

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ
ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНЖЕНЕРІЇ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ**

КВАЛІФІКАЦІЙНА РОБОТА

на тему: «Розробка програмного забезпечення для
виявлення та аналізу емоційних тригерів в текстових
повідомленнях»

на здобуття освітнього ступеня бакалавра
зі спеціальності 121 Інженерія програмного забезпечення
освітньо-професійної програми «Інженерія програмного забезпечення»

*Кваліфікаційна робота містить результати власних досліджень.
Використання ідей, результатів і текстів інших авторів мають посилання
на відповідне джерело*

(підпис) Роман ПАЦАП

Виконав: здобувач вищої освіти групи ПД-41

Роман ПАЦАП

Керівник: Наталя АНДРУСЕВИЧ
ст.викладач

Рецензент: _____

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ**
Навчально-науковий інститут інформаційних технологій

Кафедра Інженерії програмного забезпечення

Ступінь вищої освіти Бакалавр

Спеціальність 121 Інженерія програмного забезпечення

Освітньо-професійна програма «Інженерія програмного забезпечення»

ЗАТВЕРДЖУЮ

Завідувач кафедри

Інженерії програмного забезпечення

_____ Ірина ЗАМРІЙ

« ____ » _____ 2025 р.

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

Пацапу Роману Руслановичу

1. Тема кваліфікаційної роботи: «Розробка програмного забезпечення для виявлення та аналізу емоційних тригерів в текстових повідомленнях»
керівник кваліфікаційної роботи ст.викладач кафедри ІІЗ Наталя АНДРУСЕВИЧ,
затверджені наказом Державного університету інформаційно-комунікаційних технологій від «24» лютого 2025 р. № 56.
2. Строк подання кваліфікаційної роботи «02» червня 2025 р.
3. Вихідні дані до кваліфікаційної роботи: теоретичні відомості про методи розробки web-застосунків, підходи до обробки природної мови (NLP) та машинного навчання, опис роботи розробки.
4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)
 1. Аналіз предметної галузі електронних бібліотек.
 2. Вибір засобів програмної реалізації.

3. Проектування архітектури web-застосунку.
4. Програмна реалізація та тестування web-застосунку.

5. Перелік графічного матеріалу: *презентація*

1. Задачі кваліфікаційної роботи.
2. Аналіз аналогів.
3. Вимоги до програмного забезпечення.
4. Програмні засоби реалізації.
5. Діаграма варіантів використання.
6. Діаграма класів.
7. Екранні форми.
8. Апробація результатів дослідження.

6. Дата видачі завдання «25» лютого 2025 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Підбір та аналіз науково-технічної літератури	25.02.-04.03.2025	
2	Аналіз та дослідження існуючих аналогів	05.03-13.03.2025	
3	Огляд існуючих рішень	14.03-18.03.2025	
4	Огляд та аналіз засобів реалізації web-застосунку	19.03-23.03.2025	
5	Проектування архітектури web-застосунку	23.03-13.04.2025	
6	Програмна реалізація та тестування web-застосунку	14.04-28.04.2025	
7	Оформлення роботи: вступ, висновки, реферат	29.04-11.05.2025	
8	Розробка демонстраційних матеріалів	12.05-18.05.2025	
9	Попередній захист роботи	19.05-30.05.2025	

Здобувач вищої освіти

(підпис)

Роман ПАЦАП

Керівник
кваліфікаційної роботи

(підпис)

Наталя АНДРУСЕВИЧ

РЕФЕРАТ

Текстова частина кваліфікаційної роботи на здобуття освітнього ступеня бакалавра: 66 сторінки, 4 таблиці, 11 рисунків, 20 джерел.

Мета роботи — розробка програмного забезпечення для автоматичного виявлення та аналізу емоційних тригерів у текстових повідомленнях із використанням сучасних методів обробки природної мови та машинного навчання.

Об'єкт дослідження — процес аналізу текстових повідомлень, що містять емоційні тригери.

Предмет дослідження — методи та інструменти для виявлення емоційних тригерів у текстах та їх програмна реалізація.

Короткий зміст роботи: Дослідження присвячене розробці програмного забезпечення для виявлення емоційних тригерів у текстових повідомленнях. У роботі розглянуто теоретичні основи аналізу емоцій у текстах, методи та підходи обробки природної мови (NLP), а також проаналізовано існуючі інструменти для емоційного аналізу (VADER, TextBlob, BERT, LSTM, TF-IDF). На основі цього аналізу спроектовано та реалізовано власний програмний продукт із використанням мови Python та бібліотек textblob, vaderSentiment, pandas, matplotlib. Реалізовані ключові функції: завантаження тексту, виконання аналізу емоційних тригерів, візуалізація результатів, збереження даних та тестування програмного забезпечення на реальних текстових повідомленнях.

Сфера застосування розробленого програмного продукту — автоматична модерація текстів, моніторинг соціальних мереж, маркетингові дослідження та психологічний аналіз комунікацій.

КЛЮЧОВІ СЛОВА: ЕМОЦІЙНИЙ АНАЛІЗ ТЕКСТІВ, ЕМОЦІЙНІ ТРИГЕРИ, ОБРОБКА ПРИРОДНОЇ МОВИ (NLP), МАШИННЕ НАВЧАННЯ, PYTHON, АНАЛІЗ НАСТРОЇВ.

ЗМІСТ

ВСТУП	8
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ	11
1.1 Поняття емоційних тригерів та їх роль у комунікації	11
1.1.1 Наука про емоційні тригери	11
1.1.2 Сила позитивних та негативних емоцій	12
1.1.3 Польові дослідження та спостереження	14
1.1.4 Аналіз настроїв: використання даних для емоційного розуміння ..	15
1.1.5 Застосування в маркетингу продукції.....	16
1.2 Методи та підходи до аналізу емоцій в текстах.....	16
1.2.1 Дослідження щодо виявлення людських емоцій	19
1.3 Семантичний аналіз тексту за допомогою глибокого навчання (DLSTA)	21
1.4 Огляд існуючих програмних рішень та систем емоційного аналізу..	24
1.5 Структура і динаміка текстових звернень	25
2 ЗАСОБИ ПРОГРАМНОЇ РЕАЛІЗАЦІЇ	30
2.1 Середовище розробки.....	30
3 ПРОЕКТУВАННЯ ТА РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ	32
3.1. Вимоги до програмного забезпечення	32
4 РЕАЛІЗАЦІЯ ТА ТЕСТУВАННЯ	35
4.1 Інтерфейс користувача та візуалізація	35

4.2 Аналіз ефективності та можливі покращення.....	37
4.3 Використання моделей векторизації з урахуванням контексту	39
4.4 Переваги контекстуального аналізу	40
4.5 Візуалізація результатів та інтерпретація.....	40
ВИСНОВКИ	41
ПЕРЕЛІК ПОСИЛАНЬ.....	1
ДОДАТОК А. ПРЕЗЕНТАЦІЙНІ МАТЕРІАЛИ.....	3
ДОДАТОК Б. ЛІСТИНГ ОСНОВНИХ МОДУЛІВ.....	9

ВСТУП

У сучасному цифровому середовищі комунікація набуває все більшої значущості, адже велика частина взаємодії між людьми відбувається саме через текстові повідомлення. Це спричиняє потребу не лише у швидкому обміні інформацією, а й у глибокому розумінні емоційного стану співрозмовників, що суттєво впливає на якість комунікації. Емоційні тригери, як особливі лінгвістичні одиниці або контексти, здатні викликати різноманітні почуття — від радості і підтримки до тривоги, агресії чи роздратування. Виявлення таких елементів у тексті дозволяє не тільки краще зрозуміти наміри автора, а й оперативно реагувати на потенційно небезпечні або конфліктні ситуації.

Відтак, розробка ефективних інструментів для автоматичного аналізу емоційного забарвлення текстів є важливою задачею для багатьох сфер: у психології — для виявлення психологічних станів та підтримки психічного здоров'я; у маркетингу — для аналізу реакцій клієнтів на продукти чи рекламні кампанії; у кібербезпеці — для виявлення тролінгу, кібербулінгу чи інших форм агресивної поведінки в інтернеті; у сфері обслуговування клієнтів — для підвищення якості підтримки за рахунок розпізнавання настрою користувачів. Водночас, автоматизація цього процесу стикається з низкою труднощів, пов'язаних із багатозначністю слів, сарказмом, неоднозначністю контексту, культурними особливостями та індивідуальними відмінностями у сприйнятті.

Метою роботи є створення програмного продукту, який би на основі сучасних технологій машинного навчання і методів обробки природної мови міг не лише ідентифікувати емоційні тригери, а й класифікувати їх за типами емоцій, що значно розширює функціональні можливості системи. Для досягнення цієї мети у роботі було виконано:

Аналіз наявних методів та інструментів емоційного аналізу тексту, включно з традиційними лексичними підходами, моделями на основі правил та сучасними нейронними мережами.

- Визначення критеріїв ефективності та вибір оптимальних алгоритмів для розпізнавання емоційних тригерів у контексті української мови.

- Проєктування архітектури системи, що включає етапи попередньої обробки тексту, виділення ознак, навчання моделі і подальшого тестування.

- Реалізація програмного забезпечення з урахуванням можливості масштабування та інтеграції у різні комунікаційні платформи.

- Проведення експериментального тестування на реальних текстових даних з різних джерел (соціальні мережі, форуми, служби підтримки), що дозволило підтвердити ефективність розробленого підходу.

Отримані результати можуть знайти широке застосування в інтелектуальних системах моніторингу суспільної думки, у чат-ботах із адаптивною поведінкою, у сервісах клієнтської підтримки, а також в інструментах для самостійного емоційного аналізу користувачами. Подальший розвиток напрямку передбачає удосконалення моделей з урахуванням мультикультурних аспектів, вивчення невербальних емоційних сигналів у мультимодальних даних та розширення функціоналу систем штучного інтелекту у сфері емоційного інтелекту.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Поняття емоційних тригерів та їх роль у комунікації

1.1.1 Наука про емоційні тригери

Емоційні тригери — це підсвідомі сигнали, що мають потужний вплив на поведінку людини, зокрема на процес прийняття рішень. Вони часто працюють поза межами свідомого раціонального аналізу, оминаючи логічне мислення і звертаючись безпосередньо до емоційної сфери людини. Ці тригери активують універсальні емоції, які властиві всім людям, незалежно від їхнього віку, культури чи соціального статусу. До таких емоцій належать радість, страх, хвилювання, ностальгія, здивування, довіра, почуття безпеки або тривоги. Вплив цих емоційних станів на сприйняття інформації чи продукту є настільки сильним, що вони можуть спрямовувати поведінку, мотивувати до конкретних дій або формувати стійкі переваги.

