КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
БУДІВНИЦТВА І АРХІТЕКТУРИ

Автоматизації і інформаційних технологій

(факультет)

Інформаційних технологій проектування та прикладної математики

(кафедра)

ПОЯСНЮВАЛЬНА ЗАПИСКА
ДО КВАЛІФІКАЦІЙНОЇ РОБОТИ

НА ЗДОБУТТЯ ОСВІТНЬОГО РІВНЯ «БАКАЛАВР»

на тему: «Розробка нейромережевої системи аналізу тональності текстових відгуків»

ГРАФ ЄВГЕН СЕРГІЙОВИЧ
(прізвище, ім’я та по батькові студента повністю)

Київ 2026 р

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
БУДІВНИЦТВА І АРХІТЕКТУРИ

Автоматизації і інформаційних технологій

(факультет)

Інформаційних технологій проектування та прикладної математики

(кафедра)

ЗАТВЕРДЖУЮ
Завідувач кафедри ІТППМ
д.т.н., професор Бородавка Є.В.

_____

„___” _____2026 року

ПОЯСНЮВАЛЬНА ЗАПИСКА
ДО КВАЛІФІКАЦІЙНОЇ РОБОТИ
НА ЗДОБУТТЯ ОСВІТНЬОГО РІВНЯ «БАКАЛАВР»

на тему: «Розробка нейромережевої системи аналізу тональності текстових відгуків»

Виконав: студент 4-го курсу, групи ІСТ-22

Спеціальності: 126 «Інформаційні системи та технології»
(шифр і назва спеціальності)

Граф Є.С.
(прізвище та ініціали)

Керівник д.т.н., професор Бородавка Є.В.
(прізвище та ініціали)

Рецензент_
(прізвище та ініціали)

Київ, 2026 р.

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
БУДІВНИЦТВА І АРХІТЕКТУРИ

Факультет: автоматизації і інформаційних технологій
Кафедра: інформаційних технологій проєктування та ПМ
Освітній рівень: «бакалавр» за ОПП
Спеціальність: 126 «Інформаційні системи та технології»
Освітня програма: «Інформаційні системи та технології»

ЗАТВЕРДЖУЮ
Завідувач кафедри ІТППМ
д.т.н., професор Бородавка Є.В.
_____
„___” _____2026 року

З А В Д А Н Н Я
ДО ВИКОНАННЯ КВАЛІФІКАЦІЙНОЇ РОБОТИ
НА ЗДОБУТТЯ ОСВІТНЬОГО РІВНЯ «БАКАЛАВР»

Граф Євген сергійович

1. Тема роботи: Розробка нейромережевої системи аналізу тональності текстових відгуків. Затверджена наказом ректора КНУБА № 444/52-15/13.6/26 від “08” квітня 2026 року.
2. Керівник роботи: Бородавка Євгеній Володимирович, д.т.н, професор . кафедри інформаційних технологій проектування і прикладної математики.
3. Строк подання студентом роботи до захисту: червень 2026 року.
4. Зміст пояснювальної записки за розділами:
    a) Аналіз предметної області та постановка задачі
    b) Інформаційне та математичне забезпечення нейромережевої системи
    c) Проєктні рішення та розробка програмного забезпечення
    d) Ергономіка інформаційних технологій
5. Інформаційні слайди:

С.1.Аналіз тональності тексту та технології NLP

С.2. Методи та архітектури нейронних мереж для Sentiment Analysis

С.3. Розробка нейромережевої системи аналізу тональності текстових відгуків

С.4. Попередня обробка текстових даних

С.5. Архітектура моделі BiLSTM

С.6. Структура програмного забезпечення системи

С.7. Інтерфейс та результати роботи програми

6. Календарний план виконання кваліфікаційної роботи бакалавра

Види роботи та їх зміст	Дата виконання
Р. 1. Аналіз предметної області та постановка задачі	Лютий 2026
Р. 2. Інформаційне та математичне забезпечення нейромережевої системи	Квітень 2026
Р. 3. Проєктні рішення та розробка програмного забезпечення	Травень 2026
Р. 4. Ергономіка інформаційних технологій	Травень 2026
Остаточне оформлення роботи	Травень 2026
Направлення роботи на рецензування	Червень 2026
Попередній захист роботи на кафедрі	Червень 2026

7. Консультанти розділів кваліфікаційної роботи бакалавра

Розділ	Прізваще, ініціали та посада	Дата	Підпис

	консультанта, представника комісії		
Прийом програмного продукту	д.т.н., професор Бородавка Є.В		

8. Дата видачі завдання: 12 січня 2026 р.

Керівник _____ Бородавка Є.В.
(підпис) (прізвище та ініціали)

Бакалавр _____ Граф Є.С.
(підпис) (прізвище та ініціали)

Анотація

«Розробка нейромережевої системи аналізу тональності текстових відгуків».

Кваліфікаційна робота бакалавра за спеціальністю: 126 «Інформаційні системи та технології», освітня програма: «Інформаційні системи та технології». – Київський національний університет будівництва та архітектури. – Київ, 2026.

Важливою проблемою є аналіз великих обсягів текстових відгуків, коментарів та повідомлень, оскільки ручна обробка текстової інформації займає значну кількість часу, потребує уважності та не забезпечує достатньої ефективності при роботі з великими масивами даних. Саме тому актуальним є використання сучасних методів обробки природної мови та нейромережевих технологій для автоматизації аналізу тональності тексту.

Ключові слова: аналіз тональності, NLP, нейронні мережі, BiLSTM, машинне навчання, текстові відгуки, класифікація тексту, АТВ – аналіз тональності відгуків.

Summary

“Development of a Neural Network System for Sentiment Analysis of Text Reviews”.

Bachelor’s qualification thesis in specialty 126 “Information Systems and Technologies”, educational program “Information Systems and Technologies”. – Kyiv National University of Construction and Architecture. – Kyiv, 2026.

An important problem is the analysis of large volumes of text reviews, comments, and messages, since manual processing of textual information requires significant time, demands a high level of attention, and does not provide sufficient efficiency when working with large datasets. Therefore, the use of modern natural language processing methods and neural network technologies for automating sentiment analysis is highly relevant.

Keywords: sentiment analysis, NLP, neural networks, BiLSTM, machine learning, text reviews, text classification, SAR – sentiment analysis of reviews.

Зміст роботи:

Висновки

Список джерел та літератури

Додаток

Вступ

У сучасних умовах стрімкого розвитку інформаційних технологій та цифрових комунікацій значна частина взаємодії між користувачами та цифровими сервісами відбувається у мережі Інтернет. Користувачі активно залишають текстові відгуки щодо товарів, послуг, програмного забезпечення та онлайн-платформ, формуючи великі обсяги неструктурованих текстових даних. Аналіз таких повідомлень дозволяє отримувати цінну інформацію про ставлення користувачів до певного продукту, виявляти переваги та недоліки сервісів, а також оцінювати рівень задоволеності клієнтів. У зв’язку зі стрімким зростанням кількості текстової інформації виникає необхідність автоматизації процесів її обробки та аналізу.

Одним із перспективних напрямів обробки природної мови є аналіз тональності тексту, який дозволяє автоматично визначати емоційне забарвлення повідомлень та класифікувати їх як позитивні, негативні або нейтральні. Системи аналізу тональності широко застосовуються у маркетингових дослідженнях, соціальних мережах, електронній комерції та інформаційно-аналітичних сервісах. Особливої актуальності дана задача набуває в умовах активного розвитку цифрових платформ та онлайн-сервісів, де швидкий аналіз відгуків дозволяє оперативно реагувати на потреби користувачів та підвищувати якість продуктів.

Сучасні методи аналізу текстових даних дедалі частіше базуються на використанні алгоритмів машинного навчання та нейромережевих технологій. Одним із найбільш ефективних підходів до обробки текстової інформації є використання рекурентних нейронних мереж, зокрема архітектури BiLSTM, яка дозволяє враховувати контекст слів та аналізувати текст у двох напрямках.

Застосування таких моделей забезпечує підвищення точності класифікації текстових повідомлень та ефективну роботу з великими обсягами даних.

Актуальність теми дипломної роботи обумовлена необхідністю створення сучасних програмних засобів автоматизованого аналізу текстових відгуків із використанням технологій штучного інтелекту та обробки природної мови. Використання нейромережевих методів дозволяє підвищити швидкість аналізу текстової інформації та зменшити вплив людського фактору під час оцінювання повідомлень. [1], [11]

Метою дипломної роботи є розробка нейромережевої системи аналізу текстових відгуків із використанням моделі BiLSTM та засобів автоматизованої обробки природної мови.

Об’єктом дослідження є процес автоматизованого аналізу текстових даних у системах обробки природної мови.

Предметом дослідження є методи та програмні засоби аналізу тональності текстових відгуків із використанням нейромережевих технологій.

Під час виконання роботи були використані методи машинного навчання, обробки природної мови, нейромережевого моделювання та програмної інженерії. Для реалізації програмного забезпечення використовувалися мова програмування Python, бібліотеки PyTorch, Pandas, Matplotlib, а також фреймворк Streamlit для створення вебінтерфейсу користувача.

Практична цінність отриманих результатів полягає у створенні програмної системи, здатної автоматично аналізувати текстові відгуки користувачів та визначати їх емоційне забарвлення. Розроблена система може бути використана для аналізу користувацьких повідомлень у цифрових сервісах, соціальних мережах, інтернет-магазинах та ігрових платформах.

1 Аналіз предметної області та постановка задачі

1.1 Опис предметної області

На сьогоднішній день в інформаційному суспільстві постійно зростає обсяг текстових даних, що генеруються користувачами в цифровому середовищі. Відгуки клієнтів, коментарі в соціальних мережах, рецензії на товари та послуги, повідомлення на форумах і платформах утворюють значні масиви неструктурованої текстової інформації. Аналіз таких даних дозволяє отримувати цінні відомості про думки, емоції та ставлення користувачів, що є важливим чинником для прийняття управлінських та маркетингових рішень.

Одним із ключових напрямів обробки текстових даних є аналіз тональності відгуків (АТВ), який полягає у визначенні емоційного забарвлення тексту. Залежно від поставленої задачі тональність можна класифікувати як позитивну, негативну та нейтральну. У більш складних випадках можуть бути включені значення ступеня емоційної вираженості.

Предметна область АТВ належить до галузі обробки природної мови (англ. Nature Language Processing, або NLP) та тісно пов’язана з методами штучного машинного навчання та штучних нейронних мереж. Основною особливістю цієї області є робота з неструктурованими текстами, які характеризуються складною лінгвістичною природою, багатозначністю слів, контекстною залежністю, наявністю сленгу, скорочень та граматичних помилок. Це ускладнює процес автоматичної обробки та вимагає застосування спеціалізованих алгоритмів та моделей. [1], [5], [16]

Текстові відгуки як об’єкт аналізу можуть суттєво відрізнятися за довжиною, стилем викладу та мовою написання. Короткі коментарі часто містять емоційно забарвлені висловлювання без чіткої граматичної структури, тоді як розгорнуті відгуки можуть включати складні речення, порівняння та суперечливі оцінки. Крім того, часто зустрічаються іронія, сарказм та приховані емоції, що додатково ускладнює задачу конкретного визначення тональності.

Процес аналізу тональності текстових відгуків зазвичай складається з кількох взаємопов’язаних етапів. На першому етапі здійснюється збір текстових даних із відповідних джерел та формування корпусу відгуків. Далі виконується

попередня обробка тексту, яка включає очищення від спеціальних символів, нормалізацію регістру, токенізацію, видалення стоп-слів, а також лематизацію та стемінг. Стемінг та лематизація - це процеси обробки тексту, які зводять різні форми слова до одного базового варіанту, щоб комп'ютеру було легше їх аналізувати. Метою цього етапу є приведення тексту до уніфікованого вигляду, придатного для подальшого аналізу. [11], [12]

Наступним етапом є подання текстових даних у формі числових ознак, що можуть бути використані алгоритмами машинного навчання. Для цього застосуються різні методи векторизації тексту, які дозволяють відобразити сематичні та статистичні характеристики слів і фраз. Сематичні характеристики - це сукупність змістовних ознак слів, виразів або понять, які визначають їхнє значення, функції та відношення з іншими мовними одиницями. Завершальним етапом є класифікація тональності, яка здійснюється з використанням обраної моделі та дозволяє віднести кожен текстовий відгук до окремої категорії.

На Рис 1.1 наведено узагальнену схему процесу аналізу тональності текстових відгуків, що показує основні етапи обробки даних – від надходження вхідного тексту до отримання результату класифікації.

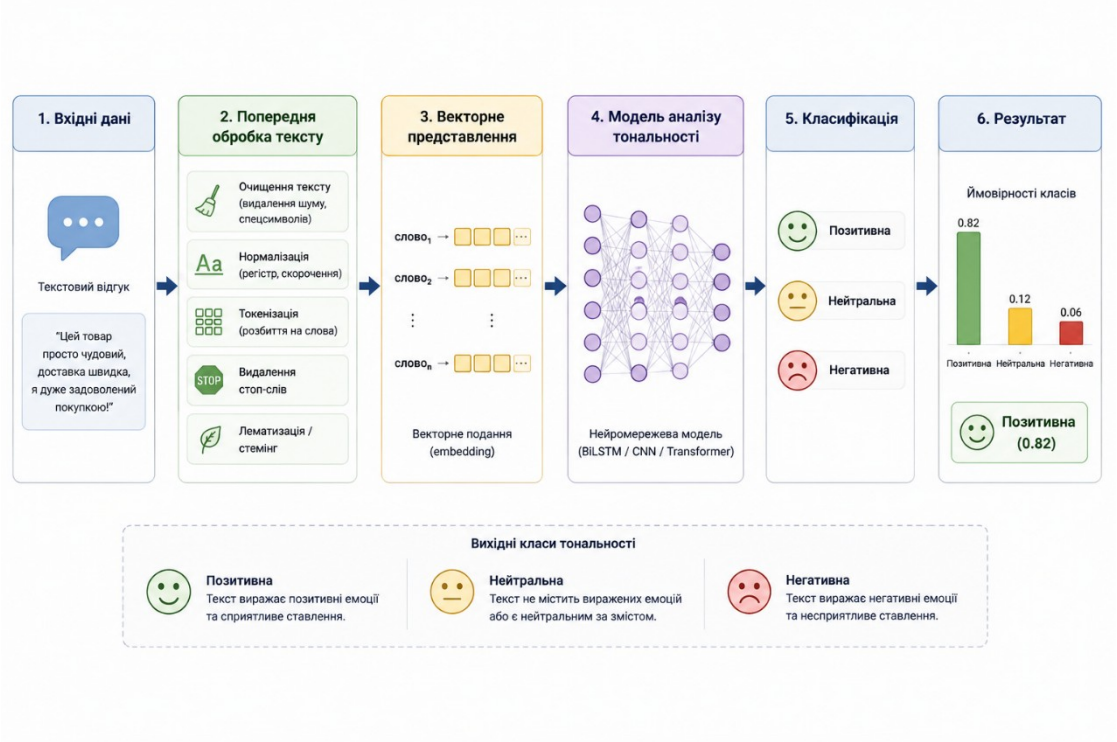


Рис1.1 Узагальнена схема процесу аналізу тональності текстових відгуків

Предметна область аналізу тональності має міждисциплінарний характер і поєднує підходи з комп’ютерної лінгвістики, статистики, інтелектуального аналізу даних та глибинного навчання. У сучасних методах все більше поширюються нейромережеві методи, які дозволяють ефективно враховувати контекст та сематичні зв’язки між словами, що є важливим для точного визначення емоційного забарвлення тексту.

В цій предметній області необхідна автоматизована обробка великих обсягів текстових даних в реальному часі, що робить дослідження актуальними. Застосування нейромережевих систем аналізу тональності дозволяє підвищити якість обробки відгуків, зменшити вплив людського фактору та забезпечити масштабованість при роботі з великими інформаційними потоками. Це підтверджує доцільність подальшого дослідження сучасних методів аналізу тональності та розробки ефективної нейромережевої системи. [1], [10]

1.2 Огляд сучасних методів АТВ

Лексиконні методи АТВ

Лексиконні методи аналізу текстових відгуків є одними з найперших підходів, що використовувалися для автоматичного визначення емоційного забарвлення тексту. Основна ідея цих методів полягає у використанні спеціально сформованих словників (лексиконів), у кожному слові заздалегідь присвоєно певну тональну оцінку – позитивну, негативну або нейтральну. Аналіз тексту здійснюється шляхом порівняння слів відгуку з елементами такого словника та обчислення загальної тональності на основі знайдених збігів.

У найпростішому випадку лексиконні методи працюють за принципом підрахунку кількості позитивно та негативно забарвлених слів у тексті. Якщо кількість позитивних слів перевищує кількість негативних, текст класифікується як позитивний, і навпаки. У більш складних реалізаціях до уваги береться вага слів, інтенсивність емоційного забарвлення, а також позиція слова в реченні.

Основною перевагою лексиконних методів є їх простота і прозорість. Результат аналізу легко інтерпретується, оскільки рішення приймаються на основі явних правил та наперед визначених словників. Крім того, такі методи не потребують великої кількості розмічених навчальних даних, що робить їх привабливими для задач, де відсутні, або обмежені корпуси текстів.

Існує велика кількість відомих лексиконів тональності, які широко застосовуються у практичних та наукових дослідженнях. До них належать SentiWordNet, NRC, Emotion Lexicon, AFINN, Opinion Lexicon та інші. Кожен з цих словників має власну структуру та підхід до оцінювання тональності, що впливає на результати аналізу. Частина лексиконів орієнтована на загальну позитивно-негативну класифікацію, тоді як інші додатково враховують емоційні категорії, такі як радість, гнів, страх або засмученість.

Попри свої переваги лексиконні методи мають низку суттєвих обмежень. Однією з основних проблем є недостатнє врахування контексту. Одне й те саме слово може мати різне емоційне забарвлення залежно від контексту використання, однак лексиконні підходи зазвичай не здатні конкретно обробляти такі випадки. Крім того, складнощі виникають при аналізі заперечень, іронії, сарказму та складних синтаксичних конструкцій.

Ще одним недоліком є залежність від повноти та якості словника. Якщо слово, яке має значний вплив на загальну тональність тексту, відсутнє у лексиконі, воно не буде враховане під час аналізу. Це особливо актуально для текстових відгуків, що містять сленг, неформальні вирази або специфічну термінологію. Крім того, більшість лексиконів орієнтовані на англомовні тексти, що ускладнює їх застосування для аналізу відгуків іншими мовами без додаткової адаптації. [11]

У результаті лексиконні методи Найчастіше використовуються як базові або допоміжні підходи в системах аналізу тональності. Вони можуть бути ефективними для швидкої оцінки загального емоційного фону тексту або для попереднього аналізу даних, однак у складних задачах, пов’язаних з великими

обсягами різнородних текстів, їх точність є обмеженою. Саме ці недоліки зумовили перехід до більш гнучких методів машинного навчання та нейромережевих підходів.

Машинні методи АТВ

Машинні методи аналізу текстових відгуків стали наступним етапом аналізу тональності текстових відгуків після лексиконних підходів. На відміну від словникових методів, ці базуються на використанні статистичних моделей, які навчаються на попередньо розмічених корпусах текстів та здатні автоматично виявляти закономірності між текстовими ознаками та відповідною тональністю.

