування системи на наближених до реальності даних, що дозволило оцінити її ефективність у виявленні настроїв та генеруванні аналітичних звітів. Результати показали високу точність та надійність роботи системи.

Результати цього дослідження мають значну практичну значущість для маркетологів та спеціалістів з управління репутацією, оскільки розроблена система дозволяє ефективно відстежувати настрої аудиторії, оперативно реагувати на зміни в репутації та оптимізувати комунікаційні стратегії. Також

система допомагає бізнесам та брендам підтримувати позитивний імідж, підвищувати конкурентоспроможність та покращувати взаємодію з клієнтами, а для дослідників та розробників інструменти та методології, розроблені в рамках цього дослідження, можуть бути використані для подальших наукових досліджень та розробок у галузі керування репутацією, аналізу даних та обробки природної мови.

Подальші дослідження можуть бути спрямовані на розширення функціональних можливостей системи, включаючи інтеграцію з новими соціальними платформами, покращення алгоритмів аналізу тональності тексту, а також розробку додаткових інструментів для антикризового управління та глибшого аналізу аудиторії.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Barnhart B. Social media reputation management: how to stay in good standing with your audience. Sprout Social. URL: https://sproutsocial.com/insights/social-media-reputation-management/ (дата звернення: 15.03.2024).
2. Szwajca D. The role of social media in corporate reputation management – the results of the polish enterprises. Foundations of management. 2017. T. 9, № 1. C. 161–174. URL: https://doi.org/10.1515/fman-2017-0013 (дата звернення: 15.03.2024).
3. Reputation management through social networking platforms for PR purposes: A SEM-based study in the Jordan / F. Al Olaimat та ін. Frontiers in communication. 2022. T. 7. URL: https://doi.org/10.3389/fcomm.2022.1009359 (дата звернення: 19.03.2024).
4. McLachlan S. Brand reputation management: 10 strategies to avoid damage. Social Media Marketing & Management Dashboard. URL: https://blog.hootsuite.com/brand-reputation/ (дата звернення: 20.03.2024).
5. ReviewTrackers. 2022 online reviews statistics and trends: a report by reviewtrackers. ReviewTrackers. URL: https://www.reviewtrackers.com/reports/online-reviews-survey/ (дата звернення: 20.03.2024).
6. Sprout Social. Crisis communications: what it is and examples brands can learn from. Sprout Social. URL: https://sproutsocial.com/insights/crisis-communication-examples/ (дата звернення: 24.03.2024).
7. Sharma H. D., Goyal P. An analysis of sentiment: methods, applications, and challenges. RAiSE-2023. Basel Switzerland, 2023. URL: https://doi.org/10.3390/engproc2023059068 (дата звернення: 26.03.2024).
8. MonkeyLearn. Sentiment analysis guide. MonkeyLearn. URL: https://monkeylearn.com/sentiment-analysis/ (дата звернення: 26.03.2024).

9. Krugmann J. O., Hartmann J. Sentiment analysis in the age of generative AI. Customer needs and solutions. 2024. Т. 11, № 1. URL: https://doi.org/10.1007/s40547-024-00143-4 (дата звернення: 30.03.2024).
10. Meta. Overview - Graph API. Meta for Developers. URL: https://developers.facebook.com/docs/graph-api/overview (дата звернення: 14.04.2024).
11. X. Getting access to the Twitter API. Use Cases, Tutorials, & Documentation | Twitter Developer Platform. URL: https://developer.x.com/en/docs/twitter-api/getting-started/getting-access-to-the-twitter-api (дата звернення: 18.04.2024).
12. TikTok. TikTok for Developers. TikTok for Developers. URL: https://developers.tiktok.com/doc/about-research-api/ (дата звернення: 20.04.2024).