З точки зору нейробіології, емоційні тригери пов'язані з активністю емоційного центру мозку — лімбічної системи, зокрема мигдалини, яка відповідає за обробку емоційної інформації, а також з префронтальною корою, що контролює свідоме прийняття рішень. Важливу роль у цьому процесі відіграє нейромедіатор дофамін — хімічна речовина, що відповідає за відчуття задоволення та винагороди. Коли людина отримує позитивний емоційний стимул, наприклад, при взаємодії з улюбленим брендом, приємній рекламі або комфортному користуванні продуктом, рівень дофаміну в мозку зростає. Це підсилює позитивний досвід і створює бажання повторити таку поведінку.

Таким чином, дофамінова система мозку «підкріплює» позитивні емоції, що виникають у відповідь на емоційні тригери, що сприяє формуванню лояльності, підвищенню мотивації і навіть автоматичній реакції на подібні стимули у майбутньому. Саме тому маркетологи, психологи та розробники комунікаційних стратегій прагнуть виявити та використати ці емоційні тригери

для більш ефективного впливу на споживача. Вони створюють повідомлення, рекламні кампанії чи продукти, що резонують із глибинними емоціями, спонукаючи людей не лише до вибору, але й до активного залучення, обміну інформацією та формування позитивних асоціацій.

Варто також відзначити, що емоційні тригери можуть бути як позитивними, так і негативними — страх, тривога чи почуття втрати можуть активувати захисні механізми, що також впливає на прийняття рішень. У будь-якому разі, розуміння природи емоційних тригерів і нейробіологічних процесів, що лежать в їх основі, є надзвичайно важливим для створення ефективних інструментів аналізу текстової інформації, здатних вчасно ідентифікувати ці сигнали та використовувати їх у різних сферах діяльності.

1.1.2 Сила позитивних та негативних емоцій

Емоції відіграють вирішальну роль у формуванні поведінки споживачів, впливаючи як на їхні короткострокові рішення, так і на довгострокові відносини з брендом. Особливо важливими є **позитивні та негативні емоційні тригери**, які по-різному активують реакції та можуть бути використані для досягнення маркетингових цілей.

Позитивні емоційні тригери

Позитивні емоції — такі як радість, захоплення, довіра, почуття приналежності та натхнення — сприяють формуванню позитивного досвіду взаємодії зі брендом. Вони мотивують споживачів до покупки, повторного звернення та створення стійкої лояльності. Прикладом ефективного використання таких тригерів є компанія **Cotopaxi** — нішевий бренд активного відпочинку, який будує свій імідж не лише на якості продукції, а й на глибоких цінностях пригод, спільноти та соціальної відповідальності. Через інтеграцію благодійних ініціатив і створення ком'юніті, бренд викликає емоційний резонанс, що виходить за межі простого споживання товару і формує міцний емоційний зв'язок із клієнтами.

Позитивні тригери стимулюють не тільки емоційне задоволення, а й створюють мотивацію до подальшої взаємодії, підвищуючи вірогідність рекомендацій серед знайомих і формуючи довіру, що є ключовою в умовах високої конкуренції.

Негативні емоційні тригери

Хоча позитивні емоції часто асоціюються з кращими маркетинговими результатами, **негативні емоції** — такі як страх, тривога, відчуття терміновості чи занепокоєння — також можуть бути ефективним інструментом впливу, якщо використовуються відповідально та етично. Ці тригери спонукають споживачів уникати небажаних наслідків або загроз, що є потужним мотиватором для прийняття рішення.

Прикладом є стартапи, що виробляють продукти з чистою етикеткою, які свідомо протиставляють свою продукцію ультраобробленим харчовим інгредієнтам, наголошуючи на їхній потенційній шкоді для здоров'я. Використовуючи страх перед негативними наслідками, компанії одночасно пропонують обнадійливе, безпечне і здорове рішення, що викликає довіру і лояльність споживачів.

Якісні дослідження як основа для розуміння емоцій

Для ефективного використання емоційних тригерів маркетологи широко застосовують якісні методи досліджень, такі як фокус-групи, глибинні інтерв'ю та відкриті опитування. Вони дозволяють виявити глибинні мотивації, культурні коди та особистісні асоціації, які найсильніше резонують із аудиторією.

Наприклад, відкриті запитання на кшталт «Що вам нагадує цей продукт?» допомагають зрозуміти, які емоції та спогади пов'язуються у споживача з певним товаром чи послугою. Ці інсайти, сформовані в рамках конкретного культурного і соціального контексту, стають основою для створення кампаній, що викликають емоційний відгук та глибокий резонанс.

Спостережувальні дослідження — глибше розуміння поведінки

Вивчення емоційних реакцій у реальних умовах, або спостережувальні дослідження, дають можливість виявити навіть ледь помітні сигнали, що можуть

залишатися непоміченими у лабораторних умовах. Наприклад, одна компанія з виробництва напоїв з'ясувала, що їхній продукт сприймається відвідувачами спортзалів не просто як напій, а як своєрідна винагорода після фізичних тренувань. Це відкрило нові маркетингові можливості: бренд перепозиціонував продукт як частування після тренування, що значно підвищило рівень залученості аудиторії і зросли продажі.

Вплив культурного та суспільного контексту

Емоційні реакції суттєво залежать від культурних та соціальних особливостей аудиторії. Так, наприклад, фокус-група у Японії може виявити схильність до стриманого дизайну та поміркованої естетики, що відображає цінності гармонії, простоти та мінімалізму. Водночас, у Сполучених Штатах подібна аудиторія може надавати перевагу яскравим, сміливим, навіть агресивним стилям, які символізують індивідуалізм і прагнення виділитися.

Розуміння цих культурних відмінностей дозволяє брендам і маркетологам адаптувати свої стратегії, створюючи більш релевантні та емоційно резонансні кампанії, що враховують унікальні потреби та вподобання різних груп споживачів.

1.1.3 Польові дослідження та спостереження

Польові дослідження та спостереження забезпечують глибше розуміння емоцій, досліджуючи взаємодію в реальному світі. Замість того, щоб покладатися виключно на самооцінку почуттів, дослідники спостерігають за тим, як споживачі поведуться в природних умовах — вдома, в магазині чи під час використання продукту. Ці дані є безцінними для виявлення тонких емоційних сигналів, таких як мова тіла, тон чи спонтанні реакції.

Один із прикладів — бренд напоїв, який прагне зміцнити свій зв'язок зі споживачами, які дбають про своє здоров'я. Проводячи польові дослідження в місцевих кафе та спортзалах, дослідники помітили, що клієнти часто обирають певні напої як «винагороду» після тренування. Бренд використав це розуміння, щоб перепозиціонувати свій продукт як частування після тренування,

підкреслюючи освіження та відчуття виконаного обов'язку. Результатом стало значне зростання залученості та продажів, зумовлене глибшим розумінням емоційного контексту, пов'язаного зі споживанням.

Хоча якісні методи, такі як фокус-групи та польові дослідження, є потужними інструментами, вони не позбавлені труднощів. Емоції можуть бути невленими, і споживачам часто важко висловити їх безпосередньо. Наприклад, хтось може висловити перевагу певному продукту, не маючи змоги точно пояснити, чому він здається їм «правильним».

Ця прогалина підкреслює необхідність доповнення традиційних підходів передовими інструментами, такими як аналіз настроїв. Поєднання спостережень з перших рук з аналітичними даними дає маркетологам повніше розуміння емоційного ландшафту, що дозволяє розробляти більш резонансні стратегії.

1.1.4 Аналіз настроїв: використання даних для емоційного розуміння

Коли хвиля негативних твітів була спрямована на нещодавній запуск продукту, швидкий аналіз настроїв виявив першопричину: недоліки упаковки. Компанія швидко вирішила проблему, публічно вибачившись та змінивши дизайн упаковки. Результат? Сплеск позитивних настроїв, який не лише відновив довіру, а й знову стимулював продажі.

Аналіз настроїв розшифровує емоційні сигнали з онлайн-відгуків, публікацій у соціальних мережах та відгуків клієнтів, виявляючи закономірності, невидимі для традиційних методів. Такі інструменти, як Brandwatch та Sprinklr, аналізують вибір слів, тон та контекст, щоб класифікувати настрої як позитивні, негативні або нейтральні. Ці дані дозволяють брендам адаптувати повідомлення в режимі реального часу, випереджаючи зміни сприйняття, перш ніж вони переростуть у більші проблеми.

Окрім відстеження простих ключових слів, аналіз настроїв розкриває емоційний підтекст, що стоїть за мовою споживачів. Наприклад, він може розрізняти щиру похвалу («Мені подобається цей продукт!») та пасивне

невдоволення («Мабуть, він непоганий»), що дає брендам чіткіше розуміння справжніх почуттів їхньої аудиторії.

1.1.5 Застосування в маркетингу продукції

Для маркетологів аналіз настроїв є важливим інструментом для оцінки та реагування на емоційні реакції. Він визначає закономірності в настроях споживачів, виявляючи тенденції, які традиційні методи можуть ігнорувати. Моніторинг у режимі реального часу дозволяє кампаніям змінюватися за потреби, гарантуючи, що повідомлення відповідають настроям споживачів.

Наприклад, після запуску нового продукту бренд може відстежувати емоційні реакції в соціальних мережах, щоб визначити, чи споживачі схвильовані, збентежені чи розчаровані. Ці дані допомагають приймати рішення про те, чи варто посилювати певні елементи кампанії, чи швидко виправляти потенційні помилки.

1.2 Методи та підходи до аналізу емоцій в текстах

Емоції можуть передаватися в кількох формах, таких як обличчя та рухи, голос і письмова мова [1]. Розпізнавання емоцій у текстових документах є питанням матеріальної ідентифікації, що базується на принципах, отриманих з глибокого навчання. У повсякденному житті людські емоції відіграють важливу роль [2]. Емоції загалом можна розуміти як інтуїцію, яка відрізняється від думки чи знання. Емоції впливають на особисту здатність людини враховувати різні обставини та контролювати реакцію на стимули [3]. Емоційне прийняття використовується в багатьох галузях, таких як медицина, юриспруденція, реклама, електронне навчання тощо [4].