Застосування методів машинного навчання вимагає подання текстових даних у числовій формі. Для цього використовуються різні методи векторизації тексту. Одним з найпростіших та найбільш поширених підходів є модель «Мішок слів» (Bag of Words), у якій текст представлено у вигляді вектора частот слів без урахування їх порядку. Більш інформативним є метод TF-IDF (Term Frequency – Inverse Document Frequency), який дозволяє зменшити вплив часто вживаних слів і підвищити значність термінів, характерних для конкретного тексту. [26], [27]

Після формування векторного подання тексту застосовуються класифікаційні алгоритми машинного навчання. Одним з найпопулярніших методів є наївний баєсівський класифікатор (Naive Bayes), який базується на ймовірісній моделі та припущенні незалежності ознак. Незважаючи на спрощеність цього припущення, метод демонструє достатньо хороші результати для задач аналізу тональності, особливо при роботі з великими обсягами текстових даних.

Ще одним поширеним алгоритмом є метод опорних векторів (Support Vector Machine, SVM). Цей підхід дозволяє будувати оптимальну розділювальну гіперплощину між класами та добре працює з високоймовірними просторами ознак, характерними для текстових даних. SVM широко застосовується в задачах класифікації тексту завдяки високій точності та стійкості до перенавчання.

Логістична регресія також є ефективним інструментом для аналізу тональності. Вона дозволяє оцінювати ймовірність належності тексту до певного класу та легко інтерпретується. У поєднанні з методами векторизації, такими як TF-IDF, логістична регресія часто демонструє конкурентні результати та використовується як базовий підхід для порівняння з більш складними моделями.

Основною перевагою класичних методів машинного навчання є їх відносна простота реалізації та зрозумілість результатів. Вони не потребують значних обчислювальних ресурсів і можуть бути ефективними навіть на середніх за обсягом навчальних вибірках. Крім того, такі методи добре масштабуються та легко інтегруються у прикладні програмні системи.

Разом з тим, класичні підходи мають низку обмежень. Головним недоліком є недостатнє врахування сематичного контексту та порядку слів у тексті. Моделі типу Bag of Words та TF-IDF не здатні конкретно обробляти довгострокові залежності між словами, а також мають труднощі з аналізом складних мовних конструкцій, іронії та сарказму. Внаслідок цього точність таких методів обмежується при роботі зі складними та неоднозначними текстовими відгуками.

Таким чином, класичні методи машинного навчання перевищують словникові методи за гнучкістю та точністю, проте поступаються нейромережевим підходам у здатності враховувати контекст та сематичні зв’язки. Це зумовлює доцільність переходу до нейронних мереж для побудови ефективних систем аналізу тональності текстових відгуків.

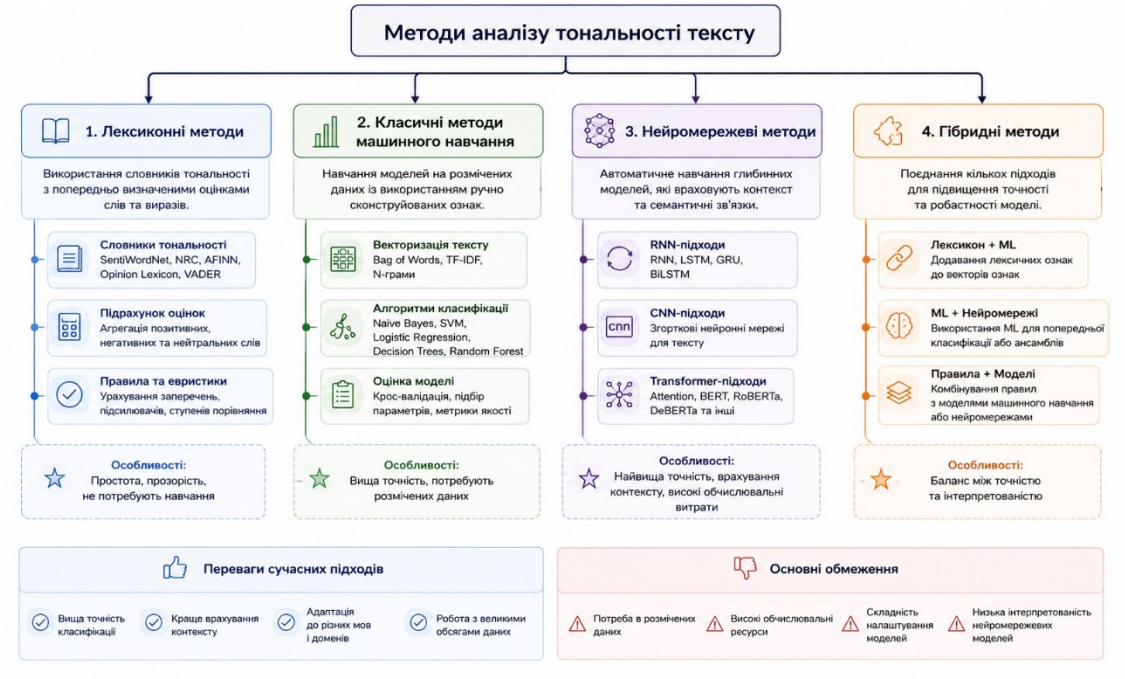


Рис 1.2 Класифікація сучасних методів аналізу тональності тексту

Як показано на Рис 1.2, сучасні методи аналізу тональності можна умовно поділити на чотири основні групи: лексиконні методи, класичні алгоритми машинного навчання, нейромережеві підходи та гібридні моделі. Лексиконні методи базуються на використанні словників тональності та наборів правил, тоді як алгоритми машинного навчання використовують статистичні ознаки тексту для побудови класифікаторів. Найбільш ефективними в задачах аналізу тексту вважаються нейромережеві моделі, які дозволяють враховувати контекст та семантичні зв’язки між словами. Гібридні підходи поєднують переваги декількох методів з метою підвищення точності класифікації.

Нейромережеві методи АТВ

Нейромережеві методи аналізу тональності текстових відгуків є сучасним етапом розвитку систем обробки природної мови та значно перевершують традиційні підходи за гнучкість й точністю. На відміну від класичних методів машинного навчання, нейронні мережі здатні автоматично формувати складні ознакові представлення тексту та враховувати контекстні залежності між словами, що є критично важливим для конкретного визначення емоційного забарвлення. [2], [4]

Однією з перших нейромережевих архітектур, які почали застосовувати для аналізу текстів, є рекурентні нейронні мережі. (Recurrent Neural Networks, RNN). Їх особливістю є наявність зворотних зв’язків, які дозволяють обробляти послідовні дані та враховувати попередній контекст під час аналізу кожного слова. У задачах аналізу тональності RNN дозволяють моделювати залежності між словами в межах речення або тексту, що суттєво підвищує якість класифікації порівняно з підходами, які ігнорують порядок слів. [3], [4]

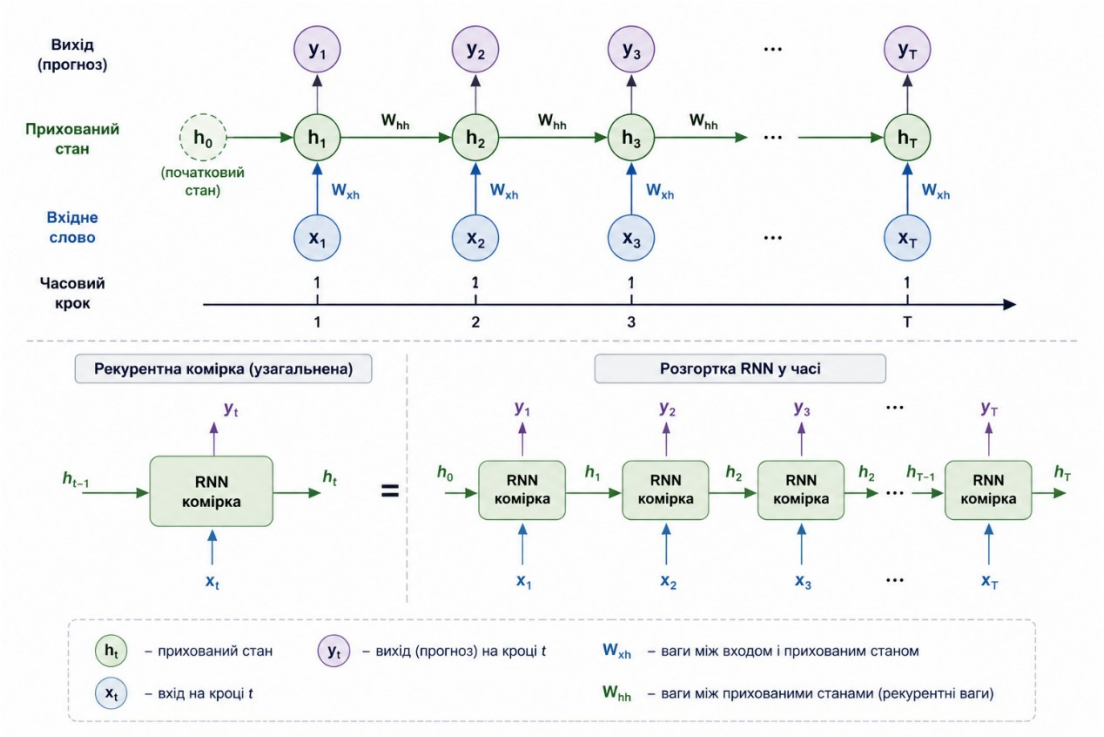


Рис 1.3 Принцип роботи рекурентної нейромережевої моделі

Як показано на Рис 1.3, рекурентна нейронна мережа обробляє текстову послідовність поетапно, передаючи прихований стан між часовими кроками. На кожному етапі модель отримує поточне слово та інформацію про попередній контекст, що дозволяє враховувати залежності між елементами тексту. Однак класичні RNN мають проблему втрати довготривалих залежностей, що ускладнює аналіз великих текстових послідовностей та знижує ефективність класифікації.

Проте класичні RNN мають обмеження, пов'язані з проблемами зникнення та вибуху градієнтів при обробці довгих послідовностей. Для подолання цих недоліків були розроблені архітектури з довготривалою пам'яттю, зокрема Long

Short-Term Memory (LSTM). LSTM-мережі використовують спеціальні керуючі механізми – так звані ворота, які дозволяють зберігати або забувати інформацію протягом тривалих послідовностей. Це робить їх ефективними для аналізу розгорнутих текстових відгуків, у яких тональність може формуватися поступово. [7]

Подальшим розвитком рекурентних моделей є двонаправлені моделі (Bidirectional LSTM), які аналізують текст одночасно у прямому та зворотному напрямках. Такий підхід дозволяє враховувати як попередній, так і наступний контекст для кожного слова, що є особливо корисним у задачах аналізу тональності, де значення слова може залежати від слів, розташованих після нього. [8]

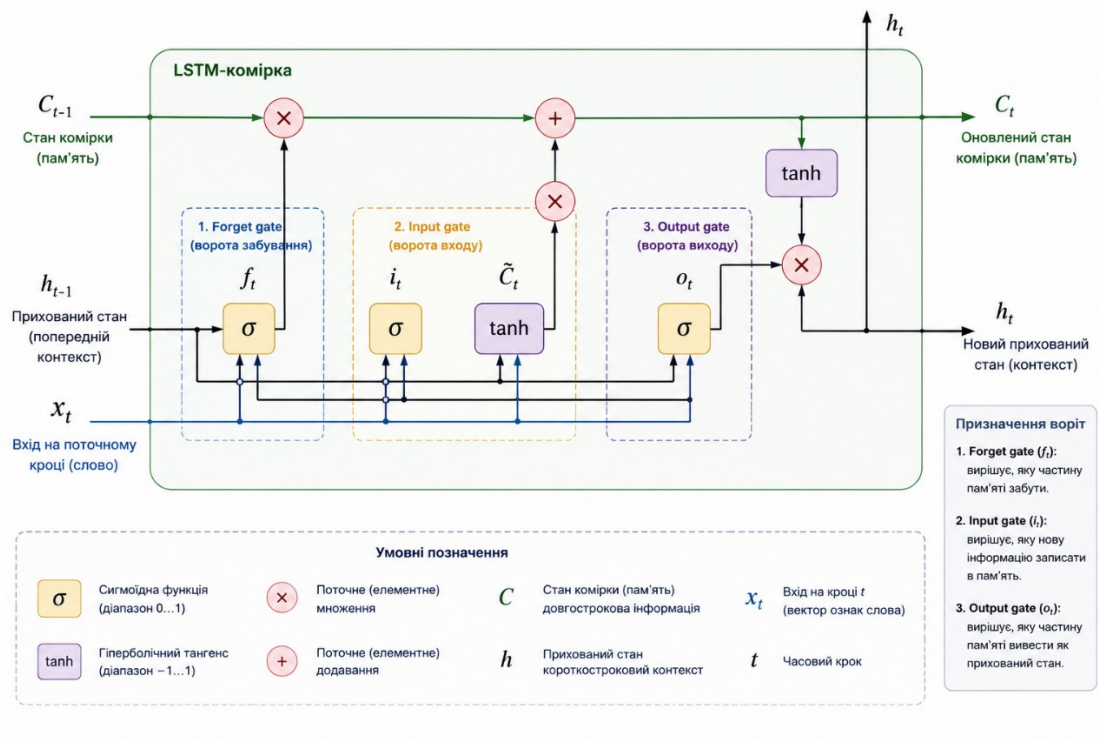


Рис 1.4 Архітектура LSTM-комірки

Як показано на Рис 1.4, LSTM-комірка містить спеціальні механізми керування інформаційними потоками - forget gate, input gate та output gate. Ворота забування визначають, яку інформацію необхідно видалити зі стану пам’яті, вхідні ворота відповідають за запис нових даних, а вихідні ворота формують прихований стан мережі. Наявність окремого стану пам’яті дозволяє моделі

ефективно зберігати довготривалі залежності між словами та підвищує якість аналізу текстових послідовностей.

Окрім рекурентних архітектур, у задачах аналізу тональності широко застосовуються згорткові нейронні мережі (Convolutional Neural Networks, CNN). Хоча CNN спочатку були розроблені для обробки зображень, вони виявилися ефективними і для аналізу текстових даних. У текстових CNN згорткові фільтри використовуються для виявлення локальних шаблонів, таких як ключові фрази або словосполучення, що мають характерне емоційне забарвлення. Це дозволяє моделі ефективно визначати тональність навіть у коротких текстових відгуках.

Суттєвим кроком у розвитку нейромережевих підходів аналізу тональності механізмів уваги та архітектури трансформерів. На відміну від рекурентних моделей, трансформери обробляють весь текст паралельно та використовують механізм самоуваги для визначення важливості кожного слова відносно інших. Це дозволяє моделі ефективно враховувати глобальний контекст і складні сематичні зв’язки у тексті. [6]

Однією з найвідоміших трансформерних моделей є BERT (Bidirectional Encoders Representations from Transformers), яка навчається на великих обсягах текстових даних і може бути адаптована до задачі аналізу тональності текстових даних шляхом додаткового навчання. Використання попередньо навчених трансформерних моделей дозволяє досягати високої точності навіть при обмежених обсягах розмічених даних, що є важливою перевагою для практичних застосувань. [9]

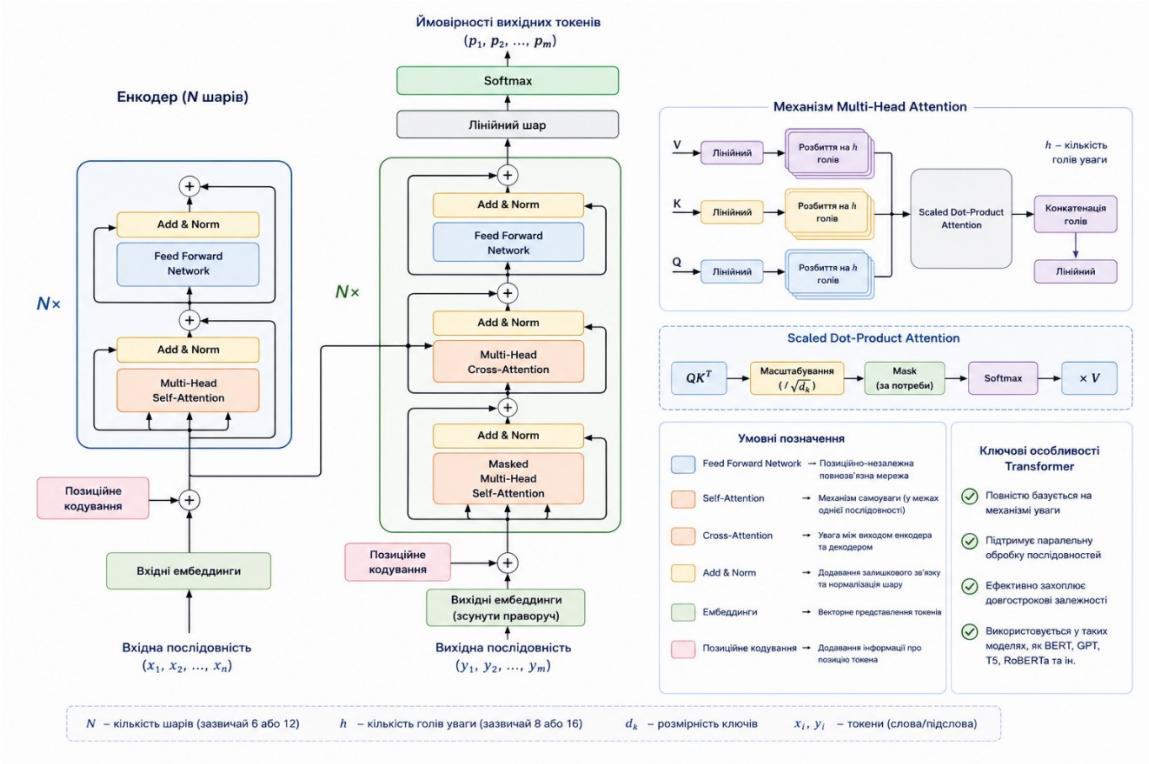


Рис 1.5 Архітектура моделі Transformer

Як показано на Рис 1.5, Transformer-архітектура складається з енкодера та декодера, що містять механізми multi-head attention і повнозв’язні шари. Основною особливістю даної архітектури є використання механізму self-attention, який дозволяє визначати важливість окремих слів у межах текстової послідовності. На відміну від рекурентних нейронних мереж, Transformer забезпечує паралельну обробку тексту та більш ефективно враховує довготривалі залежності між словами. Саме на основі Transformer-архітектур реалізовано сучасні NLP-моделі, зокрема BERT, GPT та RoBERTa.

Основною перевагою нейромережевих методів є їх здатність автоматично витягувати суттєві ознаки тексту без необхідності ручного проектування характеристик. Такі моделі краще адаптуються до різних мов, стилів і форматів текстових відгуків та демонструють високу узагальнювальну здатність. Разом з тим, нейромережеві підходи потребують значних обчислювальних ресурсів та ретельного налаштування параметрів, що може ускладнювати їх використання у деяких прикладних системах.

Отже, нейромережеві підходи є найбільш перспективним напрямом розвитку систем аналізу тональності текстових відгуків. Їх переваги у врахуванні контексту, сематики та складних мовних залежностей роблять доцільним вибір саме нейромережевого підходу для розробки системи аналізу тональності.

1.3 Порівняльний аналіз класичних та нейромережевих підходів

Після огляду основних груп методів аналізу тональності текстових відгуків, а саме лексиконних підходів, класичних методів машинного навчання та нейромережевих моделей. Кожен із цих підходів має власні переваги та обмеження, що зумовлює доцільність їх порівняльного аналізу з метою вибору оптимального рішення для побудови ефективної системи аналізу тональності.

Лексиконні методи характеризуються простотою реалізації та прозорістю результатів. Вони не потребують етапу навчання та можуть бути використані навіть за відсутності розмічених даних. Разом із тим, такі методи є малогнучкими та не здатні коректно враховувати контекст, що суттєво знижує точність аналізу при роботі зі складними текстами, зокрема з використанням іронії, заперечень або багатозначних слів.