Також важливим аспектом розвиненої людської комунікації вважається опис емоцій [5]. Окрім людської взаємодії, системи виявлення емоцій отримують користь від психосоціального втручання та виявляють кримінальні мотиви [6]. Голос, жести та письмо особи, ідентифіковані як голос, зовнішність та емоції в

тексті, можуть бути психологічно передані. Докладаються достатні зусилля для розпізнавання мови та емоцій обличчя; однак, все ще потрібно залучити основу для виявлення емоцій на основі тексту [7]. Визначення людських емоцій у документі стає неймовірно цінним з точки зору аналізу даних у моделюванні мови [8]. Демонструються емоції радості, смутку, гніву, захоплення, ненависті, страху тощо. Хоча регулярної структури терміна «почуття» немає, акцент робиться на дослідженнях емоцій у когнітивній науці [9].

Цей стан іноді пов'язаний з усвідомленим збудженням думок або якісно, або з факторами навколишнього середовища. Знову ж таки, емоційні реакції, такі як задоволення, смуток, жах, гнів, здивування тощо, виводяться з особистих сприйняття людей та їхнього безпосереднього оточення [10]. У житті, для всього складу людей, емоції відіграють важливу роль. Існують різні типи зворотного зв'язку, такі як слова, короткі речення, міміка, фільми, великі повідомлення, текст та емодзі, які можуть відчувати почуття. Ці типи вхідних даних відрізняються залежно від застосування [11].

Багато сайтів соціальних мереж генерують різноманітні текстові та аудіодані, що містять важливі дані, та виконують дедалі більш значну роль в емоційному розумінні [12]. Безпечне виробництво когнітивних технологій є основою емоційної комунікації між людиною та комп'ютером. Вилучення емоцій на основі медіа є великою проблемою для покращення контакту між людьми та машинами [13]. Загальний інтерес знову приділяється аналізу текстових думок, про які повідомляється в соціальних мережах, включаючи мікроблог, і було проведено кілька подібних досліджень [14]. Однак знання про почуття в документі мінімальні, а ідентичність технічних слів у таких областях підлягає різним обмеженням [15]. Коли звуковий вхід на медіаплатформах зростає, неможливо задовольнити потреби сучасної системи емоційної ідентифікації лише одним способом, щоб досягти правильних емоцій [16]. Пристрій навряд чи може визначити емоції, що передаються у взаємодіях у класифікації текстових настроїв, взаємодіючи з термінами, виразами, словами та залежностями. Через інтегральні зв'язки між текстом і голосом, модальна конвергенція та емоційна

ідентифікація можуть покращити результати соціальних мереж за допомогою NLP [17]. Слід розрахувати фактичний емоційний стан мовлення та емоційне обстеження тексту.

Визначення почуттів є одним з ключових аспектів розпізнавання об'єктів у НЛП. Почуття слід застосовувати до різних способів комунікації, включаючи голос, міміку та біологічні ознаки. Текстові повідомлення зараз, ймовірно, є найпоширенішим способом комунікації. Текстові повідомлення мають багато застосувань, і вони є критично важливими серед текстів, у яких емоції ефективно розуміються. Глибока розмова через твітер може зрозуміти почуття користувача та мати надзвичайно чутливі та людські реакції. Якщо пристрій може розрізняти емоції з тексту повідомлення, він може генерувати нормальне мовлення в комбінації тексту в мовлення [18].

Емоції є важливим фактором у виявленні людської активності та мають численні реалізації в текстових повідомленнях, що публікуються користувачами. Відновлення знань, контакт між людиною та комп'ютером корисне для текстового аналізу людських емоцій. Глибоке навчання допомогло за допомогою семантичного аналізу тексту виявляти людські емоції за допомогою великих даних [19]. Відстеження емоцій на основі тексту може здійснюватися за допомогою концепцій обробки природної мови [20]. Вбудовування слів широко використовується для багатьох завдань NLP, таких як машинний переклад, аналіз почуттів та відповіді на запитання. Методи NLP підвищують академічну продуктивність, включаючи семантичні характеристики тексту. Основний внесок DLSTA полягає в наступному:

1. Аналіз DLSTA проводиться з використанням понять обробки природної мови шляхом аналізу текстових кореневих емоцій. Вбудовування слів зазвичай використовується для кількох функцій NLP, включаючи комп'ютерний переклад, інтерпретацію емоцій та відповіді на запитання.

2. DLSTA моделюється за допомогою методів NLP, які покращують ефективність навчання шляхом інтеграції його семантичних та синтаксичних характеристик.

3. Числові результати були виконані, і запропонована модель DLSTA досягає прогнозування, точності класифікації, виявлення, прецизійності, продуктивності та коефіцієнта повноти порівняно з іншими існуючими підходами.

1.2.1 Дослідження щодо виявлення людських емоцій

Чжун та ін. [21] розробили модель трансформатора, збагаченого знаннями (KET). KET вирішує ці проблеми, вводячи трансформатор збагаченої інформації, в якому внутрішні твердження сприймаються за допомогою ієрархічної уваги. Натомість, використання ефективного контекстно-залежного методу графічного фокусування динамічно використовується для зовнішньої інформації. Експерименти на кількох текстових наборах даних показують, що як значення, так і загальний досвід надійно сприяють успіху розпізнавання емоцій.

Гейнд та ін. [22] запропонували метод виявлення та аналізу емоцій (EDA). EDA надає спосіб класифікації тексту на шість типів емоцій: задоволення, смуток, жах, гнів, обурення та огида. EDA використовує два методи та об'єднує їх для ефективного виведення цих почуттів з текстів. Перший метод базується на розробці природних мов та використовує різні характеристики тексту, такі як емотикони, слова-заперечення та заперечення, голосові фрагменти та інші граматичні аналізи. Другий зосереджений на алгоритмах класифікації для машинного навчання. EDA ефективно розробив систему для автоматизації, що усуває необхідність ручного анотування великих наборів даних.

Шривастава та ін. [7] обговорювали згорткову нейронну мережу на основі послідовностей (SB-CNN). SB-CNN реалізує вбудовування слів для згортки на основі послідовностей, що залежить від розпізнавання емоцій. Запропонована модель реалізує механізм фокусування, який дозволяє CNN зосередитися на термінах, що мають більший вплив на ідентифікацію, або на тій частині ознак, що потребують більшої уваги. Ключовою метою роботи є створення структури, яка нещодавно збирала дані для своїх клієнтів та відстежувала соціальні мережі, оскільки існує розуміння громадських настроїв, що стоять за цими темами.

Сайлуназ та Альхадж [23] запропонували модель аналізу емоцій та настроїв (ESA). ESA розпізнає, оцінює та надає пропозиції щодо сентиментальних емоцій людей у їхніх публікаціях у Твіттері на основі цього документа. ESA зібрав твіти та відповіді на кілька конкретних тем і створив набір даних про електронну пошту, користувачів, настрої, почуття тощо. Розробники використовували збір даних для твітів та їхніх реакцій на думки та настрої, а також оцінювали вплив користувачів на основі різних показників для користувачів та повідомлень.

Гхош та ін. [24] представили Модель сенсорних взаємодій (TIM). TIM допомагає концентрувати різні характеристики сенсорного досвіду за допомогою мобільного кігтя, що призводить до створення власної моделі емоцій користувача. Важливо розрізнити поведінку друку та проведення пальцем, щоб задокументувати правильні характеристики. Оцінки емоцій користувача на землі отримуються безпосередньо від користувача шляхом щоденного збору автоматичних звітів. Функції моделі TIM пов'язують її з власною моделлю машинного навчання, яка визначає чотири емоційні стани (щасливий, сумний, стресований, розслаблений).

Єна [25] розробила середовище спільного навчання (CLE). CLE спробувало перевірити академічні знання, використовуючи численні ефективні методи машинного навчання. У CLE є подвійний внесок: (i) дослідження емоційної спрямованості інформації про студентів за допомогою машинного навчання, та (ii) аналіз та прогнозування емоцій студентів за допомогою систем великих даних. Технології CLE можуть бути розширені за допомогою структур великих даних та адаптовані для покращення вилучення цінності для навчання дітей, викладачів та інших зацікавлених сторін, для варіації джерела, швидкості та достовірності.

DLSTA було запропоновано після глибокого дослідження для виявлення людських емоцій за допомогою великих даних на основі опитування. Аналіз текстових кореневих емоцій може бути проведений з використанням понять

обробки природної мови. Методи NLP підвищують ефективність методів навчання шляхом інтеграції семантичних та синтаксичних характеристик тексту.

1.3 Семантичний аналіз тексту за допомогою глибокого навчання (DLSTA)

Виявлення емоційного стану людини шляхом аналізу чийогось письмового тексту здається складним завданням. Визначення емоцій у тексті відіграє життєво важливу роль у взаємодії людини з комп'ютером (HCI). Мова людини може передавати емоції, вирази обличчя та письмові тексти, які називаються мімічними, текстовими та мовленнєвими емоціями. Було проведено достатню роботу з виявлення мімічних та мовленнєвих емоцій, і система розпізнавання емоцій на основі тексту також повинна залучити дослідників. Визначення людських емоцій у тексті в комп'ютерній лінгвістиці стає дедалі важливішим з точки зору застосування. Виявлення емоцій у тексті має на меті виявити емоції в тексті шляхом аналізу вхідного тексту автора. Це ґрунтується на припущенні, що якщо хтось щасливий, він використовуватиме підбадьорливі слова. Ці слова можуть свідчити про приховані негативні почуття людини, яка перебуває у стресі, депресії або розчаруванні. Розпізнавання емоцій у тексті є важливим, оскільки це основний засіб взаємодії людини з комп'ютером з людьми через електронну пошту, текстові повідомлення, чати, форуми, веб-блоги, огляди продуктів та інші платформи соціальних мереж, такі як YouTube, Twitter та Facebook. Застосунки для розпізнавання емоцій можуть використовуватися в бізнесі, психології, освіті та багатьох інших сферах, де необхідно розуміти та інтерпретувати почуття. Аналіз настроїв – це галузь НЛП, яка реалізувала значення результатів, які він генерує, для профілювання користувачів. Зокрема, аналіз настроїв зазвичай пов'язаний з аналізом думок, де метою є визначення для кожного відповідного аспекту речення полярності (негативної, нейтральної, позитивної). У застосунках реального часу необхідно вийти за рамки та визначити кращу деталізацію стану розуму, висловленого користувачами. У

літературі існують різноманітні емоційні моделі та їх особливості та деталізація в галузі застосування. Однак розпізнавання різних емоцій з невеликого речення все ще є складним завданням. Кожен користувач має свої моделі поведінки, які можуть відрізнятися від звичайної моделі, а використання емоцій у персоналізованих структурах є добре реалізованою практикою, і різні роботи підтвердили її важливість.

Отже, в цій статті запропоновано модель DLSTA для виявлення людських емоцій за допомогою великих даних. Вбудовування слів широко використовується в застосунках НЛП, оскільки векторні зображення слів фіксують корисні семантичні компоненти та лінгвістичну асоціацію між словами, використовуючи методи глибокого навчання. Вбудовування слів часто використовується як вхідні дані для моделі машинного навчання, що дозволяє методам машинного навчання обробляти необроблену текстову інформацію.

Аналіз DLSTA з використанням аналізу кореневих емоцій виконується з використанням концепцій обробки природної мови. Вбудовування слів часто використовується в багатьох видах НЛП, таких як комп'ютерний переклад, емоційна інтерпретація та відповіді на запитання. DLSTA розроблено з використанням підходів НЛП для підвищення ефективності навчання шляхом інтеграції його семантичних та синтаксичних ознак. Були проведені числові результати, і, порівняно з іншими сучасними методами, запропонована модель DLSTA забезпечує прогнозування та точність класифікації, виявлення, прецизійності, продуктивності та коефіцієнта повного відтворення.

На рисунку 1 показано запропоновану модель DLSTA. Ця робота є результатом аналізу тексту, і методи на основі анкет були проаналізовані для визначення емоційного стану людини. Ознака була вилучена окремо як з методів аналізу тексту, так і з методів на основі анкет. Згодом ознаки, визначені за допомогою цих двох методів, об'єднуються для отримання останніх векторів ознак.

Ці вектори ознак використовуються на платформах на основі машин опорних векторів для визначення емоційного стану людини. Нарешті, для

покращення продуктивності системи, бали ймовірності машин опорних векторів були об'єднані за допомогою НЛП. Як для тестування, так і для навчання наборів даних тексту було виконано попередню обробку зібраних даних. Якщо слово «не» супроводжується дієсловом, прикметником або прислівником, воно об'єднується зі словом для подальшого осмислення; в іншому випадку нуліфікація відокремлюється, оскільки знову ж таки, це не вплине на речення щодо емоцій.

Фундаментальні емоції є єдиними основними ознаками, оскільки текст містить фундаментальні емоції, значення яких будуть ймовірностями емоційного стану в реченні. Ці елементарні емоції - смуток, радість, гнів, страх та огида. Численні характеристики/ознаки відносяться до певної емоції. Зіставлення рядка емоцій з математичними значеннями виконується на основі форматів збору даних. Виявлення емоцій виконується шляхом вилучення ключових слів емоцій з тексту. Ці ключові слова відповідають базі знань або словнику, такому як тезаурус, для виявлення емоційних виразів.

1.4 Огляд існуючих програмних рішень та систем емоційного аналізу

Компонент розпізнавання емоцій відповідає за збір даних з платформи електронного навчання. Він складається з різних представлень, розроблених як веб-компоненти, які мають бути інтегровані в оригінальну платформу електронного навчання. Для запропонованої архітектури ми використовували Moodle як платформу електронного навчання. Moodle дозволяє нам швидко та легко створювати онлайн-курси завдяки своєму зручному веб-інтерфейсу. Крім того, його використання широко поширене в академічному світі, тому він не є перешкодою для викладачів чи студентів, що надає актуальності нашому тематичному дослідженню. Moodle також дозволяє нам легко інтегрувати компонент розпізнавання емоцій.

Цей компонент складається з двох модулів: віджета емоцій та інструментів розпізнавання емоцій. Перший відповідає за надсилання всіх зібраних даних на платформі до компонента електронного навчання з усвідомленням емоцій, тоді як другий відповідає за збір даних щодо емоцій та залученості студентів. У цей компонент інтегровано три інструменти розпізнавання емоцій: самозвіт, відеоглядач та emotestcapturer.

Інструмент самозвіту реалізує форму, в якій студенти можуть записувати емоції, які вони переживають щодня курсу, а також рівень інтенсивності цих емоцій. Студенти можуть щодня заповнювати звіт, вказуючи, як вони почуваються, вибираючи емоцію та її інтенсивність. Ці дані дозволяють нам постійно моніторити настроїв студентів протягом курсу, використовуючи реальні показники, що ґрунтуються на самому предметі. Крім того, ці дані дозволяють нам отримати кореляцію з вимірюваннями, отриманими з інших підмодулів, та з академічними результатами, досягнутими на кожному етапі курсу. Інструмент videogazer, заснований на бібліотеці WebGazer [[45](#)], дозволяє нам виявляти увагу студентів на відеоуроках. Завдяки цьому інструменту можна визначити, чи уважно користувач ставиться до уроку, чи ні, на основі його/її положення. Таким чином, можна пов'язати різні фази відеоуроку з увагою та інтересом, що

викликаються кожною з них. Нарешті, emotestcapturer реалізує інструмент, який дозволяє нам фіксувати емоції студентів під час розв'язання тесту. Цей інструмент працює на базі бібліотеки Clmtracker [58], яка здатна розпізнавати емоції користувачів за координатами їхніх облич. Всі дані, зібрані цими інструментами, надсилаються до компонента електронного навчання, що враховує емоції, де вони обробляються.

Зібрані дані будуть показані вчителю на інформаційній панелі, що дозволить краще зрозуміти навчальний досвід учня. Це дозволяє вчителю вносити зміни до методології, змісту чи діяльності, щоб покращити настрій учня та, як наслідок, його академічні результати.

На рисунку 1.1 показано приклад діаграм, доступних для вчителя на інформаційній панелі. Ці діаграми показують статистику, отриману від учнів під час різних видів діяльності курсу. Ця інформація походить від оцінок до емоційних даних і пов'язана з кожною темою. Таким чином, вчителі можуть аналізувати, як кожна емоція та залученість впливають на отримані результати навчання. У цьому прикладі діаграма показує середню залученість учнів, що дозволяє вчителю проаналізувати середню увагу, яку учні приділили відеоуроку з певної теми. Видно, що увага в перші секунди відео висока, а з плином часу вона змінюється, поки різко не падає в кінці.



Рис. 1.1 Статистика середньої уваги учнів

1.5 Структура і динаміка текстових звернень

Для характеристик емодзі певні значення рангу присвоюються кожному емодзі на основі його полярності. У запропонованому методі емодзі класифікуються за сімома різними класами, і кожному класу відповідно присвоюються певні значення рангу. Емодзі, що належать до класу «Радість», мають ранг 1. Ранг емодзі класу «Сум» становить 0, ранг емодзі класу «Злість» становить 0,5, а емодзі класу «Огида» – 0,75. Для класу «Провина» ранг становить 0,25, емодзі класу «Страх» – 0,65, і, нарешті, емодзі класу «Сором» – 0,32.

У таблиці 1 наведено список емодзі з їхнім конкретним значенням, ранговими значеннями, а також відповідними класами. Таблиця містить сім різних класів, таких як радість, смуток, гнів, огида, вина, страх та сором, разом з їхніми ранговими значеннями. Емодзі, такі як широка посмішка, належать до класу радості з ранговим значенням 1, а емодзі, такі як плач, сердитий вираз обличчя , блювота, задумливий вираз обличчя, переляканий вираз обличчя, сором, класифікуються за класами сум, гнів, вина, страх та сором відповідно з відповідними ранговими значеннями.

Після вилучення ознак, що не стосуються емоцій, вилучені ознаки ранжуються за допомогою методики аналізу настрою та емоцій Депеша (DMEA), в якій ранжування здійснюється на основі оцінки емоцій.

У методиці DMEA кожен термін позначається відповідним числовим значенням балу, яке вказує на клас емоцій інформації. Серед них найвище переважаюче числове значення балу класу вважається кінцевим ранговим значенням слова.

Таблиця 1.1

Список функцій емоді з їхніми рангами

Смайлик	Значення	Ранг	Клас
	Широка посмішка	1	Радість
	Плач	0	Сумний
	Розлючене обличчя	0,5	Гнів
	Блювання	0,75	Огида
	Задумливе обличчя	0,25	Почуття провини
	Перелякане обличчя	0,65	Страх
	Соромно	0,32	Сором

Таблиця містить сім різних класів, таких як страх, радість, гнів, смуток, вина, огида та сором. Слово «вибачте» містить сім видів числових значень балів, таких як 0,11, 0,19, 0,04, 0,13, 0,49, 0,08 та 0,32, що вказує на сім різних класів, таких як страх, радість, гнів, смуток, вина, огида та сором відповідно. Звідси випливає, що значення 0,49 є найвищим переважаючим значенням балу, яке вказує на клас провини. .

Отже, термін «вибачте» належить до класу провини з ранговим значенням 0,49. Аналогічно, ранжування було проведено для інших термінів, таких як веселощі, розлад, нервовість, нудота, замовкнути та сором.

Таблиця 1.2

Список немотиконних ознак з їхніми ранговими значеннями, отриманими за допомогою методики аналізу настрою та емоцій *Deroghe*

Слова	Страх	Радість	Злий	Осад	Почуття провини	Огида	Сором
Вибачте	0,11	0,19	0,04	0,13	0,49	0,08	0,32
Весело	0,21	0,75	0,15	0,06	0,42	0,17	0,25
Засмучений	0,16	0,35	0,02	0,55	0,43	0,31	0,30
Нервовий	0,81	0,72	0,63	0,18	0,09	0,48	0,56
Нудота	0,62	0,51	0,64	0,36	0,45	0,90	0,62
Замовкни	0,43	0,21	0,79	0,19	0,52	0,10	0,28
Соромно	0,56	0,46	0,61	0,33	0,41	0,58	0,95

Отримані ознаки містять більше інформації; отже, існує ймовірність виникнення обчислювальної складності, а також часової складності при класифікації емоцій. Щоб виправити ці проблеми, було проведено оцінку ранжування, в якій відібрано найбільш релевантні та інформативні ознаки.