Класичні методи машинного навчання значно перевершують лексиконні підходи за точністю та адаптивністю. Вони дозволяють автоматично виявляти статистичні залежності між ознаками тексту та його тональністю, проте потребують наявності розмічених навчальних вибірок. Основним обмеженням таких методів є спрощене подання тексту у вигляді векторів ознак, яке не враховує повною мірою семантичний контекст та порядок слів.

Нейромережеві підходи є найбільш сучасним та перспективним напрямом аналізу тональності текстових відгуків. Вони забезпечують автоматичне формування ознакових представлень тексту та здатні враховувати складні контекстні й семантичні залежності. Завдяки цьому нейромережеві моделі демонструють найвищу точність, особливо при роботі з великими та різнорідними

масивами текстових даних. Недоліками таких підходів є підвищені вимоги до обчислювальних ресурсів та складність реалізації.

Для наочного порівняння основних характеристик розглянутих підходів у таблиці 1.1 наведено узагальнені результати аналізу.

Таблиця 1.1

Критерій порівняння	**Лексиконні методи**	**Класичні ML-методи**	**Нейромережеві методи**
Необхідність навчальних даних	Не потребують	Потребують	Потребують
Урахування контексту	Низьке	Середнє	Високе
Обробка довгих текстів	Обмежена	Обмежена	Ефективна
Стійкість до шуму та сленгу	Низька	Середня	Висока
Точність аналізу	Низька–середня	Середня	Висока
Інтерпретованість результатів	Висока	Середня	Низька
Обчислювальні витрати	Низькі	Середні	Високі
Складність реалізації	Низька	Середня	Висока

Проведений порівняльний аналіз показує, що лексиконні методи доцільно використовувати лише для базового або попереднього аналізу тональності. Класичні методи машинного навчання можуть бути ефективними для задач середньої складності, проте їх можливості обмежені при роботі з великими обсягами неструктурованих текстових даних. Найбільш перспективними для побудови сучасних систем аналізу тональності є нейромережеві підходи, які забезпечують високу точність та адаптивність за рахунок глибокого аналізу контексту. [10], [11], [12]

Отримані результати порівняльного аналізу підтверджують доцільність використання нейромережевої моделі для розробки системи аналізу тональності текстових відгуків, що обґрунтовує вибір напряму подальших досліджень та визначає логічний перехід до формування постановки задачі.

1.4 Огляд існуючих програмних рішень та бібліотек

На сьогоднішній день існує велика кількість програмних рішень та бібліотек, призначених для аналізу тональності текстових даних. Вони відрізняються за функціональними можливостями, рівнем складності, підтримуваними мовами та підходами до реалізації алгоритму аналізу тональності. Використання готових інструментів дозволяє значно скоротити час розробки систем.

Серед готових програмних рішень для аналізу тональності можна виділити як комерційні, так і відкриті платформи. До комерційних сервісів належать рішення, що надаються у вигляді хмарних API та дозволяють виконувати аналіз тексту без необхідності розгортання власної інфраструктури. Такі сервіси зазвичай підтримують багатомовний аналіз, автоматичне масштабування та інтеграцію з іншими інформаційними системами.

Відкриті програмні рішення для аналізу тональності представлені різноманітними бібліотеками та фреймворками, які можуть бути інтегровані у власні програмні продукти. Їх перевагою є доступ до вихідного коду, можливість модифікації алгоритмів та адаптації під конкретні вимоги задачі. Такі рішення особливо актуальні для наукових досліджень та навчальних проєктів, де важливою є прозорість та контроль над процесом аналізу.

Важливу роль у розробці систем аналізу тональності відіграють бібліотеки машинного навчання, які забезпечують реалізацію алгоритмів класифікації тексту. До найбільш поширених бібліотек цього класу належить Scikit-learn, яка дає широкий набір інструментів для попередньої обробки тексту, векторизації та побудови моделей машинного навчання. Її перевагою є простота використання, детальна документація та висока стабільність, що робить її популярною як у наукових, так і прикладних розробках.

Для реалізації нейромережевих методів аналізу тональності широко застосовуються бібліотеки глибинного навчання. Однією з найбільш популярних є TensorFlow, який забезпечує побудову, навчання та оптимізацію складних

нейронних мереж. TensorFlow підтримує різні типи архітектур, зокрема рекурентні та розгорткові мережі, а також трансформерні моделі. Завдяки високій продуктивності та можливості використання апаратного прискорення ця бібліотека часто застосовується у промислових системах аналізу тексту.

Альтернативою TensorFlow є бібліотека PyTorch, яка набуває більшої популярності у наукових дослідженнях та експериментальних проєктах. PyTorch відрізняється динамічною обчислювальною графікою, що спрощує налагодження та модифікацію моделей. Це робить її зручною для розробки та використання нейромережевих моделей аналізу тональності, зокрема при роботі з рекурентними мережами та трансформерами.

Окрему групу становлять спеціалізовані бібліотеки для обробки природної мови, які надають інструменти для токенізації, лематизації, частиномовного аналізу та роботи з текстовими корпусами. До таких бібліотек належать NLTK, spaCy та Hugging Face Transformers. Вони значно спрощують процес попередньої обробки текстових даних та інтеграції нейромережевих моделей у прикладні системи аналізу тональності.

1.5 Постановка задачі

На підставі проведеного аналізу предметної області та сучасних методів аналізу текстових відгуків можна сформулювати основну мету та завдання цієї роботи. Зростання обсягів текстових даних, що генеруються користувачами цифрових платформ, зумовлює необхідність створення автоматизованих систем, що здатні швидко та з високою точністю визначити емоційне забарвлення текстових повідомлень. Особливо актуальним є застосування нейромережевих методів, які забезпечують глибший аналіз контексту та сематичних зв’язків у тексті.

Метою цієї роботи є розробка нейромережевої системи аналізу тональності текстових відгуків, яка забезпечує автоматичну класифікацію тексту за емоційним забарвленням із використанням сучасних методів глибинного навчання.

Для досягнення поставленої мети потрібно розв’язати такі завдання:

1. Провести аналіз предметної області АТВ.
2. Дослідити сучасні методи АТВ та здійснити їх порівняльний аналіз.
3. Визначити доцільність використання нейромережевого підходу.
4. Проаналізувати вхідні та вихідні дані системи.
5. Обрати та обґрунтувати архітектуру нейронної мережі.
6. Розробити математичну модель нейромережевого класифікатора.
7. Реалізувати програмну систему аналізу тональності.
8. Провести тестування та оцінити якість роботи моделі.

Об’єктом дослідження є процес автоматизованого аналізу тональності текстових відгуків.

Предметом дослідження є методи і моделі нейромережевої класифікації текстових даних для визначення їх емоційного забарвлення.

Вхідними даними є текстові відгуки користувачів, представлені у вигляді неструктурованих текстових повідомлень.

Вихідними даними є:

- клас тональності (позитивна, негативна, нейтральна);
- ймовірісна оцінка належності тексту до відповідного класу.

Під час розробки системи перебачається:

- використання розміченого навчального набору даних;
- попередня текстова обробка вхідних даних;
- реалізація системи у вигляді програмного застосунку з інтерфейсом користувача;
- орієнтація на тексти певною мову (українська)

Не розглядається задача багатомовного або аспектно-орієнтованого аналізу, якщо це не перебачено технологічними вимогами.

З формальної точки зору задача аналізу тональності може бути представлена як задача класифікації. Необхідно побудувати функцію f: X → Y, де X – множина текстових відгуків, а Y – множина класів тональності.

Функція f реалізується на основі нейромережевої моделі, яка навчається на розмічених прикладах та мінімізує функцію втрат під час навчання.

2 Інформаційне та математичне забезпечення нейромережевої системи

2.1 Аналіз та опис вхідних та вихідних даних

У процесі розробки нейромережевої системи аналізу тональності текстових відгуків особливе значення має детальне дослідження вхідних та вихідних даних, оскільки саме їх природа, структура та якісні характеристики визначають підходи до попередньої обробки, вибір математичних методів та архітектури моделі. Текстові відгуки, що виступають вхідними даними системи, належать до класу неструктурованої інформації, яка характеризується високим рівнем варіативності, неоднорідністю та значною залежністю від контексту. Такі дані формуються у процесі взаємодії користувачів із цифровими платформами та можуть надходити з різноманітних джерел, включаючи вебсайти електронної комерції, соціальні мережі, форуми, сервіси оцінювання товарів і послуг, а також мобільні застосунки. Унаслідок цього виникає необхідність врахування широкого спектра

особливостей текстових даних, які безпосередньо впливають на складність їх автоматизованого аналізу.

Однією з ключових характеристик вхідних текстів є їх довжина, яка може варіюватися від коротких емоційних повідомлень, що складаються з одного або кількох слів, до розгорнутих відгуків, що містять складні синтаксичні конструкції та багаторівневу аргументацію. Така неоднорідність ускладнює процес обробки, оскільки модель повинна однаково ефективно працювати як з короткими, так і з довгими текстами, зберігаючи здатність до узагальнення. Крім того, текстові відгуки характеризуються значною лінгвістичною складністю, яка проявляється у використанні багатозначних слів, складних граматичних конструкцій, заперечень, умовних висловлювань та інших мовних особливостей. Наприклад, наявність заперечення може кардинально змінювати тональність висловлювання, що потребує від моделі здатності враховувати контекстні залежності між словами.

Важливим аспектом є також наявність шуму у вхідних даних. Текстові відгуки, створені користувачами, часто містять орфографічні та граматичні помилки, неформальні вирази, сленг, скорочення, а також різноманітні символи, включаючи емодзі та знаки пунктуації, що використовуються не за правилами. У деяких випадках може спостерігатися змішування мов або використання транслітерації, що додатково ускладнює задачу автоматичного аналізу. Усе це знижує якість вихідних даних та вимагає застосування ефективних методів попередньої обробки для зменшення впливу шуму та приведення тексту до уніфікованого вигляду.

Суттєвою особливістю текстових відгуків є їх контекстна залежність та емоційна неоднозначність. Значення окремих слів або фраз не може бути визначене ізольовано, без урахування загального контексту висловлювання. Більше того, один і той самий текст може містити як позитивні, так і негативні оцінки, що формує так звану змішану тональність. У таких випадках задача класифікації значно ускладнюється, оскільки модель повинна визначити домінуючу емоційну складову або врахувати баланс оцінок. Додаткову складність

створюють явища іронії, сарказму та прихованих емоцій, які часто зустрічаються у відгуках користувачів і важко піддаються формалізації.

З формальної точки зору вхідні дані можуть бути представлені у вигляді множини текстових повідомлень, кожне з яких є послідовністю лінгвістичних одиниць. У процесі підготовки до машинного аналізу ці дані перетворюються у числове представлення, придатне для обробки алгоритмами машинного навчання. Таке перетворення дозволяє перейти від текстової форми до векторного простору ознак, у якому кожен документ описується набором числових характеристик. Вибір способу такого подання є критично важливим, оскільки він визначає, наскільки повно буде збережено семантичну та контекстну інформацію, закладену у тексті.

Вихідні дані нейромережевої системи представлені результатами класифікації тональності текстових відгуків. У межах даної роботи розглядається задача багатокласової класифікації, де кожному тексту ставиться у відповідність один із трьох можливих класів: позитивний, негативний або нейтральний. Однак сучасні підходи до аналізу тональності передбачають не лише визначення класу, але й оцінку ступеня впевненості моделі у своєму рішенні. З цією метою вихід моделі формується у вигляді вектора ймовірностей, де кожен елемент відповідає ймовірності належності тексту до певного класу. Такий підхід дозволяє отримати більш гнучке та інформативне представлення результатів, що може бути використане для подальшого аналізу або прийняття рішень.

Важливу роль у побудові ефективної нейромережевої системи відіграє якість навчальних даних. Для навчання моделі використовується розмічена вибірка, у якій кожному текстовому відгуку відповідає заздалегідь визначена мітка тональності. Проте процес розмітки текстів є суб’єктивним і може супроводжуватися неоднозначностями, особливо у випадках змішаної або слабко вираженої емоційної тональності. Це може призводити до появи помилок у навчальних даних, що негативно впливає на якість моделі. Крім того, у реальних наборах даних часто спостерігається дисбаланс класів, коли кількість відгуків

однієї категорії суттєво перевищує інші. Такий дисбаланс може призводити до зміщення моделі у бік більш представленого класу, що потребує застосування спеціальних методів балансування даних або корекції функції втрат.

Таким чином, проведений аналіз показує, що вхідні дані для задачі аналізу тональності характеризуються високим рівнем складності, неоднорідністю та залежністю від контексту, що зумовлює необхідність використання сучасних методів їх обробки. Вихідні дані, у свою чергу, повинні забезпечувати не лише класифікацію тексту, але й можливість оцінки впевненості моделі у прийнятому рішенні. Урахування зазначених особливостей є необхідною умовою для побудови ефективної нейромережевої системи аналізу тональності текстових відгуків.

2.2 Попередня обробка текстових даних

Попередня обробка текстових даних є одним із ключових етапів побудови нейромережевої системи аналізу тональності, оскільки саме на цьому етапі здійснюється перетворення неструктурованого тексту у форму, придатну для подальшого математичного аналізу та обробки алгоритмами машинного навчання. Якість виконання цього етапу безпосередньо впливає на ефективність навчання моделі, її здатність до узагальнення та точність класифікації. У випадку текстових відгуків, що характеризуються високим рівнем шуму, мовною неоднорідністю та контекстною складністю, роль попередньої обробки є особливо важливою, оскільки вона дозволяє зменшити вплив небажаних факторів та виділити релевантні ознаки.

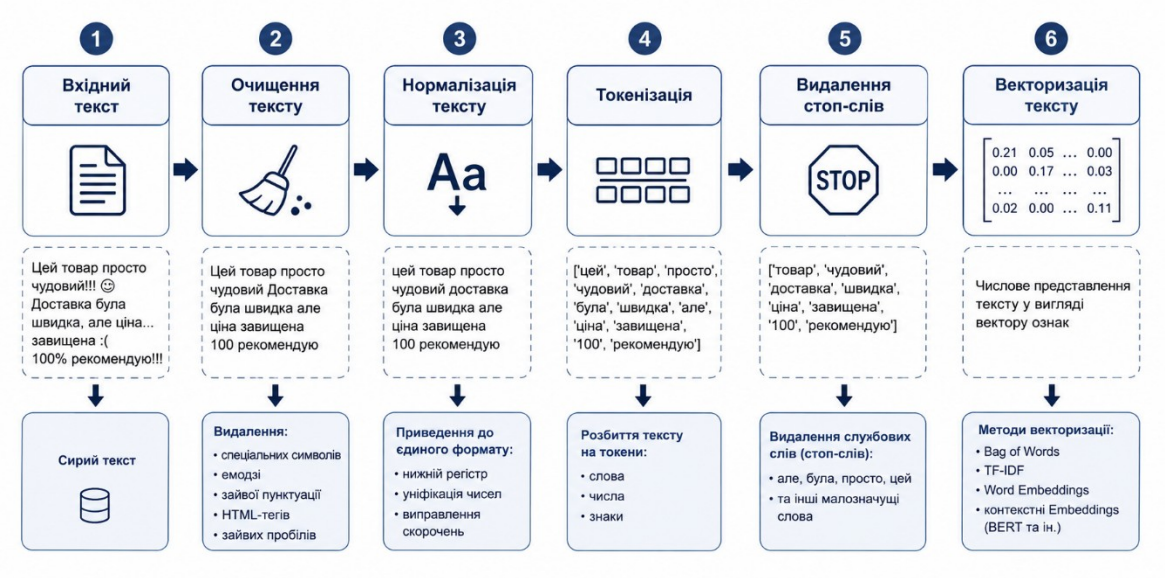


Рис 2.1 Основні етапи попередньої обробки текстових даних

Як показано на Рис 2.1, процес попередньої обробки тексту складається з декількох послідовних етапів, серед яких очищення тексту, нормалізація, токенізація, видалення стоп-слів та векторизація. Виконання наведених операцій дозволяє підвищити якість текстових даних та забезпечити ефективну роботу нейромережевої моделі аналізу тональності.

Процес попередньої обробки текстових даних зазвичай складається з кількох послідовних етапів, кожен з яких виконує окрему функцію у загальному конвеєрі обробки. На початковому етапі здійснюється очищення тексту від зайвих або нерелевантних елементів. Це включає видалення спеціальних символів, HTML-тегів, надлишкових пробілів, а також символів пунктуації, які не несуть суттєвого смислового навантаження. Водночас слід зазначити, що у деяких випадках пунктуація може відігравати роль індикатора емоційності (наприклад, знаки оклику), тому рішення про її видалення повинно прийматися з урахуванням специфіки задачі.

Наступним етапом є нормалізація тексту, яка передбачає приведення всіх символів до єдиного формату. Найчастіше це включає переведення тексту до нижнього регістру, що дозволяє уникнути дублювання слів, які відрізняються лише регістром написання. Крім того, на цьому етапі можуть виконуватися додаткові операції, такі як уніфікація написання чисел, скорочень та окремих

мовних конструкцій. Нормалізація сприяє зменшенню розмірності словникового простору та підвищенню ефективності подальших етапів обробки.

Важливим кроком є токенізація тексту, тобто розбиття його на окремі одиниці — токени. У найпростішому випадку токенами виступають окремі слова, однак у сучасних підходах можуть використовуватися також підсловні одиниці або навіть символи. Токенізація є основою для подальшого аналізу, оскільки саме вона визначає, яким чином текст буде представлений у вигляді послідовності елементів. Для української мови цей процес може ускладнюватися морфологічною варіативністю слів та наявністю різних форм одного й того ж лексичного елемента. [16]

Після токенізації часто виконується видалення стоп-слів — службових частин мови, які не несуть значного смислового навантаження і не впливають суттєво на визначення тональності тексту. До таких слів належать сполучники, прийменники, займенники тощо. Видалення стоп-слів дозволяє зменшити обсяг даних та сфокусувати увагу моделі на більш інформативних елементах тексту. Водночас у задачах аналізу тональності необхідно обережно підходити до цього етапу, оскільки деякі службові слова можуть впливати на зміст висловлювання, зокрема у випадках заперечення.

Наступним важливим етапом є приведення слів до їх базової форми, що може здійснюватися за допомогою стемінгу або лематизації. Стемінг передбачає механічне відсікання закінчень слів, у той час як лематизація базується на використанні словників та мовних правил для визначення початкової форми слова. Лематизація є більш точним методом, особливо для мов зі складною морфологією, таких як українська, проте вона потребує більших обчислювальних ресурсів та наявності відповідних лінгвістичних інструментів. Використання цих методів дозволяє зменшити різноманітність словоформ та покращити узагальнюючі властивості моделі.

Ключовим етапом попередньої обробки є перетворення тексту у числове представлення, яке може бути використане алгоритмами машинного навчання. У

традиційних підходах для цього застосовуються методи векторизації, такі як модель “мішок слів” (Bag of Words) та TF-IDF. У моделі “мішок слів” текст представлено у вигляді вектора, де кожна компонента відповідає частоті появи певного слова у документі. Незважаючи на простоту, цей підхід не враховує порядок слів та контекстні залежності.

Більш інформативним є метод TF-IDF, який дозволяє оцінити важливість слова як у межах окремого документа, так і в усьому корпусі текстів. Суть цього підходу полягає у врахуванні не лише частоти появи слова в конкретному тексті, але й ступеня його поширеності серед усіх документів. Таким чином, слова, що часто зустрічаються в одному документі, але рідко — в інших, отримують більшу вагу, ніж загальновживані слова.