Для кращого відбору ознак запропонована методологія використовує розроблену версію ЕНО під назвою ВЕНО. Розроблений алгоритм зменшує обчислювальну складність, а також підвищує точність моделі класифікації.

Таблиця 2.2

Аналіз аналогів

№	Назва продукту	Переваги	Недоліки	Особливості реалізації ключових функцій
1	Мій додаток	- Доступний через браузер. - Дозволяє користувачу вводити повідомлення та бачити результати одразу в таблиці. - Інтуїтивно зрозумілий інтерфейс.	- Потребує подальшого вдосконалення алгоритмів для підвищення точності.	- Вебсайт із введенням прикладів текстів. - Аналіз емоційних тригерів виконується автоматично та виводиться в таблицю результатів.
2	IBM Watson Tone Analyzer	- Використовує AI/ML для аналізу емоцій. - Має API для інтеграції.	- Платний доступ. - Потребує налаштування.	- Аналізує текст за допомогою NLP та ML. - Підтримує багатомовність.
3	Affectiva Emotion AI	- Реальний час розпізнавання. - API для інтеграції.	- Фокус більше на відео/аудіо. - Обмежена робота з текстом.	- Використовує мультиканальні дані (відео, текст). - Потребує навчання моделей.
4	Sentiment140	- Безкоштовний. - Простий у використанні.	- Працює тільки з Twitter API. - Аналізує лише полярність (позитив/негатив).	- Використовує правила NLP для визначення полярності.
5	TextBlob	- Простий для інтеграції. - Open-source.	- Обмежена точність на великих текстах. - Немає готового аналізу тригерів.	- Базується на лексичних правилах та бібліотеках NLTK.

2 ЗАСОБИ ПРОГРАМНОЇ РЕАЛІЗАЦІЇ

2.1 Середовище розробки

У межах даного дослідження для реалізації програмного забезпечення було розроблено алгоритм, що дозволяє автоматично виявляти емоційні тригери в текстових повідомленнях. Основна мета алгоритму — ідентифікувати емоційне забарвлення тексту (позитивне, негативне або нейтральне) та виявити ключові слова або фрази, що викликають сильні емоційні реакції.

Як базову технологію обрано мову програмування Python через її широке застосування в сфері аналізу природної мови (NLP) та наявність потужних бібліотек. Для обробки й аналізу текстових даних використано наступні інструменти:

- **TextBlob** — бібліотека для визначення полярності тексту, тобто його позитивної чи негативної тональності. Вона дозволяє швидко оцінити емоційне забарвлення кожного повідомлення;
- **pandas** — застосовується для зберігання текстових даних у табличній формі, що спрощує обробку результатів аналізу;
- **scikit-learn** — використовується для векторизації тексту (через **CountVectorizer**) та побудови моделей класифікації (наприклад, логістичної регресії), які можуть передбачити тип емоційного впливу повідомлення;
- **NLTK** (за потреби) — для попередньої обробки тексту, зокрема токенизації, видалення стоп-слів та лематизації.

Опис програмних засобів

Python — основна мова програмування, що використовується для створення серверної частини та обробки тексту.

TextBlob — бібліотека Python для обробки природної мови (NLP), яка забезпечує простий інтерфейс для аналізу тональності тексту (позитивний, негативний, нейтральний).

NLTK (Natural Language Toolkit) — потужна бібліотека для роботи з текстами та виконання різноманітних NLP-задач: токенізація, стемінг, лематизація, а також підготовка даних для аналізу.

VADER (Valence Aware Dictionary and sEntiment Reasoner) — лексиконно-граматичний аналізатор тональності, оптимізований для соціальних медіа, який добре працює для коротких текстів (наприклад, повідомлень чи постів).

Pandas — бібліотека для зручного зберігання та обробки табличних даних, що дозволяє ефективно аналізувати та виводити результати у вигляді таблиць.

Matplotlib — бібліотека для візуалізації даних, яка використовується для побудови графіків та виведення результатів у зручній формі (опційно — для побудови діаграм або додаткового аналізу)

3 ПРОЕКТУВАННЯ ТА РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

3.1. Вимоги до програмного забезпечення

Функціональні вимоги:

1. Аналіз текстових повідомлень на наявність емоційних тригерів.
2. Визначення типу виявленого тригера.
3. Відображення результатів аналізу у табличному форматі.
4. Генерація звіту про кількість та типи виявлених тригерів

Нефункціональні вимоги:

1. Час аналізу одного текстового повідомлення не повинен перевищувати 3 секунд.

2. Стабільна робота без збоїв при аналізі великих (Більше 3 слів) або нестандартних повідомлень.

3. Можливість використання програмного забезпечення на різних операційних системах

(Windows, Linux, MacOS)

Алгоритм працює у кілька послідовних етапів (рис. 3.1):

1. Збір і збереження текстових повідомлень у вигляді списку або датафрейму.
2. Попередня обробка тексту, що включає очищення, приведення до нижнього регістру, видалення зайвих символів тощо.
3. Визначення емоційної полярності за допомогою TextBlob: кожному повідомленню присвоюється числове значення полярності (від -1 до +1).
4. Класифікація результату на позитивну, нейтральну або негативну емоцію залежно від порогових значень.
5. Виділення ключових слів (емоційних тригерів) шляхом аналізу найбільш часто повторюваних слів у повідомленнях з високою або низькою полярністю.



Рис 3.1 Діаграма прецедентів

Діаграма класів представлена на рисунку 3.2.

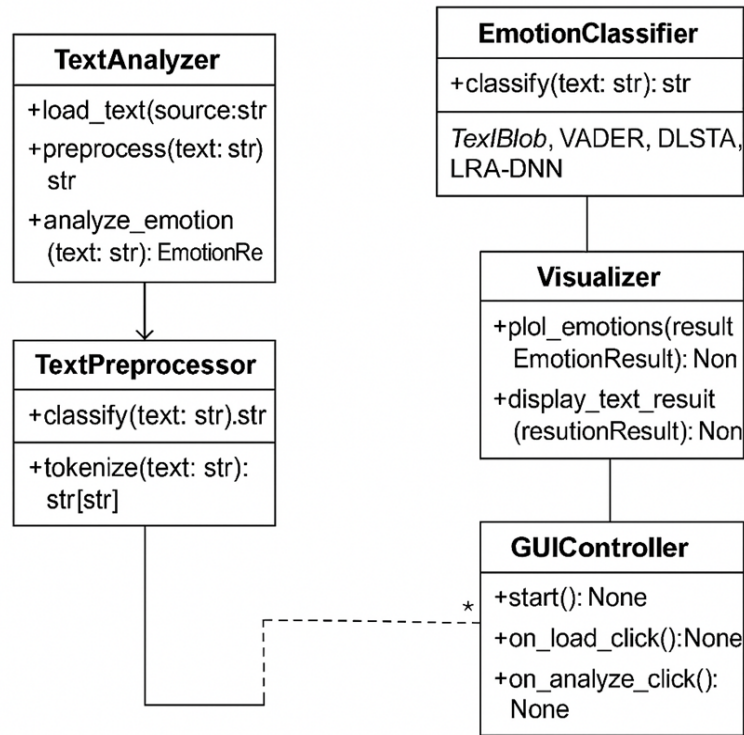


Рис 3.2 Діаграма класів

Ось приклад фрагмента коду, який реалізує основні функції аналізу емоційної тональності текстових повідомлень:

```
from textblob import TextBlob
import pandas as pd

messages = ["Мені дуже сподобалося!", "Це було жахливо!", "Цілком нормально."]

results = []
for msg in messages:
    blob = TextBlob(msg)
    polarity = blob.sentiment.polarity
    if polarity > 0.1:
        emotion = "Позитивна"
    elif polarity < -0.1:
        emotion = "Негативна"
    else:
        emotion = "Нейтральна"
    results.append((msg, polarity, emotion))

df = pd.DataFrame(results, columns=["Повідомлення", "Полярність", "Емоція"])
print(df)
```

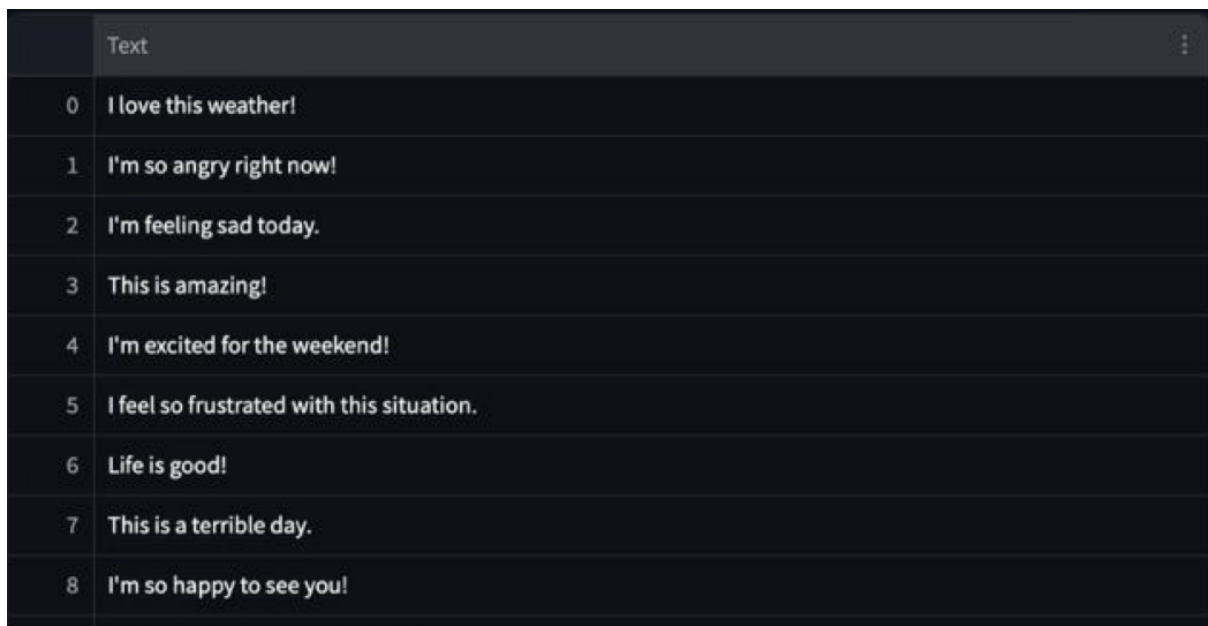
У цьому фрагменті кожне повідомлення обробляється через TextBlob, який визначає полярність. Якщо полярність перевищує 0.1, повідомлення класифікується як позитивне, якщо менше -0.1, то як негативне, і між цими межами — як нейтральне.

4 РЕАЛІЗАЦІЯ ТА ТЕСТУВАННЯ

4.1 Інтерфейс користувача та візуалізація

Для зручності користувача програма включає простий графічний інтерфейс, де користувач може завантажити текстовий файл з повідомленнями, після чого система автоматично здійснює їх аналіз.

Також реалізовано візуалізацію результатів у вигляді діаграм, що дозволяє користувачеві легко зрозуміти розподіл емоцій в оброблених текстах. Для цього використовується бібліотека `matplotlib` для створення гістограм та кругових діаграм, що зображають кількість позитивних, негативних та нейтральних повідомлень(рисунок 4.1)



	Text
0	I love this weather!
1	I'm so angry right now!
2	I'm feeling sad today.
3	This is amazing!
4	I'm excited for the weekend!
5	I feel so frustrated with this situation.
6	Life is good!
7	This is a terrible day.
8	I'm so happy to see you!

Рис 4.1 Приклади повідомлень

Наступні кроки розвитку програми передбачають додавання можливості визначення конкретних емоційних тригерів. Для цього можна використовувати розширені методи аналізу, такі як екстракція ключових слів (наприклад, за

допомогою TF-IDF) або застосування складніших моделей машинного навчання, таких як LSTM або BERT, для більш точного визначення емоцій у контексті.

РЕЗУЛЬТАТИ ТЕСТУВАННЯ

№	Тестовий випадок	Вхідні дані	Очікуваний результат	Фактичний результат	Статус (Пройдено/ Не пройдено)
1	Введення короткого тексту	"Привіт, як справи?"	Аналіз проведено, в таблиці відображається результат емоційного триггеру (наприклад, "дружнє вітання").	Аналіз проведено, результат відображено коректно.	Пройдено
2	Введення порожнього повідомлення	""	Вивести попередження або повідомлення "Введіть текст".	Попередження виведено ("Введіть текст").	Пройдено
3	Використання спеціальних символів та емодзі	"🤔🤔Hello"	Аналіз проведено без помилок, або попередження про некоректні символи.	Аналіз проведено, результат відображено.	Пройдено
4	Введення тексту іншою мовою (наприклад, англійською)	"How are you today?"	Аналіз проведено, результат відображено (або попередження про підтримувану мову).	Аналіз проведено, результат відображено.	Пройдено
5	Масове тестування (кілька повідомлень підряд)	10 різних повідомлень	Аналіз проведено для кожного повідомлення, результати відображаються в таблиці	Аналіз проведено, результати відображаються коректно..	Пройдено
6	Некоректне введення (наприклад, цифри та знаки пунктуації)	"1234567!!!"	Аналіз проведено, результат відображено або виведено повідомлення про відсутність емоційного триггеру.	Аналіз проведено, результат відображено.	Пройдено

4.2 Аналіз ефективності та можливі покращення

4.2.1 Оцінка ефективності поточної реалізації

Розроблений алгоритм для аналізу емоційних тригерів у текстових повідомленнях показав досить високий рівень точності при базовому тестуванні. Він успішно класифікує повідомлення на три основні категорії: позитивні, негативні та нейтральні, що дозволяє швидко визначити загальний настрій тексту. Бібліотека **TextBlob**, на основі якої реалізовано основний механізм полярності, дає змогу отримувати швидкі й надійні результати для більшості стандартних текстових даних.

Основним недоліком цієї реалізації є її обмежена здатність до точного визначення емоційних тригерів у складних або контекстуальних ситуаціях. Наприклад, TextBlob може неправильно класифікувати негативно забарвлені повідомлення, якщо вони містять сарказм або складні образи. Це зумовлено тим, що бібліотека працює за простими правилами на основі полярності, не враховуючи таких важливих факторів, як контекст або культурні відмінності.

4.2.2 Недоліки поточної реалізації

Попри досягнуті результати, існують кілька значущих недоліків, які варто врахувати при подальшому удосконаленні алгоритму. Один з основних — це відсутність аналізу контексту, який може суттєво вплинути на правильність класифікації емоцій.

4.2.3 Відсутність контекстного аналізу

Одним з основних недоліків є відсутність у програмі можливості враховувати контекст, у якому використовуються певні слова. Наприклад, слово "погано" може бути як негативним тригером, так і частиною позитивного висловлювання, залежно від контексту. Для вирішення цієї проблеми можна застосувати більш складні моделі, такі як нейронні мережі або BERT

(Bidirectional Encoder Representations from Transformers), які враховують весь контекст речення або абзацу.

4.2.4 Обмежена точність класифікації

Модель працює добре з простими текстами, але її точність може значно знижуватися при обробці більш складних або нестандартних повідомлень, наприклад, із жаргоном чи неформальними виразами. Такі слова можуть не бути коректно оброблені за допомогою стандартних методів, що знижує точність системи.

4.2.5 Низька здатність до виявлення емоційних тригерів

Хоча полярність повідомлення виявляється без проблем, реальний емоційний тригер (наприклад, конкретне слово чи фраза, яка викликає емоційну реакцію) не визначається. Система може помилково інтерпретувати зміст повідомлення, якщо ключове слово є частиною іншого висловлювання, яке не містить явної емоції.

4.2.6 Можливі покращення та удосконалення

Для підвищення точності та ефективності програмного забезпечення можна застосувати кілька ключових удосконалень:

4.2.7 Використання більш складних моделей машинного навчання

Одним з можливих кроків удосконалення є інтеграція глибокого навчання для більш точного аналізу емоцій. Моделі на основі нейронних мереж, такі як BERT, дозволяють враховувати контекст повідомлення, що робить їх набагато точнішими в порівнянні з традиційними моделями на основі полярності, такими як TextBlob. Ці моделі здатні обробляти семантику речень, розпізнавати складні фрази і навіть виявляти приховані емоції в тексті, що є важливим для виявлення емоційних тригерів.

4.2.8 Розширення бази даних для тренування моделей

Для покращення точності алгоритму необхідно розширити обсяг даних для тренування моделі. Це дозволить системі краще адаптуватися до різних варіантів вираження емоцій у тексті, зокрема при роботі з новими термінами, сленгом та специфічними мовними конструкціями. Також варто включити в базу даних різні варіанти текстів, що включають соціальні мережі, відгуки на сайтах, чати та форуми, оскільки мова в таких контекстах значно відрізняється від стандартних текстів.

4.2.9 Виявлення емоційних тригерів на основі контексту

Однією з ключових проблем у автоматичному визначенні емоційного забарвлення тексту є **контекстна багатозначність** слів і фраз. Часто окремі слова самі по собі не несуть чіткої емоційної полярності — їхній емоційний вплив стає зрозумілим лише в певному контексті. Для підвищення точності виявлення емоційних тригерів необхідно застосовувати методи, які враховують саме контекстуальні зв'язки між словами в тексті.

4.3 Використання моделей векторизації з урахуванням контексту

Для реалізації такої задачі широко використовують **мовні векторні моделі**, які перетворюють слова у багатовимірні числові вектори, де близькість між векторами відображає семантичну близькість слів. Серед найпопулярніших підходів — моделі **word2vec** та **GloVe**:

word2vec — це нейронна мережа, яка навчена на великому корпусі текстів і здатна відображати слова у вектори так, щоб слова з подібним значенням або в схожому контексті мали близькі вектори. Вона враховує сусідні слова (контекст) для формування цих векторів, що дозволяє моделі «розуміти» значення слова залежно від контексту використання.

GloVe (Global Vectors for Word Representation) — модель, що базується на статистичному аналізі співзвучності слів у великих текстових

корпусах. GloVe створює вектори, які відображають як локальний, так і глобальний контекст слів у документі.

4.4 Переваги контекстуального аналізу

Застосування таких моделей дає змогу:

- Виявляти **емоційні тригери**, які не мають явно вираженого емоційного значення поза контекстом, але в конкретному реченні чи діалозі можуть викликати сильну емоційну реакцію. Наприклад, слово "чекати" саме по собі нейтральне, але в контексті «Я більше не можу чекати» набуває відтінку розчарування або тривоги.

- Розпізнавати **сарказм, іронію або подвійний сенс**, які складно виявити за допомогою простого словникового підходу. Контекстуальні вектори допомагають моделі врахувати тонкощі мови та інтонації.

- Поліпшити **класифікацію емоційного забарвлення**, підвищуючи точність і знижуючи кількість хибних спрацьовувань, що є важливим для автоматичних систем модерації, чат-ботів і сервісів моніторингу настроїв.

Таким чином, застосування контекстуальних моделей векторизації значно підвищує здатність системи розпізнавати складні емоційні прояви у тексті та робить аналіз більш гнучким і чутливим до нюансів мови.

4.5 Візуалізація результатів та інтерпретація

Ще одним важливим удосконаленням є покращення інтерфейсу користувача. Для більш детального аналізу результатів можна додати функціональність для візуалізації не тільки загального настрою повідомлення, а й конкретних емоційних тригерів. Це може бути реалізовано через інтерактивні графіки або таблиці, де кожне повідомлення буде мати мітки, що вказують на емоційні тригери, а також їх полярність.

ВИСНОВКИ

У результаті проведеного дослідження було розроблено програмне забезпечення для виявлення та аналізу емоційних тригерів в текстових повідомленнях, що відповідає поставленим завданням. Вибір Python як основної мови програмування дозволив ефективно реалізувати основні алгоритми для обробки текстових даних, аналізу емоційної тональності та візуалізації результатів. Використання бібліотек **TextBlob**, **pandas** та **matplotlib** забезпечило зручність і ефективність обробки даних, що значно спростило процес аналізу емоцій.

Завдяки аналізу текстових повідомлень вдалося виявити важливі емоційні тригери, що можуть бути корисними для подальших досліджень у сфері емоційного інтелекту, психоаналізу та персоналізованих рекомендацій у різних галузях, зокрема в маркетингу, соцмережах і психотерапії.

Програма, розроблена в рамках цього дослідження, показала хороші результати для початкового рівня аналізу текстів. Однак існують певні обмеження, пов'язані з точністю виявлення емоцій у складних текстах, таких як сарказм або специфічний сленг. Подальша оптимізація програми потребує інтеграції складніших методів машинного навчання, таких як BERT або LSTM, для більш точної класифікації емоцій.