Вага терміна у методі TF-IDF визначається як добуток двох складових: термчастоти (TF), яка відображає частоту входження слова в документ, та оберненої частоти документів (IDF), що зменшує вагу слів, які часто зустрічаються у всьому корпусі. У результаті такого підходу формується більш інформативне представлення тексту, яке дозволяє виділити найбільш значущі слова для подальшого аналізу.

Однак сучасні нейромережеві підходи дедалі частіше використовують більш складні методи подання тексту у вигляді щільних векторних представлень — так званих embedding-векторів. На відміну від розріджених векторів TF-IDF, embeddings дозволяють враховувати семантичні зв’язки між словами. Методи на кшталт Word2Vec або GloVe формують векторні представлення слів таким чином, що слова з подібним значенням мають близькі вектори у просторі ознак. [13], [14]

Подальшим розвитком цього підходу є контекстні embedding-моделі, зокрема трансформерні архітектури, які формують представлення слова з урахуванням його оточення у тексті. Це дозволяє більш точно відображати значення слів у різних контекстах та значно підвищує якість аналізу тональності.

Узагальнюючи, процес попередньої обробки текстових даних можна представити як послідовний конвеєр перетворень, що включає очищення,

нормалізацію, токенізацію, фільтрацію та векторизацію. Кожен із цих етапів відіграє важливу роль у підготовці даних та формуванні якісного вхідного представлення для нейромережевої моделі.

Таким чином, ефективна попередня обробка текстових даних є необхідною умовою для побудови високоточної системи аналізу тональності. Вона дозволяє зменшити вплив шуму, врахувати мовні особливості та сформувати інформативне представлення тексту, що забезпечує покращення результатів класифікації.

На Рис 2.2 зображено основні етапи обробки даних. Від надходження тексту – до класифікованих відгуків.

2.2 Інфографіка етапів обробки даних

2.3 Вибір та обґрунтування архітектури нейронної мережі

Вибір архітектури нейронної мережі є одним із визначальних етапів розробки системи аналізу тональності текстових відгуків, оскільки саме структура моделі визначає її здатність ефективно обробляти текстові дані, враховувати контекст та забезпечувати достатній рівень точності класифікації. З огляду на специфіку задачі, пов’язаної з обробкою природної мови, особливу увагу слід приділити моделям, здатним працювати з послідовними даними та враховувати залежності між словами у тексті.

Текстовий відгук у загальному випадку можна розглядати як послідовність слів або токенів, кожен з яких після етапу векторизації подається у вигляді числового вектора. Таким чином, вхідні дані для нейронної мережі мають послідовну структуру, що зумовлює доцільність використання моделей, орієнтованих на обробку часових або послідовних залежностей. У цьому контексті одними з найбільш поширених є рекурентні нейронні мережі, які дозволяють враховувати інформацію про попередні елементи послідовності під час аналізу кожного наступного слова.

Однак класичні рекурентні нейронні мережі мають низку обмежень, пов’язаних зі складністю навчання на довгих послідовностях. Зокрема, при збільшенні довжини тексту виникають проблеми з передачею інформації між віддаленими елементами, що може призводити до втрати важливого контексту. Це суттєво знижує ефективність таких моделей при аналізі розгорнутих текстових відгуків, у яких тональність формується поступово.

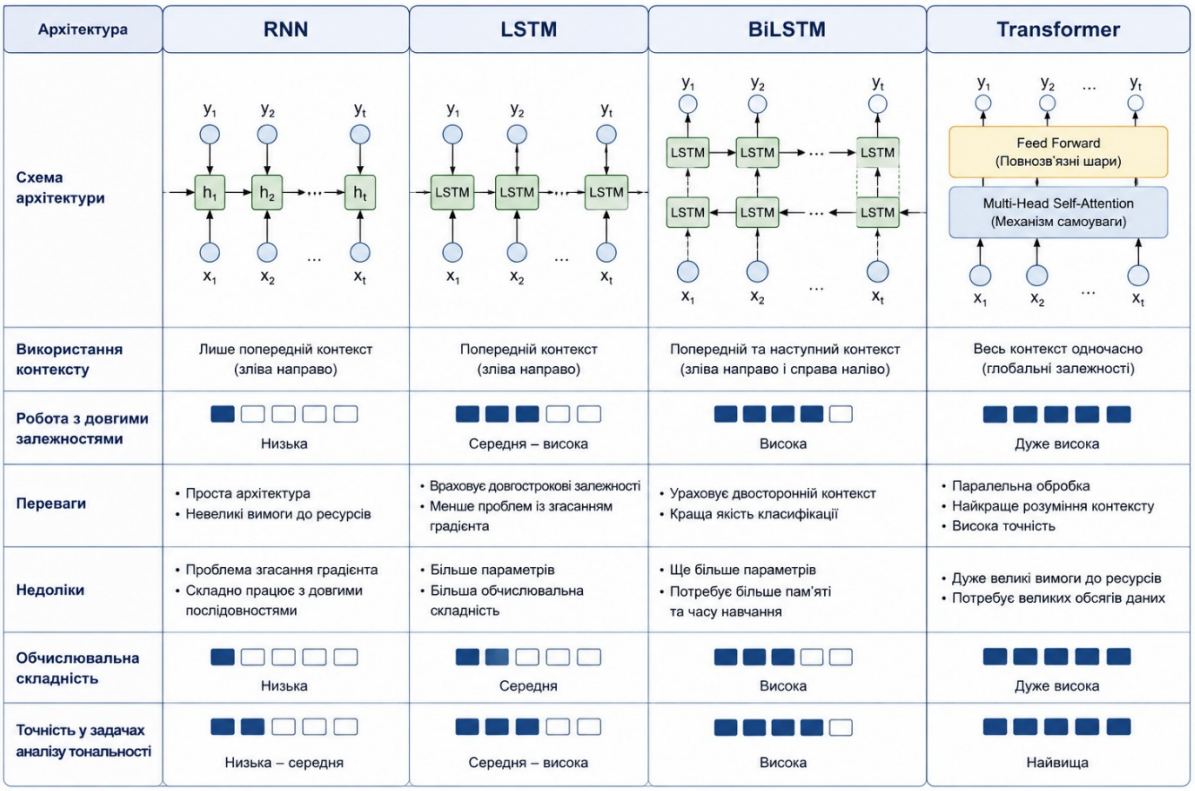


Рис 2.3 Порівняння архітектур нейронних мереж для задач аналізу тексту

Як показано на Рис 2.3, архітектура BiLSTM поєднує переваги класичних рекурентних мереж та механізмів довготривалого запам’ятовування, забезпечуючи врахування контексту як у прямому, так і у зворотному напрямках. У порівнянні зі звичайними RNN та LSTM, модель BiLSTM демонструє вищу ефективність у задачах аналізу тональності тексту, оскільки дозволяє враховувати залежності між словами у межах всього речення. Саме тому дана архітектура була обрана як основа нейромережевої моделі розробленої системи.

З метою подолання зазначених недоліків доцільно використовувати більш складні архітектури, зокрема мережі з довготривалою пам’яттю (LSTM). Основною перевагою LSTM є наявність спеціального механізму керування інформаційними потоками, який дозволяє моделі зберігати або ігнорувати інформацію залежно від її важливості. Завдяки цьому такі мережі здатні ефективно враховувати як короткострокові, так і довгострокові залежності між словами, що є критично важливим для задач аналізу тональності.

Подальшим розвитком рекурентних моделей є двонаправлені архітектури, які забезпечують обробку тексту одночасно у прямому та зворотному напрямках. Це дозволяє враховувати не лише попередній контекст, але й інформацію, що знаходиться після певного слова у реченні. Такий підхід є особливо корисним у випадках, коли значення слова визначається його оточенням, що часто зустрічається у природній мові.

На завершальному етапі обробки сформоване нейромережею представлення тексту передається до повнозв’язного шару, який виконує класифікацію та визначає належність тексту до одного з класів тональності. Для отримання ймовірнісного розподілу між класами використовується функція нормалізації, яка дозволяє інтерпретувати вихід моделі як набір імовірностей. Остаточне рішення про клас тональності приймається на основі найбільшого значення цієї ймовірності.

Окрім рекурентних моделей, у сучасних системах аналізу тексту все більшого поширення набувають трансформерні архітектури, які базуються на

механізмі самоуваги. На відміну від рекурентних мереж, такі моделі обробляють увесь текст паралельно та здатні ефективно враховувати взаємозв’язки між усіма словами в межах послідовності. Це дозволяє досягати високої точності аналізу, особливо при роботі з великими обсягами даних. Водночас використання трансформерів потребує значних обчислювальних ресурсів, що може бути обмеженням у прикладних системах.

З урахуванням проведеного аналізу та особливостей поставленої задачі доцільним є використання нейромережевої архітектури на основі LSTM або двонаправленої LSTM. Такий вибір обумовлений здатністю цих моделей ефективно обробляти послідовні дані, враховувати контекст та забезпечувати прийнятний баланс між точністю та обчислювальною складністю. У разі наявності достатніх ресурсів альтернативним рішенням може бути застосування трансформерних моделей, які демонструють найвищі результати у задачах аналізу тональності. [2], [4]

Таким чином, обрана архітектура нейронної мережі відповідає особливостям вхідних даних та вимогам до системи, що дозволяє забезпечити ефективний аналіз тональності текстових відгуків та досягти високої якості класифікації.

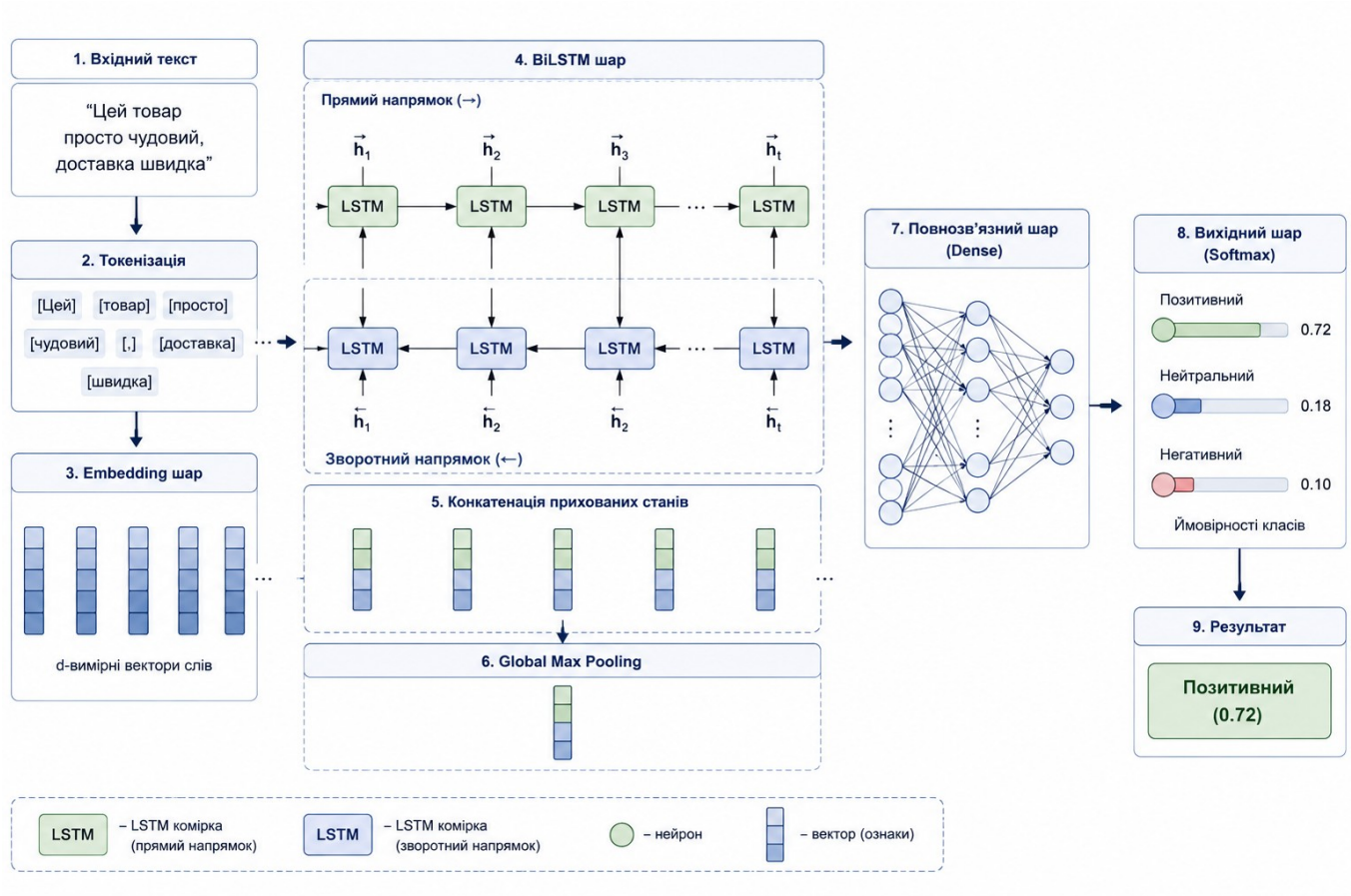


Рис 2.4 Структура нейромережевої архітектури BiLSTM

Як показано на рисунку 2.3, процес аналізу тексту в моделі BiLSTM починається з токенізації вхідного повідомлення та формування embedding-представлення слів. Далі текстова послідовність обробляється двонапрямленими LSTM-шарами, які аналізують контекст як у прямому, так і у зворотному напрямках. Отримані приховані стани об’єднуються та передаються до повнозв’язного шару для подальшої класифікації. На вихідному шарі Softmax формується ймовірність належності тексту до одного з класів тональності — позитивного, нейтрального або негативного.

2.4 Математична модель нейромережевої системи

Математична модель нейромережевої системи аналізу тональності текстових відгуків визначає формальний підхід до обробки вхідних даних, структуру перетворень та спосіб отримання кінцевого результату класифікації. У межах даної роботи задача аналізу тональності розглядається як задача

багатокласової класифікації, у якій кожному текстовому відгуку ставиться у відповідність один із можливих класів емоційного забарвлення.

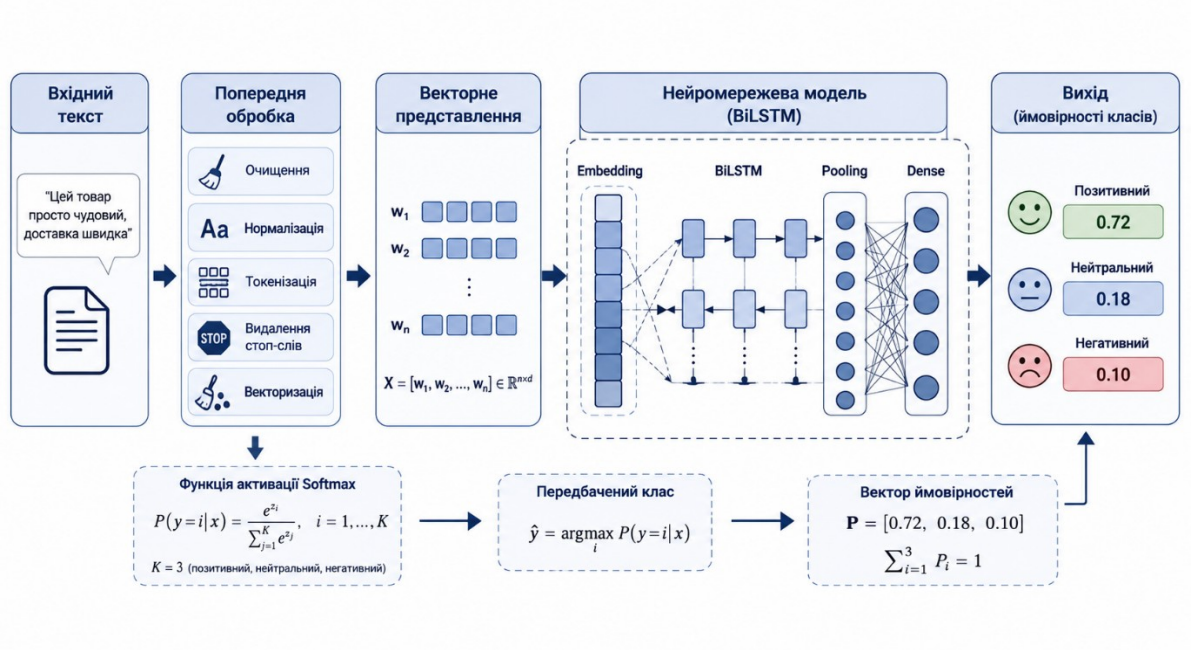


Рис 2.5 Математична модель класифікації тональності

Як показано на Рис 2.5, процес класифікації текстових повідомлень складається з декількох послідовних етапів. Спочатку текст проходить попередню обробку та векторизацію, після чого embedding-представлення передаються до нейромережевої моделі BiLSTM. Далі виконується агрегування прихованих станів та обчислення ймовірностей належності повідомлення до одного з класів тональності за допомогою функції Softmax. Отримані значення використовуються для визначення позитивної, нейтральної або негативної емоційної тональності текстового повідомлення.

З формальної точки зору модель реалізує відображення множини вхідних даних у множину вихідних класів. Вхідні дані представлені текстовими відгуками, які після попередньої обробки та векторизації подаються у вигляді числових векторів або послідовностей векторів. У свою чергу, вихід моделі формується у вигляді оцінки належності кожного тексту до певного класу тональності.

У загальному вигляді роботу моделі можна описати як послідовність перетворень, у ході яких вхідний текст проходить через декілька рівнів обробки. На першому етапі здійснюється перетворення тексту у векторне представлення, яке відображає семантичні властивості слів. Далі ці представлення подаються на вхід нейронної мережі, яка виконує їх обробку з урахуванням контексту та внутрішніх залежностей між елементами послідовності.

У процесі проходження через нейромережеву модель формується узагальнене представлення тексту, яке містить найбільш значущу інформацію для визначення його тональності. Це представлення використовується на завершальному етапі для прийняття рішення щодо класу, до якого належить текстовий відгук.

Для отримання інтерпретованого результату використовується механізм нормалізації вихідних значень, який дозволяє представити їх у вигляді ймовірностей. Це забезпечує можливість не лише визначити найбільш ймовірний клас, але й оцінити ступінь впевненості моделі у прийнятому рішенні. Кінцева класифікація здійснюється шляхом вибору класу з максимальною ймовірністю.

Процес навчання моделі базується на використанні розмічених даних, де для кожного текстового відгуку відомий правильний клас тональності. У ході навчання модель поступово коригує свої внутрішні параметри таким чином, щоб мінімізувати різницю між передбаченими результатами та фактичними значеннями. Це досягається шляхом ітеративного оновлення вагових коефіцієнтів на основі помилки класифікації. [30]

З практичної точки зору важливим є забезпечення узагальнюючої здатності моделі, тобто її здатності правильно класифікувати нові, раніше невідомі дані. Для цього необхідно уникати перенавчання, яке може виникати у випадку надмірної адаптації моделі до навчальної вибірки. З цією метою можуть застосовуватися різні підходи, зокрема регуляризація, використання валідаційних вибірок та контроль кількості епох навчання.

Таким чином, математична модель нейромережевої системи аналізу тональності текстових відгуків базується на послідовному перетворенні текстових даних у числове представлення, їх обробці за допомогою нейронної мережі та подальшій класифікації за визначеними класами. Такий підхід дозволяє ефективно враховувати складну структуру тексту та забезпечує достатню точність аналізу у практичних застосуваннях.

2.5 Алгоритм навчання та оцінки якості моделі

Алгоритм навчання нейромережевої моделі аналізу тональності текстових відгуків визначає порядок налаштування параметрів моделі з метою досягнення максимальної точності класифікації. У межах даної роботи навчання здійснюється на основі розміченої вибірки, що містить текстові відгуки та відповідні їм мітки тональності. Такий підхід належить до класу навчання з учителем і передбачає поступове вдосконалення моделі шляхом мінімізації помилки прогнозування. [27]

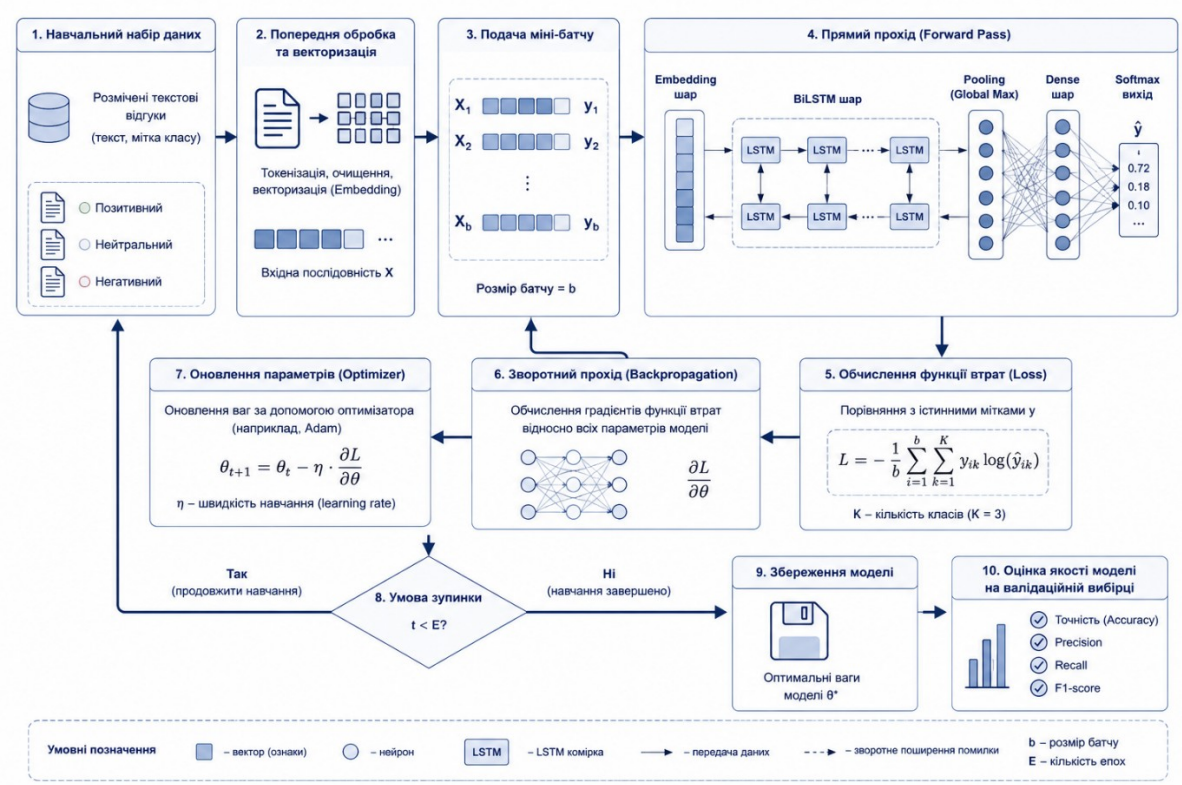


Рис 2.6 Алгоритм навчання нейромережевої моделі

На Рис 2.6 зображено математичну модель класифікації тональності тексту на основі нейромережевого підходу. Схема демонструє послідовність

перетворення текстового повідомлення від етапу попередньої обробки та векторизації до формування embedding-представлення і подальшої обробки нейромережевою моделлю BiLSTM. На завершальному етапі виконується класифікація тексту за допомогою функції Softmax, яка визначає ймовірність належності повідомлення до одного з класів тональності — позитивного, нейтрального або негативного. Схема також відображає процес формування вихідного вектора ймовірностей та прийняття остаточного рішення щодо класу текстового відгуку.

Процес навчання починається з формування навчальної та тестової вибірок. Вхідні дані попередньо обробляються та перетворюються у числове представлення, яке подається на вхід нейронної мережі. На кожній ітерації навчання модель формує прогноз для вхідних даних, після чого отриманий результат порівнюється з еталонними значеннями. Різниця між ними використовується для корекції внутрішніх параметрів моделі.

Оновлення параметрів виконується за допомогою методів оптимізації, що базуються на обчисленні градієнта функції помилки. У практичних реалізаціях широко застосовуються адаптивні алгоритми оптимізації, які дозволяють прискорити збіжність та підвищити стабільність навчання. У результаті багаторазового повторення цього процесу модель поступово наближається до такого стану, при якому її прогнози максимально відповідають реальним значенням. [30]

Оцінка якості моделі є невід’ємною частиною процесу навчання та дозволяє визначити ефективність її роботи на нових даних. Для цього використовується тестова або валідаційна вибірка, яка не бере участі у навчанні. Основною метою цього етапу є перевірка здатності моделі до узагальнення, тобто правильного прогнозування для раніше невідомих прикладів.

У задачах класифікації текстових даних одним із ключових показників якості є точність класифікації, яка визначається як відношення кількості правильно класифікованих прикладів до загальної кількості розглянутих об’єктів.

Даний показник дозволяє оцінити загальну ефективність моделі при виконанні задачі аналізу тональності. [26], [27]

Окрім точності, для більш детального аналізу можуть використовуватися додаткові метрики, такі як повнота, точність (precision) та F-міра, які дозволяють оцінити якість моделі з урахуванням особливостей розподілу класів. Це особливо важливо у випадках, коли навчальні дані є незбалансованими.

Важливим аспектом є контроль процесу навчання з метою запобігання перенавчанню. Для цього застосовуються такі підходи, як розділення даних на навчальну та валідаційну вибірки, моніторинг значення функції втрат та зупинка навчання при відсутності покращення результатів. Це дозволяє забезпечити стабільну роботу моделі та підвищити її здатність до узагальнення.

Таким чином, алгоритм навчання нейромережевої моделі базується на ітеративному коригуванні параметрів на основі помилки прогнозування, а оцінка якості здійснюється за допомогою відповідних метрик, що відображають ефективність класифікації. Застосування цих підходів дозволяє отримати модель, здатну з високою точністю визначати тональність текстових відгуків у практичних умовах.

3. Проєктні рішення та розробка програмного забезпечення

3.1 Архітектура нейромережевої системи АТВ

Архітектура програмної системи аналізу тональності текстових відгуків була розроблена з урахуванням необхідності автоматизованої обробки неструктурованих текстових даних, отриманих із зовнішніх джерел. Основною метою створення системи є забезпечення автоматичного визначення емоційного забарвлення текстових повідомлень користувачів, а також формування узагальненої оцінки змісту відгуків на основі сучасних методів обробки природної мови та нейромережевого аналізу. При проєктуванні архітектури особлива увага приділялася модульності, масштабованості та можливості інтеграції з зовнішніми джерелами даних, що дозволило реалізувати гнучку

структуру програмного забезпечення з можливістю подальшого розширення функціоналу системи.

Розроблена система побудована за модульним принципом та складається з декількох взаємопов’язаних компонентів, кожен із яких відповідає за окремий етап обробки інформації. Загальна логіка функціонування системи передбачає послідовне проходження текстового повідомлення через етапи отримання даних, попередньої обробки, перекладу тексту, нейромережевого аналізу, додаткової оцінки тональності на основі словникових правил та подальшого формування результатів аналізу. Такий підхід дозволяє поєднати переваги глибокого навчання та rule-based методів аналізу тексту, що позитивно впливає на точність визначення тональності відгуків. [2]

На першому етапі роботи системи виконується отримання вхідних текстових даних. Джерелом інформації можуть бути окремі текстові повідомлення користувача, набори відгуків у форматі CSV або дані, автоматично отримані з онлайн-платформ за допомогою API. У межах практичної реалізації системи було реалізовано підтримку отримання відгуків із платформи Steam через офіційний веб-інтерфейс API. Використання зовнішнього API дозволяє працювати з реальними текстовими відгуками користувачів, що значно підвищує практичну цінність розробленої системи та забезпечує можливість тестування нейромережевої моделі в умовах реальних даних. [3]

Після отримання текстових повідомлень виконується етап попередньої обробки тексту. Необхідність цього етапу пояснюється тим, що необроблені текстові дані можуть містити спеціальні символи, HTML-елементи, зайві пробіли, числові значення та інші компоненти, які негативно впливають на ефективність роботи нейромережевої моделі. У процесі попередньої обробки виконується очищення тексту від службових символів, нормалізація регістру символів, усунення надлишкових пробілів та приведення тексту до уніфікованого вигляду. Додатково реалізовано механізм токенізації тексту, що дозволяє перетворити

текстове повідомлення у послідовність окремих токенів для подальшої передачі до нейромережевої моделі.

Однією з важливих особливостей розробленої системи є підтримка багатомовного аналізу текстових повідомлень. Оскільки нейромережеву модель було навчено переважно на англомовному наборі даних, до архітектури системи було інтегровано модуль автоматичного перекладу тексту. Перед виконанням аналізу тональності текст автоматично перекладається англійською мовою за допомогою зовнішнього сервісу машинного перекладу. Реалізація такого підходу дозволяє виконувати аналіз україномовних, російськомовних та інших текстових повідомлень без необхідності окремого навчання моделі для кожної мови. Крім того, використання проміжного перекладу забезпечує універсальність системи та спрощує подальше масштабування програмного забезпечення.

Центральним компонентом архітектури є нейромережевий модуль аналізу тональності, побудований на основі двонапрямленої рекурентної нейронної мережі типу BiLSTM. Використання двонапрямленої архітектури дозволяє моделі враховувати контекст слів як у прямому, так і у зворотному напрямку речення, що суттєво підвищує ефективність аналізу природної мови. Нейромережа отримує на вхід числове представлення тексту після токенізації та виконує класифікацію повідомлення за трьома основними класами тональності: позитивна, нейтральна та негативна. Результатом роботи моделі є прогнозований клас тональності, який використовується для подальшого формування фінального результату аналізу.

З метою підвищення точності роботи системи до архітектури було додатково інтегровано rule-based модуль аналізу тональності. Принцип його роботи базується на використанні словників позитивно та негативно забарвлених слів і фраз. Під час аналізу система додатково перевіряє наявність у тексті емоційно виражених конструкцій, які можуть свідчити про позитивне або негативне ставлення користувача. Такий підхід дозволяє компенсувати окремі помилки нейромережевої моделі та підвищує стабільність роботи системи під час аналізу коротких або неоднозначних повідомлень. Крім того, реалізовано

механізм виявлення змішаної тональності тексту, коли повідомлення одночасно містить позитивні та негативні характеристики.

Фінальним етапом роботи системи є формування та візуалізація результатів аналізу. Для взаємодії користувача із системою було реалізовано графічний інтерфейс на основі програмної платформи Streamlit. Інтерфейс дозволяє вводити окремі текстові повідомлення, завантажувати набори відгуків та автоматично отримувати результати аналізу у зручному для користувача вигляді. Додатково система підтримує побудову графіків розподілу тональності відгуків, що забезпечує наочне представлення результатів роботи нейромережевої моделі та спрощує процес аналізу великих обсягів текстових даних.

Таким чином, розроблена архітектура нейромережевої системи аналізу тональності поєднує сучасні методи глибокого навчання, алгоритми обробки природної мови, rule-based підхід та засоби інтеграції із зовнішніми джерелами даних. Запропонована структура програмного забезпечення забезпечує достатній рівень гнучкості, масштабованості та ефективності, що дозволяє використовувати систему для автоматизованого аналізу реальних текстових відгуків у різних предметних областях.

3.2 Середовище розробки програмного забезпечення

Розробка нейромережевої системи аналізу тональності текстових відгуків потребувала використання сучасних програмних засобів, орієнтованих на реалізацію задач машинного навчання, обробки природної мови, аналізу текстових даних та створення інтерактивного користувацького інтерфейсу. Під час вибору середовища розробки основна увага приділялася таким характеристикам, як підтримка нейромережевих бібліотек, наявність інструментів для роботи з текстовими даними, можливість інтеграції із зовнішніми сервісами та простота створення прикладного програмного забезпечення. У результаті для реалізації програмної системи було обрано стек технологій, який забезпечує ефективну взаємодію між компонентами системи та дозволяє реалізувати повний

цикл обробки текстових даних — від отримання відгуків до візуалізації результатів аналізу.

Для забезпечення модульності, масштабованості та зручності обробки даних у розробленій системі було реалізовано багаторівневу архітектуру, що включає модулі отримання, попередньої обробки, аналізу та візуалізації текстових відгуків. Загальна структура системи наведена на Рис 3.1.

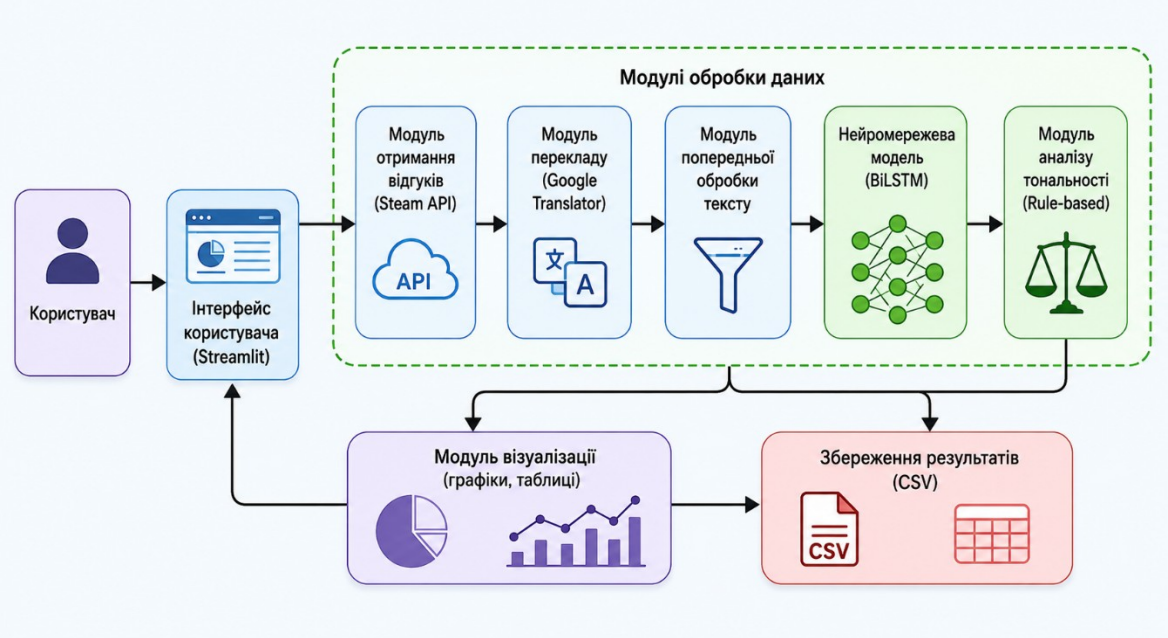


Рис 3.1 Архітектура системи АТВ

Основною мовою програмування під час розробки системи було обрано мову Python. Дане рішення обумовлене широким поширенням Python у сфері машинного навчання, аналізу даних та обробки природної мови. Важливою перевагою Python є наявність великої кількості спеціалізованих бібліотек для побудови нейромережевих моделей, роботи з API, аналізу текстів та візуалізації даних. Крім того, мова характеризується зрозумілим синтаксисом, високою швидкістю розробки програмного забезпечення та активною підтримкою з боку міжнародної спільноти розробників. Завдяки цьому Python широко використовується під час створення сучасних систем штучного інтелекту та програмних рішень у сфері обробки природної мови. [17]

Для реалізації нейромережевої моделі аналізу тональності було використано бібліотеку PyTorch. Даний фреймворк є одним із найбільш популярних засобів

розробки систем глибокого навчання та забезпечує гнучкі механізми створення нейронних мереж різної складності. Використання PyTorch дозволило реалізувати двонапрямлену рекурентну нейронну мережу типу BiLSTM, що застосовується для аналізу послідовностей текстових даних. Важливою перевагою фреймворку є підтримка динамічних обчислювальних графів, що значно спрощує процес налагодження та модифікації архітектури нейромережі. Крім того, PyTorch забезпечує високу швидкодію обчислень, підтримує використання графічних процесорів та містить вбудовані інструменти для навчання й тестування моделей машинного навчання. [18]

Для підготовки текстових даних та формування числового представлення тексту було використано інструменти бібліотек TensorFlow та Keras. Зокрема, у системі було реалізовано механізм токенізації тексту, який дозволяє перетворити текстові повідомлення у послідовність числових індексів для подальшої передачі до нейромережевої моделі. Додатково використовувались засоби автоматичного вирівнювання довжини текстових послідовностей, що забезпечило уніфіковане представлення вхідних даних під час навчання та виконання прогнозування. [17], [18]

Для роботи з наборами даних та структурованою інформацією використовувалась бібліотека Pandas. Застосування Pandas дозволило ефективно виконувати завантаження CSV-файлів, обробку текстових наборів даних, формування таблиць результатів та збереження результатів аналізу. Дана бібліотека забезпечує зручні засоби роботи з табличними структурами даних, що значно спрощує виконання операцій фільтрації, групування та статистичного аналізу інформації. Крім того, Pandas активно використовується під час підготовки навчальних вибірок для моделей машинного навчання. [21]

Візуалізація результатів роботи системи реалізована за допомогою бібліотеки Matplotlib. Використання засобів візуалізації дозволило реалізувати побудову гістограм та кругових діаграм розподілу тональності текстових відгуків. Застосування графічного представлення результатів аналізу суттєво покращує

сприйняття статистичної інформації та спрощує аналіз великих наборів текстових даних. Побудовані графіки використовуються як у процесі тестування нейромережевої моделі, так і під час демонстрації роботи системи користувачу. [22]

Для реалізації графічного інтерфейсу користувача було використано програмний фреймворк Streamlit. Даний інструмент забезпечує можливість швидкого створення інтерактивних вебзастосунків без необхідності використання складних frontend-технологій. За допомогою Streamlit у системі було реалізовано форму введення тексту, інтерактивні елементи керування, візуалізацію результатів аналізу, побудову графіків та механізм завантаження результатів у форматі CSV. Використання даного фреймворку дозволило суттєво скоротити час розробки користувацького інтерфейсу та забезпечити просту взаємодію користувача із системою аналізу тональності. [23]

Особливу роль у роботі системи відіграє механізм інтеграції із зовнішніми сервісами. Для отримання реальних текстових відгуків було реалізовано інтеграцію з офіційним Steam API, що дозволило автоматизувати процес завантаження користувацьких коментарів із платформи Steam. Отримання даних через API забезпечує роботу системи з актуальними текстовими повідомленнями та значно підвищує практичну цінність розробленого програмного забезпечення. Крім того, для підтримки багатомовного аналізу тексту було інтегровано сервіс автоматичного перекладу GoogleTranslator, який забезпечує автоматичний переклад текстових повідомлень англійською мовою перед виконанням нейромережевого аналізу. [24], [25]

Безпосередня розробка програмного забезпечення виконувалася у середовищі PyCharm. Використання даного інтегрованого середовища розробки дозволило забезпечити зручне редагування програмного коду, автоматичне підсвічування синтаксису, підтримку системи керування пакетами Python та засоби налагодження програмного забезпечення. Додатковою перевагою PyCharm

є можливість інтеграції з системами контролю версій та підтримка запуску локальних серверів для тестування вебзастосунків Streamlit.