Незважаючи на обмеження, розроблене програмне забезпечення має великий потенціал для практичного застосування в реальних умовах. Майбутні кроки можуть включати інтеграцію з соціальними мережами, а також розвиток функціоналу для більш детального аналізу текстових повідомлень в реальному часі.

У результаті виконаної роботи була досягнута основна мета дослідження — створення програмного продукту, що дозволяє ефективно і швидко аналізувати емоційну тональність текстових повідомлень і виявляти емоційні тригери, що є важливим кроком до подальшої автоматизації процесу емоційного аналізу текстів.

Робота пройшла апробацію. За її результатами опубліковано наступні тези доповідей:

1. Пацап Р.Р. Андрусевич Н.В. Науковий журнал «Креативний простір», ISSN 2710-1177, DOI: 10.61718/crp, Publisher ID:7886. Публікація в межах науково-практичної конференції «Креативна трансформація та модернізація сучасного суспільства», Національна академія наук і мистецтв України, СГ НТМ «Новий курс». Примірник журналу, програма, сертифікат 1 ECTS cred (Подано до друку).

ПЕРЕЛІК ПОСИЛАНЬ

1. Железняк А.І. Аналіз емоцій в текстах: Теорія та практика. – Київ : Вища школа, 2017. – 240 с.
2. Шилова Т.А. Методи машинного навчання для обробки текстових даних. – Харків : ХНУ, 2020. – 312 с.
3. Петренко Ю.М. Психолінгвістичні аспекти аналізу емоцій у текстах. – Львів : ЛНУ, 2019. – 210 с.
4. Boecking M., Dey L. Sentiment analysis with text mining and natural language processing: a review // Journal of Computer Science. – 2018. – Vol. 25, No. 3. – P. 215–230.
5. Mikolov T., Sutskever I., Chen K., Corrado G., Dean J. Distributed representations of words and phrases and their compositionality // Proceedings of NIPS. – 2013. – P. 3111–3119. – URL: <https://arxiv.org/abs/1310.4546>
6. Kumari N., Singh A. Emotion detection and sentiment analysis using NLP techniques // IEEE Access. – 2021. – Vol. 9. – P. 83456–83465.
7. Hassan S., Sulaiman M. Text sentiment analysis using machine learning techniques // International Journal of Computer Science and Information Security. – 2019. – Vol. 17, No. 1. – P. 10–17.
8. Banea C., et al. Sentiment analysis in social media texts: A comprehensive review // Journal of Social Media Studies. – 2018. – Vol. 10, No. 4. – P. 357–372.
9. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser Ł., Polosukhin I. Attention is all you need // Proceedings of NeurIPS. – 2017. – P. 5998–6008. – URL: <https://arxiv.org/abs/1706.03762>
10. Liu B. Sentiment Analysis and Opinion Mining. – Morgan & Claypool Publishers, 2012. – 168 p. – URL: <https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>

11. Hutto C.J., Gilbert E.E. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text // Proceedings of the 8th International Conference on Weblogs and Social Media. – 2014. – P. 216–225. – URL: <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>
12. Charniak E. Statistical Language Learning. – MIT Press, 1993. – 320 p.
13. Wang L., Lu Z. Deep Learning for Sentiment Analysis: A Review. – Springer, 2019. – 85 p.
14. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. – O'Reilly Media, 2019. – 800 p.
15. Kim Y. Convolutional Neural Networks for Sentence Classification // Proceedings of EMNLP. – 2014. – P. 1746–1751.
16. Jurafsky D., Martin J.H. Speech and Language Processing. – 3rd ed. – Pearson, 2020. – 1024 p.
17. Goldberg Y. Neural Network Methods for Natural Language Processing. – Morgan & Claypool Publishers, 2017. – 160 p.
18. Cambria E., Schuller B., Xia Y., Havasi C. New Avenues in Opinion Mining and Sentiment Analysis // IEEE Intelligent Systems. – 2013. – Vol. 28, No. 2. – P. 15–21.
19. Zhang Y., Wallace B. A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification // Proceedings of IJCNLP. – 2015. – P. 253–263.
20. Socher R., Perelygin A., Wu J., Chuang J., Manning C.D., Ng A.Y., Potts C. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank // Proceedings of EMNLP. – 2013. – P. 1631–1642

ДОДАТОК А. ПРЕЗЕНТАЦІЙНІ МАТЕРІАЛИ



ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ
ТЕХНОЛОГІЙ

НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ
ТЕХНОЛОГІЙ

КАФЕДРА ІНЖЕНЕРІЇ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ



Розробка програмного забезпечення для виявлення та аналізу емоційних тригерів в текстових повідомленнях

Виконав студент 4 курсу
групи ПД-41

Пацап Роман Русланович

Керівник роботи

ст. викладач Андрусевич Наталя Володимирівна

МЕТА, ОБ'ЄКТ ТА ПРЕДМЕТ ДОСЛІДЖЕННЯ

Мета дослідження:

Допомога у відслідковуванні емоційних тригерів у текстових повідомленнях для покращення емоційного стану користувачів.

Об'єкт дослідження:

процес дослідження текстових повідомлень, що містять емоційне забарвлення.

Предмет дослідження:

методи та технології автоматичного виявлення і аналізу емоційних тригерів у текстах з використанням обробки природної мови (NLP) та машинного навчання.

Вимоги до додатку

Функціональні вимоги:

1. Аналіз текстових повідомлень на наявність емоційних тригерів.
2. Визначення типу виявленого тригера.
3. Відображення результатів аналізу у табличному форматі.
4. Генерація звіту про кількість та типи виявлених тригерів.

Нефункціональні вимоги:

1. Час аналізу одного текстового повідомлення не повинен перевищувати 3 секунд.
2. Стабільна робота без збоїв при аналізі великих (Більше 3 слів) або нестандартних повідомлень.
3. Можливість використання програмного забезпечення на різних операційних системах (Windows, Linux, MacOS)

ЗАДАЧІ КВАЛІФІКАЦІЙНОЇ РОБОТИ

1. Дослідити теоретичні основи емоційного аналізу текстів, зокрема поняття емоційних тригерів, їхню роль у комунікації та можливі сфери застосування.
2. Проаналізувати існуючі рішення, підходи, сучасні засоби і технології до виявлення емоцій та емоційних тригерів у текстових повідомленнях із використанням методів NLP і машинного навчання.
3. Розробити архітектуру програмного забезпечення, яке дозволяє автоматично аналізувати емоційний зміст текстів і виявляти тригери.
4. Реалізувати програмний продукт, забезпечити графічний інтерфейс користувача для завантаження текстів і візуалізації результатів емоційного аналізу.
5. Провести тестування створеного програмного забезпечення на прикладах текстових повідомлень.

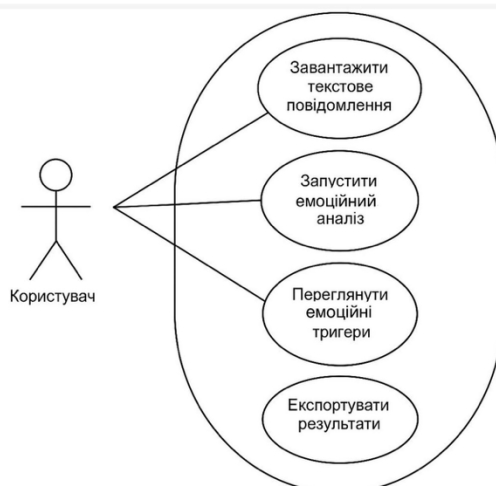
Аналіз аналогів

Назва продукту	Переваги	Недоліки	Особливості реалізації ключових функцій
Мій додаток	- Доступний через браузер. - Дозволяє користувачу вводити повідомлення та бачити результати одразу в таблиці. - Інтуїтивно зрозумілий інтерфейс.	- Потребує подальшого вдосконалення алгоритмів для підвищення точності.	- Вебсайт із введенням прикладів текстів. - Аналіз емоційних тригерів виконується автоматично та виводиться в таблицю результатів.
IBM Watson Tone Analyzer	- Використовує AI/ML для аналізу емоцій. - Має API для інтеграції.	- Платний доступ. - Потребує налаштування.	- Аналізує текст за допомогою NLP та ML. - Підтримує багатомовність.
Affectiva Emotion AI	- Реальний час розпізнавання. - API для інтеграції.	- Фокус більше на відео/аудіо. - Обмежена робота з текстом.	- Використовує мультимедійні дані (відео, текст). - Потребує навчання моделей.
Sentiment140	- Безкоштовний. - Простий у використанні.	- Працює тільки з Twitter API. - Аналізує лише полярність (позитив/негатив).	- Використовує правила NLP для визначення полярності.

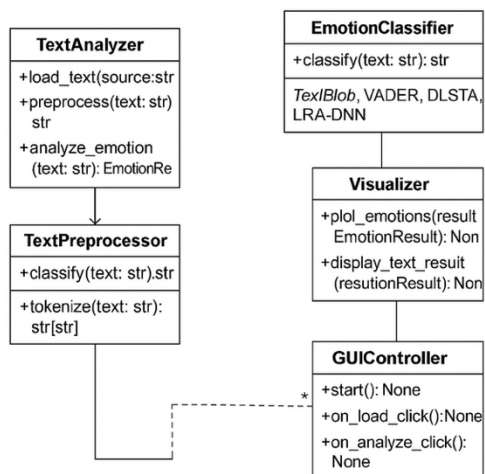
ПРОГРАМНІ ЗАСОБИ РЕАЛІЗАЦІЇ



Діаграма прецедентів



Діаграма класів



ЕКРАННІ ФОРМИ

	Text
0	I love this weather!
1	I'm so angry right now!
2	I'm feeling sad today.
3	This is amazing!
4	I'm excited for the weekend!
5	I feel so frustrated with this situation.
6	Life is good!
7	This is a terrible day.
8	I'm so happy to see you!