Таким чином, обране середовище розробки поєднує сучасні програмні засоби для реалізації задач машинного навчання, обробки природної мови, аналізу текстових даних та створення інтерактивного користувацького інтерфейсу. Використання наведеного стеку технологій забезпечило ефективну реалізацію нейромережевої системи аналізу тональності текстових відгуків та створило основу для подальшого розвитку й масштабування програмного забезпечення.

3.3 Проєктування програмних модулів системи

Проєктування програмних модулів нейромережевої системи аналізу тональності текстових відгуків виконувалося з урахуванням принципів модульності, масштабованості та повторного використання програмного коду. Використання модульної архітектури дозволило розділити систему на окремі функціональні компоненти, кожен із яких відповідає за реалізацію конкретного етапу обробки даних. Такий підхід суттєво спрощує процес супроводу програмного забезпечення, дозволяє незалежно модифікувати окремі елементи системи та забезпечує можливість подальшого розширення функціоналу без необхідності повної зміни структури застосунку.

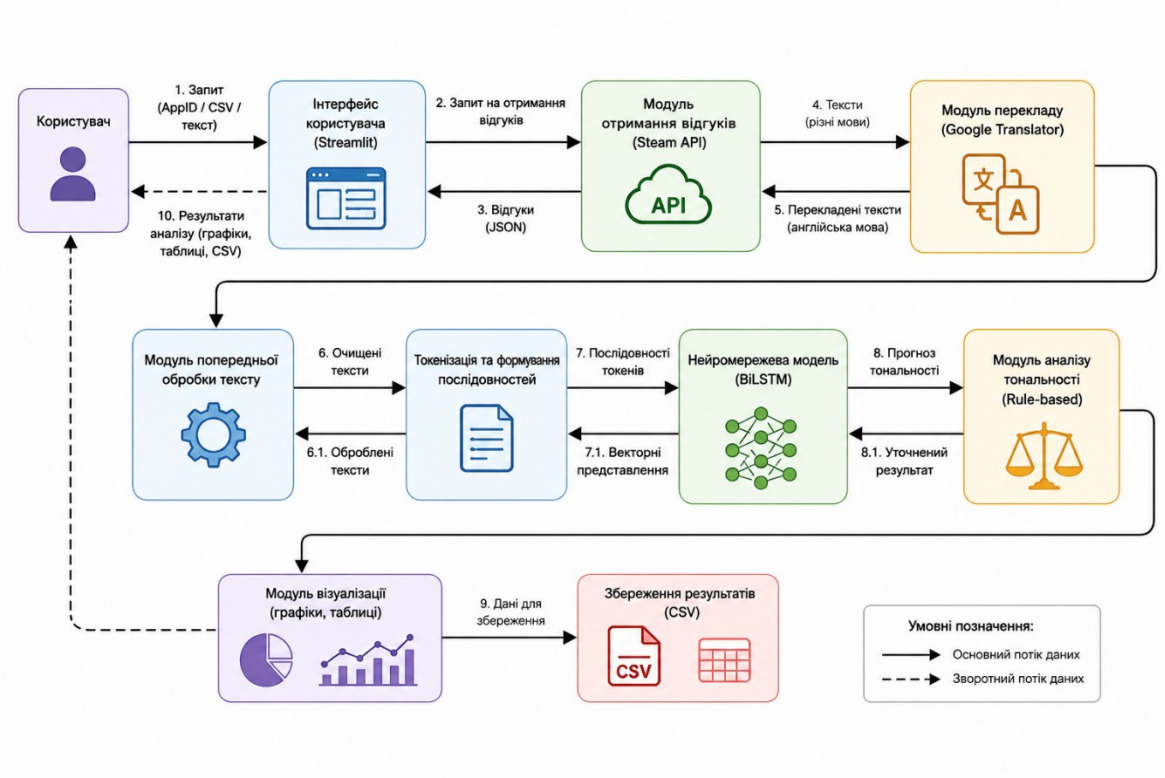


Рис 3.2 Діаграма потоків даних (DFD)

Як показано на Рис 3.2, основний потік даних починається із введення користувачем AppID або текстового повідомлення через вебінтерфейс системи. Після отримання відгуків із Steam API текстові дані проходять етап автоматичного перекладу та попередньої обробки, включаючи очищення тексту та токенізацію. Далі підготовлені дані передаються до нейромережевої моделі BiLSTM для визначення емоційної тональності повідомлень. Результати аналізу використовуються для побудови графіків, таблиць та збереження статистичної інформації у форматі CSV.

У процесі проєктування програмного забезпечення система була розділена на декілька основних модулів: модуль попередньої обробки тексту, модуль нейромережевого аналізу, модуль rule-based аналізу, модуль інтеграції із зовнішніми сервісами, модуль візуалізації результатів та модуль графічного інтерфейсу користувача. Кожен із наведених компонентів виконує окремий набір функцій та взаємодіє з іншими модулями через передачу підготовлених даних. Завдяки цьому забезпечується послідовна обробка текстових повідомлень та формування кінцевого результату аналізу тональності.

Одним із ключових компонентів системи є модуль попередньої обробки тексту. Основним призначенням даного модуля є підготовка текстових повідомлень до подальшого аналізу нейромережевою моделлю. У межах модуля реалізовано очищення тексту від спеціальних символів, HTML-конструкцій, числових значень та надлишкових пробілів. Додатково виконується нормалізація регістру символів та токенізація текстових повідомлень. Після завершення обробки текст перетворюється у послідовність токенів, які надалі використовуються для формування числового представлення вхідних даних. Реалізація окремого модуля попередньої обробки дозволяє забезпечити уніфіковану структуру текстових повідомлень та підвищує ефективність роботи нейромережевої моделі.

Наступним важливим компонентом системи є модуль нейромережевого аналізу тональності. Даний модуль реалізований у вигляді окремого програмного компонента, який відповідає за завантаження навченої моделі BiLSTM, виконання прогнозування та формування результатів класифікації тексту. У процесі роботи модуль отримує підготовлену текстову послідовність, перетворює її у числове представлення та передає до нейронної мережі. Результатом роботи модуля є визначення одного з трьох класів тональності: позитивної, нейтральної або негативної. Для забезпечення стабільної роботи системи модуль також реалізує механізм автоматичного завантаження ваг нейромережевої моделі та перевірку коректності структури вхідних даних.

Окрему роль у структурі системи виконує модуль rule-based аналізу тональності. Необхідність реалізації даного компонента пояснюється тим, що нейромережеві моделі можуть допускати помилки під час аналізу коротких або неоднозначних текстових повідомлень. З метою підвищення точності класифікації було реалізовано механізм додаткової перевірки тексту на наявність позитивно або негативно забарвлених слів та фраз. У межах модуля використовуються спеціальні словники емоційно забарвлених конструкцій, які дозволяють виявляти явні ознаки позитивного або негативного змісту тексту. Крім того, rule-based

модуль забезпечує підтримку аналізу змішаної тональності повідомлень, що особливо важливо під час роботи з реальними користувацькими відгуками.

Для забезпечення роботи системи з реальними текстовими даними було реалізовано модуль інтеграції із зовнішніми сервісами. Основною функцією даного компонента є автоматичне отримання текстових відгуків із зовнішніх платформ за допомогою API. У межах практичної реалізації системи було інтегровано Steam API, що дозволяє автоматично завантажувати користувацькі відгуки до комп’ютерних ігор без необхідності ручного введення даних. Отримані повідомлення передаються до модулів попередньої обробки та нейромережевого аналізу для подальшого визначення тональності. Додатково у системі реалізовано інтеграцію із сервісом автоматичного перекладу тексту, що забезпечує підтримку багатомовного аналізу повідомлень.

Важливим компонентом програмного забезпечення є модуль візуалізації результатів. Основним призначенням даного модуля є представлення результатів аналізу тональності у графічному вигляді. У процесі роботи система автоматично формує гістограми та кругові діаграми розподілу тональності текстових повідомлень. Крім того, реалізовано механізм формування таблиць результатів аналізу та можливість експорту даних у формат CSV. Використання засобів візуалізації дозволяє значно покращити сприйняття результатів роботи системи та спрощує аналіз великих обсягів текстових даних. [21]

Для забезпечення взаємодії користувача із системою було реалізовано окремий модуль графічного інтерфейсу користувача. Інтерфейс системи створено за допомогою фреймворку Streamlit, який дозволяє швидко розробляти інтерактивні вебзастосунки на основі мови Python. Реалізований інтерфейс містить поле введення тексту, засоби завантаження відгуків із Steam API, елементи керування параметрами аналізу та блоки відображення статистичної інформації. Додатково користувач має можливість переглядати результати аналізу у вигляді таблиць та графіків, а також виконувати експорт результатів для подальшої обробки. [22]

Для реалізації автоматизованого аналізу текстових відгуків у системі було реалізовано послідовний алгоритм обробки текстових даних. Загальну блок-схему алгоритму роботи системи наведено на Рис 3.3.

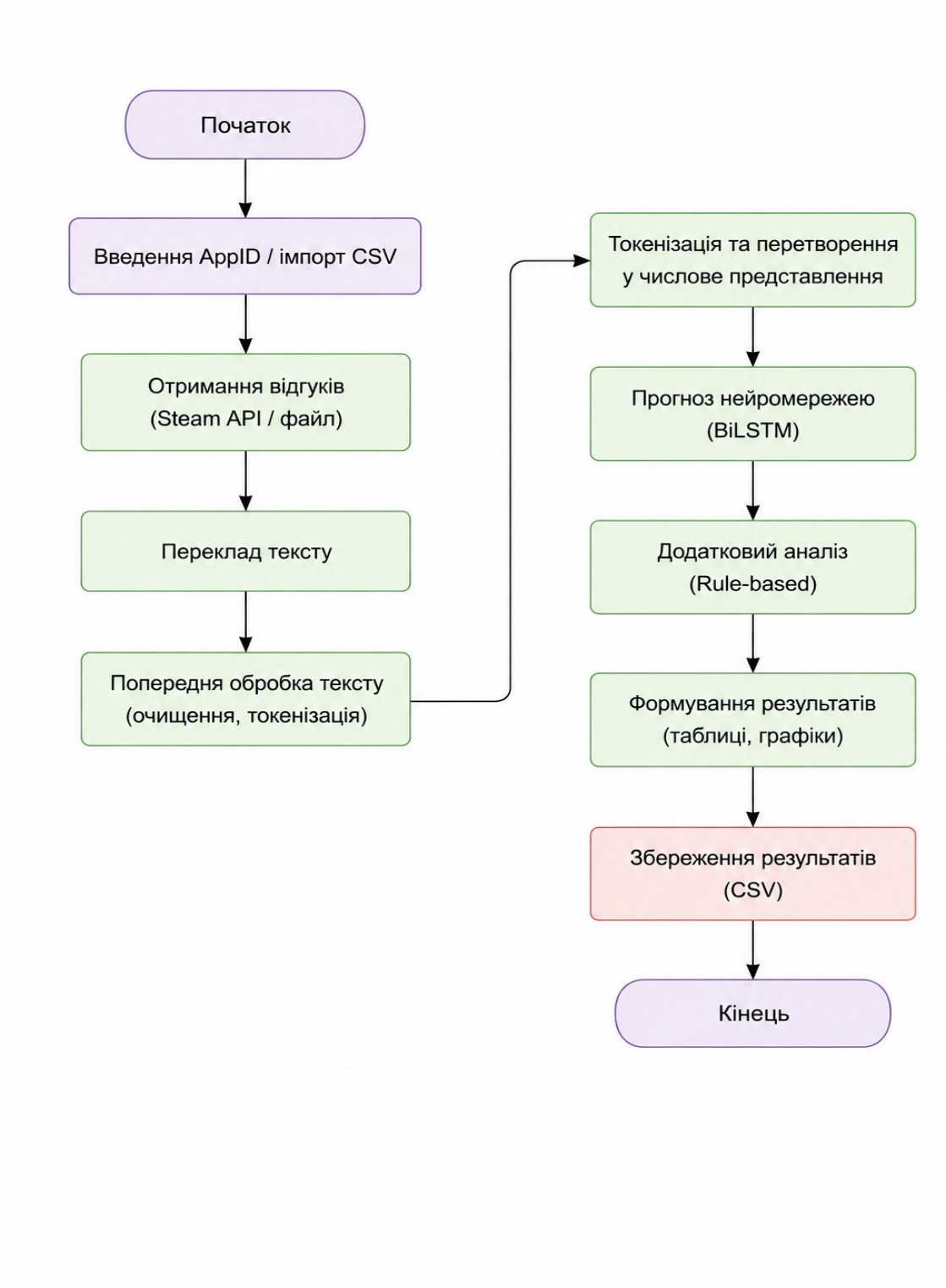


Рис 3.3 Блок схема алгоритму АТВ

У процесі проєктування програмних модулів особлива увага приділялася забезпеченню взаємодії між компонентами системи. Для передачі даних між модулями використовувалися стандартизовані структури даних, що дозволило забезпечити узгоджену роботу всіх елементів програмного забезпечення. Такий підхід спрощує тестування окремих компонентів системи та забезпечує можливість подальшого масштабування програмного забезпечення шляхом додавання нових функціональних модулів.

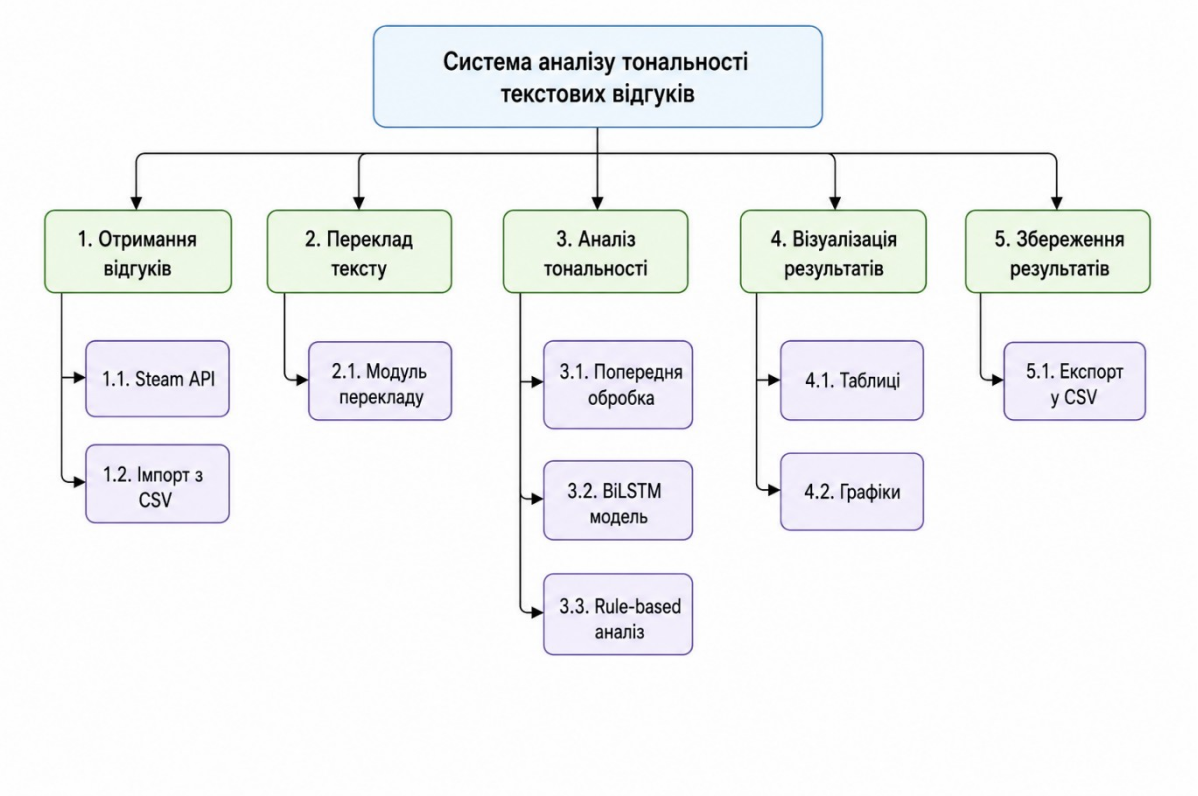


Рис 3.4 Дерево функцій системи

Як показано на Рис 3.4, система складається з декількох основних функціональних підсистем, серед яких модулі отримання відгуків, перекладу тексту, аналізу тональності, візуалізації результатів та збереження даних. Кожна підсистема реалізує окремий набір функцій, необхідних для забезпечення повного циклу обробки текстових повідомлень. Центральним елементом системи є модуль аналізу тональності, який включає механізми попередньої обробки тексту, нейромережеву модель BiLSTM та rule-based аналіз для уточнення результатів класифікації.

Таким чином, проєктування програмних модулів системи аналізу тональності виконано відповідно до сучасних принципів побудови програмного забезпечення для задач штучного інтелекту та обробки природної мови. Реалізована модульна структура забезпечує гнучкість, масштабованість та зручність супроводу системи, а також створює основу для подальшого розвитку функціональних можливостей програмного забезпечення.

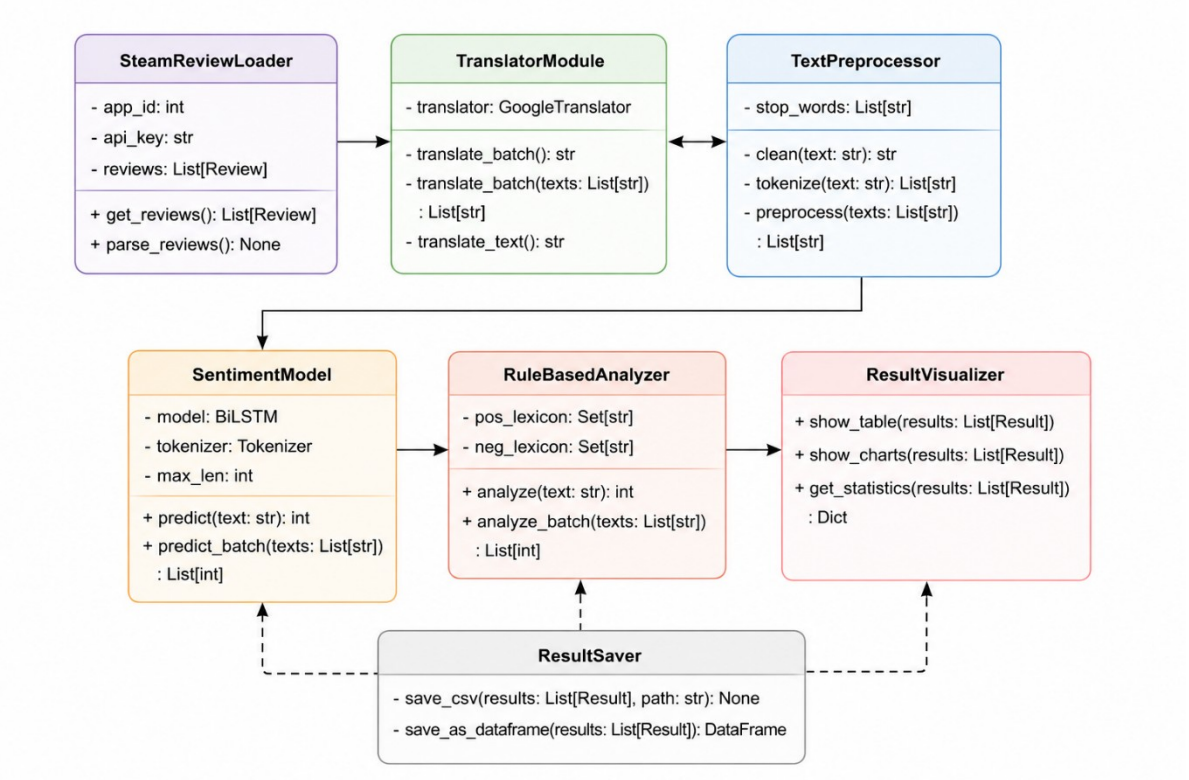


Рис 3.5 Діаграма класів системи (UML)

Як показано на Рис 3.5, програмна система складається з декількох основних класів, що реалізують окремі функціональні підсистеми. Клас SteamReviewLoader відповідає за отримання текстових відгуків через Steam API, TranslatorModule реалізує автоматичний переклад тексту, а TextPreprocessor виконує очищення та токенізацію текстових даних. Основний аналіз тональності виконується класом SentimentModel, який реалізує нейромережеву модель BiLSTM. Додатково у системі використовується клас RuleBasedAnalyzer для rule-based аналізу текстових повідомлень та уточнення результатів класифікації. Формування таблиць і графіків виконується класом ResultVisualizer, а збереження результатів аналізу забезпечується класом ResultSaver.

3.4 Реалізація нейромережевої моделі аналізу тональності

Реалізація нейромережевої моделі аналізу тональності є одним із ключових етапів розробки програмної системи, оскільки саме цей компонент забезпечує автоматичне визначення емоційного забарвлення текстових повідомлень користувачів. У межах виконання роботи було реалізовано модель глибокого навчання на основі двонапрямленої рекурентної нейронної мережі типу BiLSTM, яка забезпечує ефективну обробку послідовностей текстових даних та враховує контекст слів у межах речення. Використання даного підходу дозволило підвищити точність класифікації текстів та забезпечити стабільну роботу системи під час аналізу реальних користувацьких відгуків.

Перед початком навчання нейромережевої моделі було виконано підготовку набору даних. Для навчання системи використовувався датасет текстових відгуків, що містить приклади позитивних, нейтральних та негативних повідомлень. Кожному текстовому повідомленню у наборі даних відповідає мітка класу тональності, яка використовується під час навчання моделі. Оскільки вхідні текстові дані представлені у символьному форматі, перед навчанням нейромережі було виконано процедуру токенізації тексту та перетворення слів у числові індекси. Додатково застосовувалося вирівнювання довжини текстових послідовностей, що дозволило сформувати уніфікований формат вхідних даних для нейромережевої моделі.

Одним із важливих етапів реалізації системи стало формування словника токенів. У процесі токенізації найбільш уживані слова текстового корпусу отримували унікальні числові індекси, які надалі використовувалися для формування числового представлення тексту. Для обмеження складності моделі та зменшення обчислювального навантаження використовувався фіксований розмір словника. Крім того, у процесі реалізації моделі було застосовано механізм автоматичного доповнення текстових послідовностей до однакової довжини, що забезпечило коректну передачу даних до нейромережі під час навчання та прогнозування.

Архітектура реалізованої нейромережевої моделі складається з декількох послідовних шарів. Першим етапом обробки є embedding layer, основним призначенням якого є перетворення числових індексів слів у векторне представлення фіксованої розмірності. Використання embedding-представлень дозволяє моделі виявляти семантичні зв’язки між словами та покращує якість аналізу тексту. Після формування векторного представлення текстові дані передаються до двонапрямленого рекурентного шару BiLSTM, який виконує аналіз текстової послідовності у прямому та зворотному напрямках. Такий підхід дозволяє враховувати контекст слів не лише відносно попередніх елементів речення, але й відносно наступних слів, що суттєво підвищує ефективність аналізу природної мови.

Для класифікації результатів аналізу використовується повнозв’язний шар нейронної мережі, який формує ймовірності належності текстового повідомлення до одного з класів тональності. У системі реалізовано три основні класи: позитивна, нейтральна та негативна тональність. Остаточне рішення щодо належності тексту до певного класу приймається на основі максимального значення ймовірності, отриманої після проходження вхідних даних через нейромережеву модель.

Під час навчання моделі використовувався алгоритм оптимізації Adam, який є одним із найбільш ефективних методів оптимізації нейронних мереж у задачах машинного навчання. Для оцінювання помилки класифікації застосовувалася функція втрат CrossEntropyLoss, що широко використовується у задачах багатокласової класифікації текстових даних. Навчання моделі виконувалося протягом декількох епох із поступовим зменшенням значення функції втрат, що свідчило про покращення здатності нейромережі виконувати класифікацію текстових повідомлень.

У процесі практичного тестування було встановлено, що використання виключно нейромережевого підходу не завжди забезпечує достатню точність аналізу коротких або емоційно неоднозначних повідомлень. Зокрема, модель

могла помилково визначати окремі негативні висловлювання як нейтральні через недостатню кількість подібних прикладів у навчальному наборі даних. Для підвищення ефективності роботи системи було реалізовано додатковий rule-based механізм аналізу тональності, який використовує словники позитивно та негативно забарвлених слів і фраз. Комбінування нейромережевого та словникового підходів дозволило підвищити стабільність роботи системи та покращити результати аналізу коротких текстових повідомлень.

Важливою особливістю реалізованої системи є підтримка багатомовного аналізу тексту. Оскільки нейромережеву модель було навчено переважно на англомовному наборі даних, до структури програмного забезпечення було інтегровано механізм автоматичного перекладу тексту англійською мовою перед виконанням аналізу тональності. Реалізація такого підходу дозволила забезпечити підтримку аналізу україномовних та російськомовних текстових повідомлень без необхідності окремого навчання нейромережевої моделі для кожної мови.

Ни Рис 3.6 зображено результат процесу тренування нейромережевої моделі. Кількість епох, втрати даних та точність.

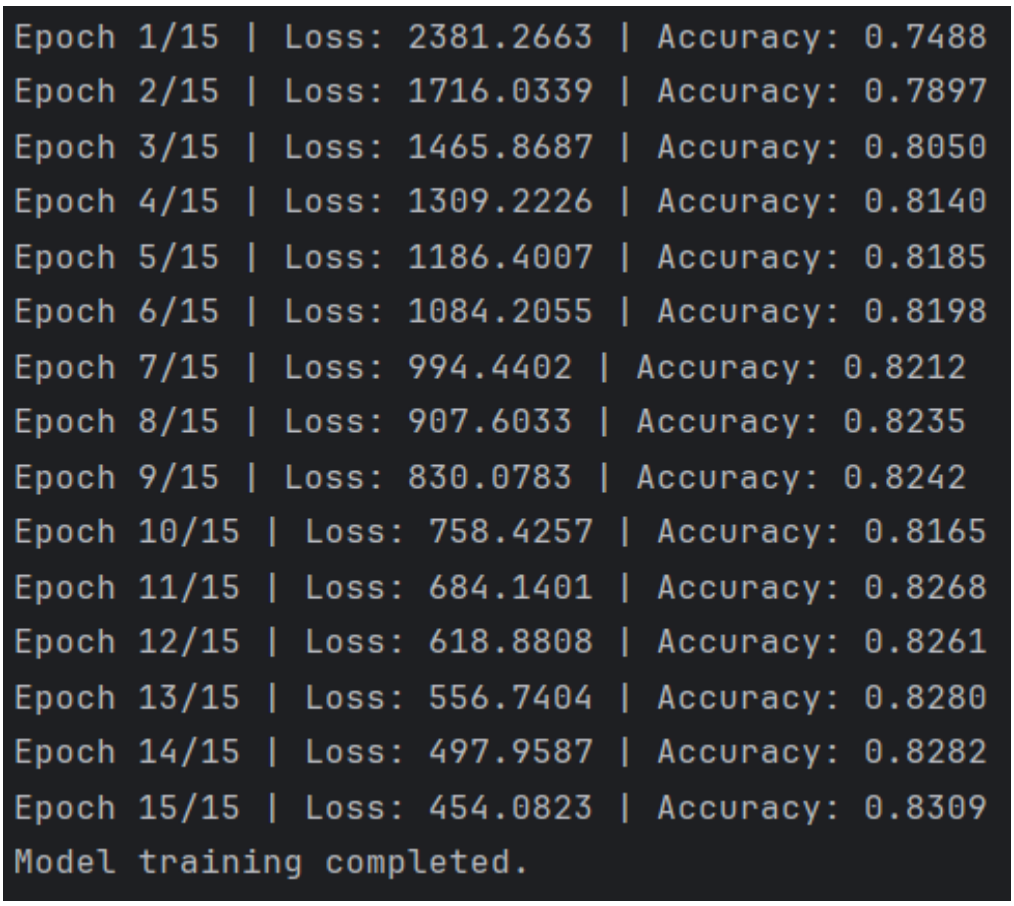


Рис 3.6 Результат навчання нейромережевої моделі

Для збереження навченої моделі використовувався механізм серіалізації ваг нейромережі у файл формату .pth. Завдяки цьому система отримала можливість повторного використання навченої моделі без необхідності повторного запуску процесу навчання. Під час запуску програмного забезпечення виконується автоматичне завантаження ваг моделі, після чого система готова до виконання аналізу текстових повідомлень у режимі реального часу.

Таким чином, реалізована нейромережа на основі архітектури BiLSTM забезпечує ефективне визначення тональності текстових повідомлень та дозволяє виконувати аналіз реальних користувацьких відгуків. Поєднання методів глибокого навчання, rule-based аналізу та автоматичного перекладу тексту дозволило створити гнучку систему аналізу тональності, придатну для роботи з багатомовними текстовими даними та інтеграції із зовнішніми інформаційними сервісами.

3.5 Реалізація програмного інтерфейсу користувача

Одним із важливих етапів розробки нейромережевої системи аналізу тональності текстових відгуків стала реалізація програмного інтерфейсу користувача. Основним призначенням інтерфейсу є забезпечення зручної взаємодії користувача із системою, спрощення процесу введення текстових даних, запуску аналізу та перегляду результатів роботи нейромережевої моделі. Під час проєктування інтерфейсу особлива увага приділялася простоті використання, наочності представлення результатів та можливості інтеграції з іншими функціональними компонентами програмного забезпечення.

Для реалізації користувацького інтерфейсу було використано фреймворк Streamlit, який дозволяє створювати інтерактивні вебзастосунки на основі мови Python без необхідності використання складних frontend-технологій. Використання Streamlit суттєво спростило процес створення графічного інтерфейсу та забезпечило можливість швидкої інтеграції нейромережевої моделі із засобами візуалізації результатів аналізу. Крім того, даний фреймворк підтримує автоматичне оновлення елементів інтерфейсу та забезпечує зручне відображення графіків, таблиць і текстових повідомлень.

Головне вікно розробленого застосунку містить елементи керування, необхідні для виконання аналізу текстових відгуків. У межах інтерфейсу реалізовано поле введення Steam AppID, яке дозволяє користувачу автоматично завантажувати текстові відгуки для обраного програмного продукту або комп’ютерної гри. Додатково користувач має можливість вказувати кількість відгуків, що підлягають аналізу, а також запускати процес аналізу за допомогою спеціальної кнопки керування. Після запуску аналізу система автоматично отримує відгуки через Steam API, виконує їх попередню обробку та передає текстові дані до нейромережевої моделі для визначення тональності.

На Рис 3.7 зображено інтерфейс головного вікна програми. В ньому можна визначити відгуки якого самого продукту в Steam будуть аналізовані та їх кількість.

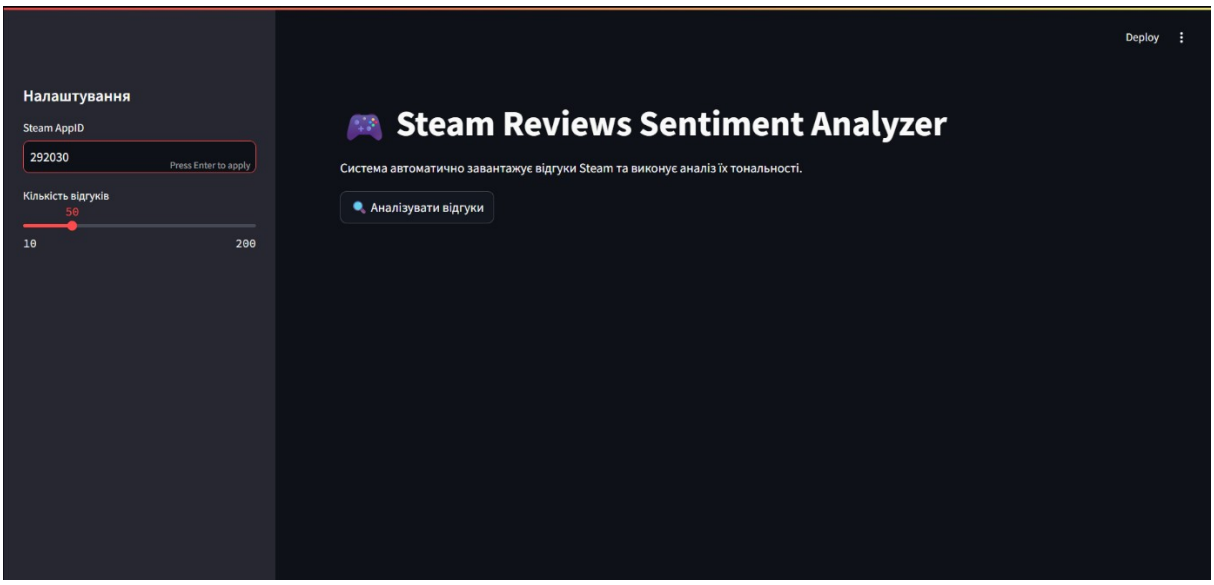


Рис 3.7 Головне вікно

У процесі виконання аналізу інтерфейс відображає службову інформацію про поточний стан роботи системи. Для цього було реалізовано механізм індикатора прогресу, який дозволяє користувачу контролювати процес обробки текстових повідомлень у режимі реального часу. Наявність такого механізму підвищує зручність використання програмного забезпечення та забезпечує кращу взаємодію користувача із системою під час аналізу великих наборів текстових даних.

Однією з ключових особливостей реалізованого інтерфейсу є підтримка графічної візуалізації результатів аналізу тональності. У системі автоматично формуються гістограми та кругові діаграми, які відображають співвідношення позитивних, нейтральних та негативних відгуків. Побудова графіків здійснюється за допомогою бібліотеки Matplotlib, інтегрованої із Streamlit. Використання графічного представлення результатів дозволяє значно спростити аналіз великих наборів текстових повідомлень та забезпечує наочне відображення статистичної інформації.

На Рис 3.8 зображено гістограми, стовпчасту та кругову, аналізованих відгуків. За їх допомогою можна продивитися кількість відгуків кожної тональності та їх відсоткове відношення.

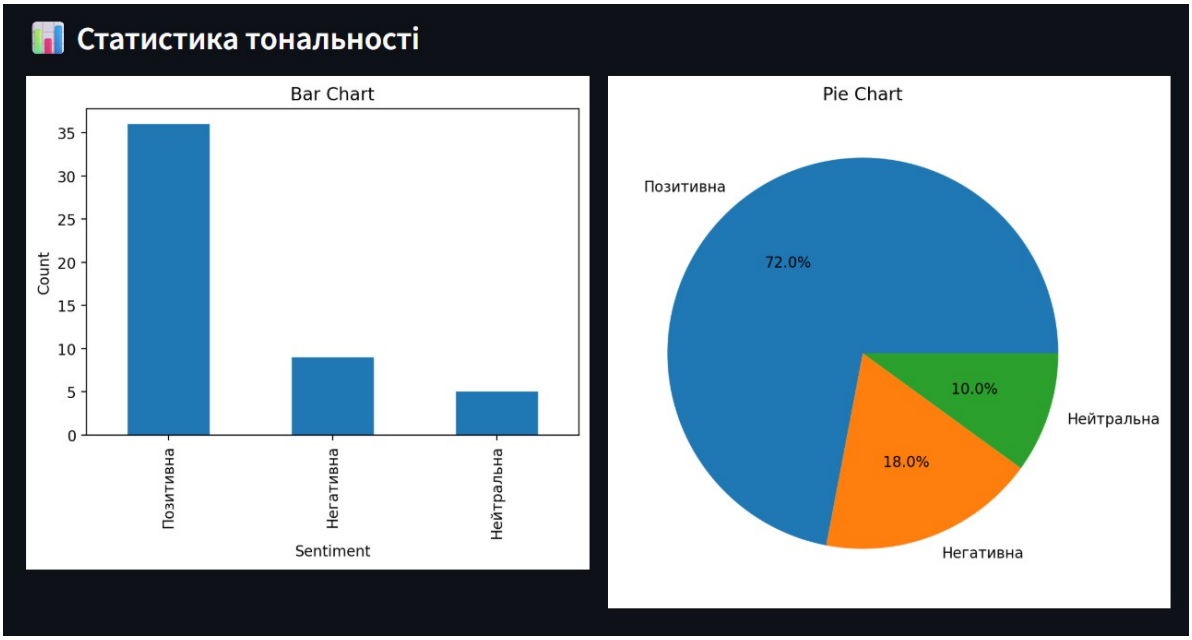
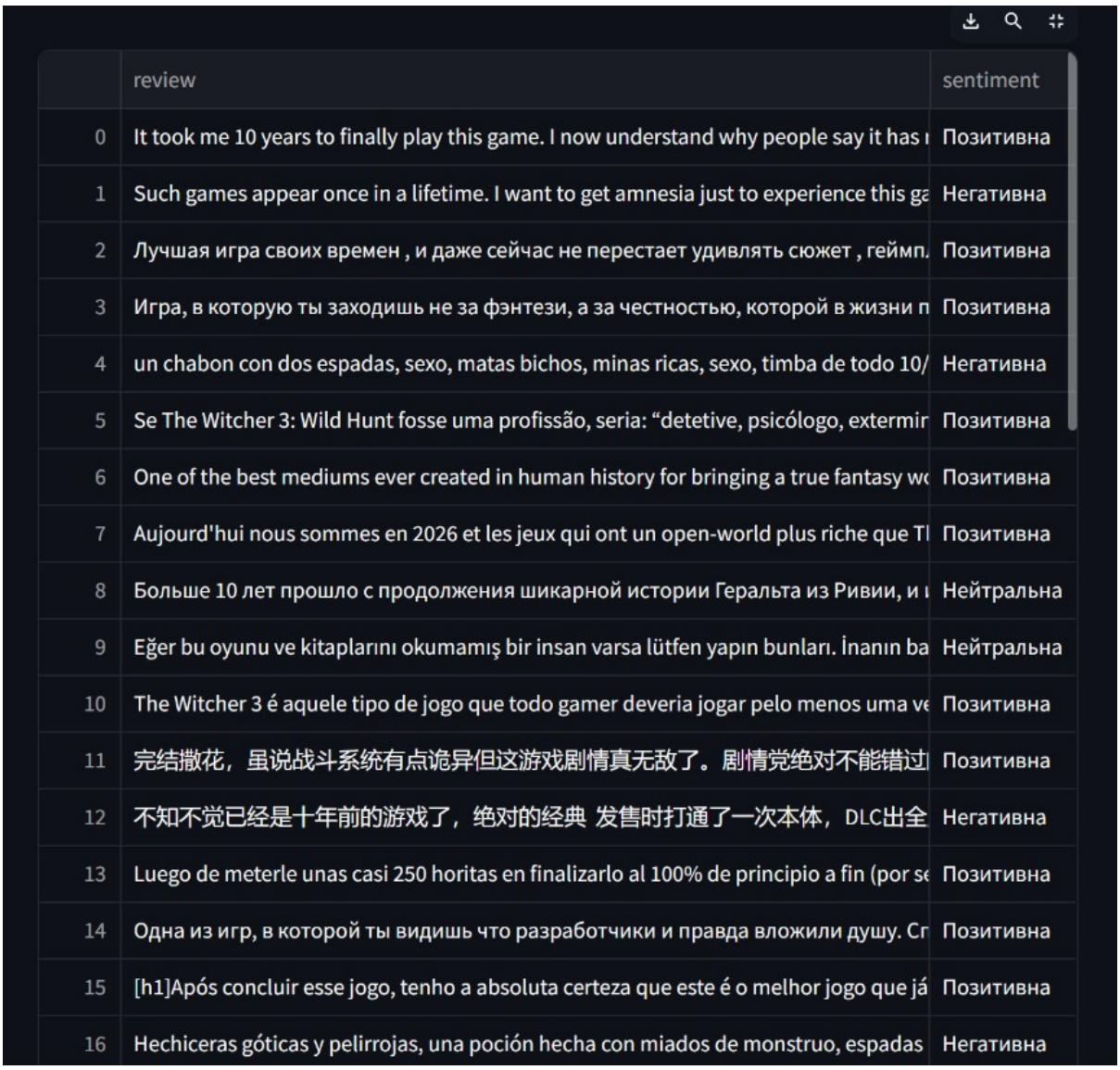


Рис 3.8 Гістограми

Крім графічної візуалізації, інтерфейс системи підтримує табличне представлення результатів аналізу. Після завершення обробки текстових повідомлень користувач отримує таблицю, яка містить оригінальні відгуки та визначену для кожного повідомлення тональність. Реалізація табличного представлення даних дозволяє швидко переглядати результати аналізу та виконувати додатковий контроль коректності роботи нейромережевої моделі.