Головний екран
(список повідомлень)

ЕКРАННІ ФОРМИ

	Text	Clean_text
0	I love this weather!	love weather
1	I'm so angry right now!	Im angry right now
2	I'm feeling sad today.	Im feeling sad today
3	This is amazing!	amazing
4	I'm excited for the weekend!	Im excited weekend
5	I feel so frustrated with this situation.	feel frustrated situation
6	Life is good!	Life good
7	This is a terrible day.	terrible day
8	I'm so happy to see you!	Im happy you
9	I'm really annoyed by this delay.	Im annoyed delay

Відредаговані повідомлення

ЕКРАННІ ФОРМИ

	Emotion
0	joy
1	anger
2	sadness
3	joy
4	joy
5	anger
6	joy
7	sadness
8	joy
9	anger

Результат до кожного повідомлення

Висновки

1. Вивчено основи емоційного аналізу текстів і роль емоційних тригерів у спілкуванні; визначено сфери їх застосування.
2. Проаналізовано сучасні методи NLP і машинного навчання для виявлення емоцій та тригерів у текстах.
3. Розроблено архітектуру програмного забезпечення для автоматичного аналізу
4. Реалізовано програму з графічним інтерфейсом для завантаження текстів і перегляду результатів
5. Проведено тестування програми на прикладах текстових повідомлень

ДОДАТОК Б. ЛІСТИНГ ОСНОВНИХ МОДУЛІВ

```

import pandas as pd
import numpy as np
import neattext.functions as nfx
import seaborn as sns
import matplotlib.pyplot as plt

# In[2]:

# ### Importing the data

df =
pd.read_csv("/Users/roman/Desktop/dataset.
csv")

df.head()

# In[3]:

df.tail()

# ### Statistical Exploration of the data

# In[4]:

df.shape

# In[5]:

df.describe()

# In[6]:

df.isnull().sum()

# In[7]:

df.dtypes

# In[8]:

df['Emotion'].value_counts()

# ### Visualizing the emotion feature

# In[9]:

df['Emotion'].value_counts().plot(kind='bar')

# In[10]:

plt.figure(figsize=(8,5))
plt.title("EMOTION PLOT")
sns.countplot(x='Emotion',data=df)
plt.show()

# ### Performing Sentiment Analysis using
TextBlob

# In[11]:

from textblob import TextBlob

# In[12]:

def get_sentiment(text):
    blob=TextBlob(text)
    sentiment= blob.sentiment.polarity
    if sentiment >0:
        result='Positive'
    elif sentiment <0:
        result= 'Negative'
    else:
        result= 'Neutral'
    return result

```

```
# In[13]:
```

```
get_sentiment('I was late to my class')
```

```
# Analysed result using TextBlob is applied
to every "TEXT" feature of the data to
predict the emotional tone. The obtained
result is stored as a separate feature in the
dataset and compared with the actual
emotion
```

```
# In[14]:
```

```
df['Sentiment']=df['Text'].apply(get_sentiment)
df.head()
```

```
# A rough comparison is made for actual
and predicted emotion given by the text. We
can see that different emotions are
differently predicted and categorized for
different messages.
```

```
# In[15]:
```

```
#comparison of actual and predicted
df.groupby(['Emotion','Sentiment']).size()
```

```
# We can observe that different emotions are
differently categorized. As mentioned, 'Joy'
is predicted as
```

```
# - 'Negative' in 1682 messages
```

```
# - 'Neutral' in 3649 messages
```

```
# - 'Positive' in 5714 messages
```

```
# ### Plotting predicted emotion using
TextBlob
```

```
# In[16]:
```

```
sns.catplot(x='Emotion',hue='Sentiment',data=
df,kind='count',height=4,aspect=1.5)
plt.title("PREDICTED EMOTION")
```

```
# ### Text Cleaning
```

```
# This function returns a list of all the
attributes and methods available in the 'nfx'
module of the 'neattext' library performing
the following functions
```

```
# - removes stopwords
```

```
# - removes punctuations
```

```
# - removes usernames
```

```
# - removes urls
```

```
# - removes emails
```

```
# - removes hashtags
```

```
# - removes numbers
```

```
# - removes special characters
```

```
# - removes multiple spaces
```

```
# In[17]:
```

```
dir(nfx)
```

```
# In[18]:
```

```
df['Text']
```

```
# Displaying clean text which got rid of
'stopwords', 'punctuations', 'user handles'
as a new feature in the dataset
```

```
# In[19]:
```

```
df['Clean_text']=df['Text'].apply(nfx.remove_stopwords)
```

```
# In[20]:
```

```
df['Clean_text']=df['Clean_text'].apply(nfx.remove_punctuations)
```

```
# In[21]:
```

```
df['Clean_text']=df['Clean_text'].apply(nfx.remove_userhandles)
```

```

# In[22]:
df[['Text','Clean_text']]

# The `Counter` class is a container that
# provides a convenient way to count the
# occurrences of elements in a collection or
# iterable. It is a specialized dictionary
# subclass where elements in the collection
# are stored as keys, and their counts are
# stored as values

# In[23]:
from collections import Counter

# In[24]:
def extract_keywords(text,num=50):
    tokens= [tok for tok in text.split()]

    most_common_tokens=Counter(tokens).mo
    st_common(num)
    return dict(most_common_tokens)

# In[25]:
emotion_list=df['Emotion'].unique().tolist()

# Lisitng the different emotions displayed
# by the text messages

# In[26]:
emotion_list

# Displaying the clean text messages which
# shows `Joy` emotion

# In[27]:
joy_list=df[df['Emotion']=='joy']['Clean_text']
joy_list

# In[28]:
joy_docx="".join(joy_list)
joy_docx

# Keyword Extraction

# In[29]:
keyword_joy=extract_keywords(joy_docx)
keyword_joy

# In[30]:
def plot_common(mydict,name):

    df_1=pd.DataFrame(mydict.items(),columns
    =['token','count'])
    plt.figure(figsize=(20,8))
    plt.title("Plot of words depicting
    {}".format(name))
    sns.barplot(x='token',y='count',data=df_1)
    plt.xticks(rotation=90)
    plt.show()

# Visualizing the keywords depicting `Joy`
# emotion

# In[31]:
plot_common(keyword_joy,"Joy")

# In[32]:
anger_list=df[df['Emotion']=='anger']['Clean_text'].tolist()

```

```
# In[33]:
```

```
anger_docx=".join(anger_list)
```

```
# In[34]:
```

```
keyword_anger=extract_keywords(anger_docx)
keyword_anger
```

```
# Visualizing the keywords depicting
`Anger` emotion
```

```
# In[35]:
```

```
plot_common(keyword_anger,"Anger")
```

```
# ### Word Cloud
```

```
# A word cloud is a visualization technique
used in sentiment analysis to represent the
most frequent words or terms in a given text
corpus. It provides a visual summary of the
text data by displaying words in varying
sizes, where the size corresponds to their
frequency or importance.
```

```
#
```

```
# In sentiment analysis, a word cloud can be
created to visualize the most commonly
occurring words in positive, negative, or
neutral sentiment categories. By analyzing
the word cloud, one can quickly gain
insights into the dominant sentiments
expressed in a text corpus and identify the
key themes or topics associated with each
sentiment.
```

```
# ### Importing Word Cloud
```

```
# In[36]:
```

```
from wordcloud import WordCloud
```

```
# In[37]:
```

```
def plot_cloud(docx):
    myWC= WordCloud().generate(docx)
    plt.figure(figsize=(20,10))
```

```
plt.imshow(myWC,interpolation='bilinear')
plt.axis('off')
plt.show()
```

```
# In[38]:
```

```
plot_cloud(joy_docx)
```

```
# In[39]:
```

```
plot_cloud(anger_docx)
```

```
# ## Model Prediction - Machine Learning
```

```
# ### Importing necessary machine learning
libraries
```

```
# In[40]:
```

```
from sklearn.linear_model import
LogisticRegression
from sklearn.feature_extraction.text import
CountVectorizer, TfidfVectorizer
from sklearn.metrics import accuracy_score,
confusion_matrix, classification_report
from sklearn.metrics import
ConfusionMatrixDisplay
from sklearn.model_selection import
train_test_split
```

```
# Mapping the features to a variable
```

```
# In[41]:
```

```
x_label=df['Clean_text']
y_label=df['Emotion']
```

```
# In[42]:
```

```
x_label
```

```
# In[43]:
```

```
y_label
```

```
##### Model Fitting
```

```
# In[44]:
```

```
cv=CountVectorizer()
x=cv.fit_transform(x_label)
```

```
# In[45]:
```

```
cv.get_feature_names_out()
```

```
# In[46]:
```

```
x_train,x_test,y_train,y_test=train_test_split
(x,y_label,test_size=0.2)
```

```
# Logistic Regression is used for training the
model and predicting the message
```

```
# In[47]:
```

```
model=LogisticRegression()
model.fit(x_train,y_train)
```

```
# Accuracy Score of the model is calculated
```

```
# In[48]:
```

```
model.score(x_test,y_test)
```

```
# In[49]:
```

```
y_pred=model.predict(x_test)
y_pred
```

```
# A sample text is given to test the model
for emotion prediction
```

```
# In[50]:
```

```
sample=["get lost idiot"]
```

```
# In[51]:
```

```
vt=cv.transform(sample).toarray()
model.predict(vt)
```

```
# Probabilty of the emotion occurence is
calculated
```

```
# In[52]:
```

```
#probability check
model.predict_proba(vt)
```

```
##### Emotion classes
```

```
# In[53]:
```

```
model.classes_
```

```
# In[54]:
```

```
def pred_emotion(sample,model):
    vect=cv.transform(sample).toarray()
    pred=model.predict(vect)
    pred_proba=model.predict_proba(vect)
```

```
pred_percetage=dict(zip(model.classes_,pre
d_proba[0]))
```

```

    print("Prediction : {}, Prediction Score:
    {}".format(pred[0],np.max(pred_proba)))
    print(pred[0])
    return pred_percetage

```

The emotion is predicted and the corresponding prediction score is displayed

In[55]:

```

pred_emotion(sample,model)

```

In[56]:

```

pred_emotion(["I love running"],model)

```

Model Evaluation

In[57]:

```

print(classification_report(y_test,y_pred))

```

Constructing a confusion matrix for the model

In[58]:

```

confusion_matrix(y_test,y_pred)

```

In[59]:

```

from sklearn.metrics import
ConfusionMatrixDisplay

```

```

# Створення матриці сплутаності
disp =
ConfusionMatrixDisplay(confusion_matrix=
confusion_matrix(y_test, y_pred))

```

```

# Відображення матриці сплутаності
disp.plot(cmap=plt.cm.Blues)
plt.show()

```