На Рис 3.9 зображено табличне представлення відгуків. Воно дозволяє дізнатись яку саме тональність отримав окремий відгук.



	review	sentiment
0	It took me 10 years to finally play this game. I now understand why people say it has	Позитивна
1	Such games appear once in a lifetime. I want to get amnesia just to experience this ga	Негативна
2	Лучшая игра своих времен , и даже сейчас не перестает удивлять сюжет , геймп.	Позитивна
3	Игра, в которую ты заходишь не за фэнтези, а за честностью, которой в жизни п	Позитивна
4	un chabon con dos espadas, sexo, matas bichos, minas ricas, sexo, timba de todo 10/	Негативна
5	Se The Witcher 3: Wild Hunt fosse uma profissão, seria: “detetive, psicólogo, extermir	Позитивна
6	One of the best mediums ever created in human history for bringing a true fantasy wo	Позитивна
7	Aujourd'hui nous sommes en 2026 et les jeux qui ont un open-world plus riche que Tl	Позитивна
8	Больше 10 лет прошло с продолжения шикарной истории Геральта из Ривии, и	Нейтральна
9	Eğer bu oyunu ve kitaplarını okumamış bir insan varsa lütfen yapın bunları. İnanın ba	Нейтральна
10	The Witcher 3 é aquele tipo de jogo que todo gamer deveria jogar pelo menos uma ve	Позитивна
11	完结撒花，虽说战斗系统有点诡异但这游戏剧情真无敌了。剧情党绝对不能错过	Позитивна
12	不知不觉已经是十年前的游戏了，绝对的经典 发售时打通了一次本体，DLC出全	Негативна
13	Luego de meterle unas casi 250 horitas en finalizarlo al 100% de principio a fin (por se	Позитивна
14	Одна из игр, в которой ты видишь что разработчики и правда вложили душу. Сг	Позитивна
15	[h1]Após concluir esse jogo, tenho a absoluta certeza que este é o melhor jogo que já	Позитивна
16	Hechiceras góticas y pelirrojas, una poción hecha con miados de monstruo, espadas	Негативна

Рис 3.9 Табличне представлення

Додатковою функціональною можливістю інтерфейсу є підтримка експорту результатів аналізу у формат CSV. Після завершення роботи системи користувач може завантажити сформований файл із результатами аналізу для подальшої обробки або використання у сторонніх програмних засобах. Реалізація механізму експорту даних суттєво підвищує практичну цінність розробленої системи та забезпечує можливість інтеграції результатів аналізу із зовнішніми інформаційними системами.

Під час реалізації інтерфейсу особлива увага приділялася забезпеченню простоти використання програмного забезпечення. Всі основні елементи керування розташовані у логічній послідовності, що дозволяє користувачу

швидко виконувати аналіз текстових повідомлень без необхідності попереднього вивчення структури системи. Крім того, Streamlit автоматично забезпечує адаптацію вебінтерфейсу до різних розмірів екрана, що покращує зручність використання застосунку на різних типах пристроїв.

У процесі тестування було встановлено, що реалізований інтерфейс забезпечує стабільну взаємодію з нейромережевою моделлю та дозволяє ефективно виконувати аналіз великих обсягів текстових даних у режимі реального часу. Інтеграція із Steam API, підтримка графічної візуалізації та можливість автоматичного експорту результатів значно розширюють функціональні можливості системи та підвищують рівень її практичного застосування.

Таким чином, реалізований програмний інтерфейс користувача забезпечує ефективну взаємодію між користувачем та нейромережевою системою аналізу тональності текстових відгуків. Використання сучасних засобів веброзробки дозволило створити зручний інтерактивний застосунок, який підтримує автоматичний аналіз текстових повідомлень, графічне представлення результатів та інтеграцію із зовнішніми джерелами даних.

3.6 Тестування та аналіз результатів роботи системи

Заключним етапом розробки нейромережевої системи аналізу тональності текстових відгуків стало проведення тестування програмного забезпечення та аналіз результатів роботи моделі. Основною метою тестування була перевірка коректності функціонування всіх програмних модулів системи, оцінка якості визначення тональності текстових повідомлень, а також аналіз стабільності роботи програмного забезпечення під час обробки реальних користувацьких відгуків. Особлива увага приділялася перевірці роботи системи в умовах аналізу текстів різної довжини, стилістики та емоційного забарвлення.

Тестування системи виконувалося поетапно. На першому етапі проводилася перевірка коректності роботи модулів попередньої обробки тексту. Зокрема, тестувалася робота механізмів очищення тексту від спеціальних символів,

нормалізації регістру, токенізації та формування числового представлення текстових повідомлень. У процесі тестування було встановлено, що реалізовані алгоритми забезпечують стабільну підготовку текстових даних до подальшої передачі у нейромережеву модель та дозволяють коректно обробляти текстові повідомлення різного формату.

Наступним етапом тестування стала перевірка роботи нейромережевої моделі аналізу тональності. Для оцінювання якості класифікації використовувався набір тестових текстових повідомлень, що містив позитивні, негативні та нейтральні відгуки. У процесі тестування оцінювалася здатність моделі правильно визначати емоційне забарвлення тексту та стабільно працювати з повідомленнями різної складності. Під час навчання моделі спостерігалося поступове зменшення значення функції втрат, що свідчило про покращення здатності нейромережі до класифікації текстових повідомлень.

У ході практичного тестування було встановлено, що нейромережева модель демонструє достатньо високий рівень точності під час аналізу текстових повідомлень із чітко вираженим позитивним або негативним змістом. Наприклад, система успішно класифікувала повідомлення з позитивною оцінкою якості продукту або негативною критикою програмного забезпечення. Разом із тим було виявлено, що короткі або неоднозначні повідомлення можуть створювати труднощі для нейромережевої моделі. Зокрема, окремі короткі негативні висловлювання іноді помилково визначалися як нейтральні повідомлення через недостатню кількість подібних прикладів у навчальному наборі даних.

Для підвищення якості роботи системи було реалізовано додатковий rule-based механізм аналізу тональності, який використовує словники позитивно та негативно забарвлених слів і фраз. У процесі тестування було встановлено, що комбінування нейромережевого та rule-based підходів дозволяє суттєво покращити результати аналізу коротких повідомлень та підвищує стабільність роботи системи загалом. Особливо ефективним даний підхід виявився під час

аналізу емоційно виражених текстових конструкцій, що містять явні ознаки позитивного або негативного ставлення користувача.

Окрему увагу було приділено тестуванню механізму автоматичного перекладу тексту. Оскільки нейромережеву модель було навчено переважно на англомовному наборі даних, система використовує автоматичний переклад текстових повідомлень англійською мовою перед виконанням аналізу. Під час тестування перевірялася якість визначення тональності україномовних та російськомовних повідомлень після автоматичного перекладу. Результати тестування показали, що інтеграція модуля перекладу дозволяє забезпечити підтримку багатомовного аналізу текстових даних без необхідності окремого навчання нейромережевої моделі для кожної мови.

Для перевірки роботи системи в умовах реальних даних було реалізовано інтеграцію із Steam API та виконано аналіз користувацьких відгуків до комп’ютерних ігор. У процесі тестування система автоматично завантажувала текстові повідомлення користувачів, виконувала їх аналіз та формувала статистику розподілу тональності відгуків. Результати роботи системи відображалися у вигляді гістограм, кругових діаграм та таблиць результатів. Використання реальних текстових повідомлень дозволило оцінити практичну ефективність системи та перевірити стабільність її роботи під час аналізу великих обсягів даних.

Під час тестування графічного інтерфейсу користувача було встановлено, що реалізований вебзастосунок забезпечує стабільну взаємодію з нейромережевою моделлю та дозволяє виконувати аналіз текстових повідомлень у режимі реального часу. Інтерфейс коректно відображає результати аналізу, підтримує побудову графіків та забезпечує можливість експорту результатів у формат CSV. Крім того, використання Streamlit дозволило реалізувати інтуїтивно зрозумілу структуру взаємодії користувача із системою.

За результатами проведеного тестування було встановлено, що розроблена нейромережа забезпечує достатній рівень ефективності для автоматизованого

аналізу тональності текстових відгуків. Реалізована система успішно виконує аналіз реальних користувацьких повідомлень, підтримує багатомовну обробку тексту та забезпечує візуалізацію результатів аналізу. Разом із тим результати тестування показали, що точність класифікації може залежати від якості навчального набору даних, складності текстових конструкцій та особливостей автоматичного перекладу повідомлень.

Таким чином, проведене тестування підтвердило працездатність розробленої системи аналізу тональності текстових відгуків та продемонструвало можливість її практичного використання для автоматизованого аналізу користувацьких повідомлень. Поєднання методів глибокого навчання, rule-based аналізу, автоматичного перекладу тексту та сучасних засобів візуалізації забезпечило створення функціональної програмної системи, придатної для подальшого розвитку та використання у задачах обробки природної мови.

4 Ергономіка інформаційних технологій

4.1 Ергономічні вимоги до робочого місця користувача

Ефективність роботи користувача з інформаційними системами значною мірою залежить від умов організації робочого місця та дотримання ергономічних вимог під час використання комп’ютерної техніки. Неправильна організація робочого простору може призводити до підвищеної втомлюваності, погіршення концентрації уваги, зниження продуктивності праці та виникнення професійних захворювань. Саме тому під час експлуатації програмного забезпечення для аналізу тональності текстових відгуків необхідно враховувати сучасні вимоги ергономіки інформаційних технологій та забезпечувати безпечні умови праці користувача.

Робоче місце користувача програмної системи повинно бути організоване таким чином, щоб забезпечувати комфортне та безпечне виконання роботи протягом тривалого часу. Основним елементом робочого місця є персональний комп’ютер або ноутбук із засобами введення та відображення інформації. При

цьому особливу увагу необхідно приділяти правильному розташуванню монітора, клавіатури, маніпулятора типу «миша» та інших периферійних пристроїв.

Монітор повинен розташовуватися безпосередньо перед користувачем на відстані приблизно 50–70 сантиметрів від очей. Верхня межа екрана має знаходитися на рівні очей або трохи нижче, що дозволяє зменшити навантаження на шийний відділ хребта та запобігти швидкій втомі під час тривалої роботи. Крім того, важливим є правильний вибір параметрів яскравості та контрастності дисплея, оскільки надмірно яскраве або недостатньо освітлене зображення може негативно впливати на зір користувача.

Клавіатура повинна розташовуватися на поверхні столу таким чином, щоб руки користувача перебували у природному положенні без надмірного навантаження на кисті та плечові суглоби. Маніпулятор «миша» має знаходитися поруч із клавіатурою на одному рівні, що забезпечує мінімізацію зайвих рухів під час роботи. Для запобігання перевтомі рекомендується використовувати ергономічні периферійні пристрої, які зменшують навантаження на опорно-руховий апарат.

Важливою складовою організації робочого місця є правильний вибір меблів. Робочий стіл повинен мати достатню площу для розміщення комп’ютерної техніки та допоміжних матеріалів. Висота столу має відповідати антропометричним характеристикам користувача та забезпечувати правильне положення рук під час роботи з клавіатурою. Робоче крісло повинно бути регульованим за висотою та кутом нахилу спинки, що дозволяє підтримувати правильне положення хребта та зменшує навантаження на спину під час тривалої роботи за комп’ютером.

Особливу увагу необхідно приділяти освітленню робочого місця. Освітлення повинно бути достатнім для комфортного сприйняття інформації на екрані монітора та документах. Рекомендується використовувати комбіноване освітлення, яке поєднує природне та штучне джерела світла. При цьому слід уникати потрапляння прямих сонячних променів на поверхню екрана, оскільки це

може створювати відблиски та погіршувати сприйняття інформації. Для роботи з програмним забезпеченням бажано використовувати нейтральне або тепле освітлення, яке знижує навантаження на органи зору.

Одним із важливих факторів ергономіки є забезпечення оптимального мікроклімату у приміщенні. Температура повітря у робочій зоні повинна знаходитися у межах 20–24 °C, а відносна вологість — у межах 40–60 %. Недотримання цих параметрів може призводити до погіршення самопочуття користувача та зниження працездатності. Крім того, приміщення повинно регулярно провітрюватися для забезпечення достатнього рівня вентиляції та підтримання комфортних умов праці.

Під час роботи з інформаційними системами важливим є також дотримання режиму праці та відпочинку. Тривала безперервна робота за комп’ютером може спричиняти перевтому, зниження концентрації уваги та негативний вплив на органи зору. Для зменшення навантаження рекомендується робити короткі перерви кожні 45–60 хвилин роботи. Під час перерв доцільно виконувати вправи для очей та змінювати положення тіла, що сприяє покращенню кровообігу та зменшенню м’язового напруження.

Під час використання системи аналізу тональності текстових відгуків користувач переважно працює із великими обсягами текстової інформації та графічними елементами інтерфейсу. Саме тому особливого значення набуває забезпечення зручності візуального сприйняття інформації. Для цього у програмному забезпеченні повинні використовуватися достатній розмір шрифтів, контрастні кольорові схеми та зрозуміле розташування елементів інтерфейсу. Дотримання цих вимог дозволяє знизити когнітивне навантаження на користувача та підвищити ефективність взаємодії із системою.

Таким чином, дотримання ергономічних вимог до організації робочого місця користувача є важливою умовою ефективної та безпечної роботи з програмною системою аналізу тональності текстових відгуків. Раціональна організація робочого простору, забезпечення комфортних умов праці та

дотримання режиму роботи дозволяють знизити рівень втомлюваності користувача, підвищити продуктивність праці та забезпечити безпечне використання інформаційних технологій. [17]

4.2 Вимоги до інтерфейсу користувача

Однією з важливих складових ефективності програмного забезпечення є якість реалізації інтерфейсу користувача. Саме інтерфейс забезпечує взаємодію між користувачем та програмною системою, впливає на швидкість виконання операцій, зручність використання функціоналу та загальну продуктивність роботи. У процесі розробки системи аналізу тональності текстових відгуків особлива увага приділялася створенню простого, зрозумілого та ергономічного інтерфейсу, який забезпечує комфортну роботу користувача з великими обсягами текстових даних.

Інтерфейс користувача повинен відповідати основним принципам ергономіки програмного забезпечення та забезпечувати мінімізацію когнітивного навантаження під час роботи. Для досягнення цієї мети необхідно дотримуватися принципів логічної структури інтерфейсу, уніфікації елементів керування та зрозумілого представлення інформації. Усі основні елементи системи повинні бути розташовані таким чином, щоб користувач міг швидко отримувати доступ до необхідних функцій без виконання зайвих дій.

Під час розробки інтерфейсу системи аналізу тональності текстових відгуків було використано фреймворк Streamlit, який дозволяє створювати інтерактивні вебзастосунки з інтуїтивно зрозумілим інтерфейсом. Основною перевагою використання Streamlit є можливість швидкого створення програмних засобів для візуалізації результатів аналізу даних без необхідності застосування складних frontend-технологій. Реалізований інтерфейс забезпечує просту взаємодію користувача із системою та підтримує виконання основних функцій аналізу текстових повідомлень.

Однією з основних вимог до інтерфейсу користувача є забезпечення зрозумілості навігації. У межах розробленого програмного забезпечення всі

основні елементи керування розташовані у логічній послідовності. Поле введення Steam AppID, кнопка запуску аналізу, таблиці результатів та графічні елементи візуалізації знаходяться у межах однієї робочої області, що дозволяє користувачу швидко орієнтуватися у функціоналі системи. Такий підхід сприяє зменшенню часу навчання користувача та підвищує загальну ефективність роботи із програмним забезпеченням.

Важливим аспектом ергономіки інтерфейсу є правильний вибір кольорової гами та параметрів відображення інформації. Для забезпечення комфортного сприйняття тексту використовуються контрастні кольори фону та шрифтів, що дозволяє знизити навантаження на органи зору користувача. Крім того, у програмному забезпеченні застосовуються достатні розміри шрифтів та інтервали між елементами інтерфейсу, що покращує читабельність текстової інформації та полегшує сприйняття результатів аналізу.

Особливу увагу під час проєктування інтерфейсу було приділено засобам візуалізації результатів роботи системи. Для відображення статистичної інформації використовуються гістограми та кругові діаграми, побудовані за допомогою бібліотеки Matplotlib. Графічне представлення результатів дозволяє користувачу швидко оцінювати співвідношення позитивних, негативних та нейтральних відгуків без необхідності детального аналізу великих обсягів текстових даних. Використання візуальних елементів суттєво покращує сприйняття інформації та підвищує ефективність роботи із системою.

Ще однією важливою вимогою до інтерфейсу користувача є забезпечення швидкого зворотного зв’язку між системою та користувачем. У розробленому програмному забезпеченні реалізовано механізми відображення поточного стану виконання операцій, зокрема індикатори завантаження та повідомлення про завершення аналізу. Це дозволяє користувачу контролювати процес роботи системи та своєчасно отримувати інформацію про результати виконання операцій.

Для забезпечення зручності роботи з великими обсягами даних у системі реалізовано табличне представлення результатів аналізу. Користувач має

можливість переглядати текстові повідомлення разом із визначеною для них тональністю, що дозволяє швидко перевіряти результати роботи нейромережевої моделі. Крім того, реалізовано можливість експорту результатів аналізу у формат CSV для подальшого використання у сторонніх програмних засобах.

Важливою характеристикою сучасного інтерфейсу є адаптивність до різних умов використання. Streamlit автоматично забезпечує коректне відображення елементів інтерфейсу на різних типах пристроїв та при різних роздільних здатностях екрана. Це дозволяє використовувати систему не лише на персональних комп’ютерах, але й на ноутбуках або інших пристроях із веббраузером.

Під час розробки інтерфейсу також враховувалися вимоги щодо мінімізації помилок користувача. Для цього у системі використовуються зрозумілі текстові підказки, логічно організовані елементи керування та автоматична перевірка введених даних. Наприклад, перед виконанням аналізу система перевіряє коректність введеного Steam AppID та повідомляє користувача про можливі помилки у випадку некоректного введення інформації.

Таким чином, реалізований інтерфейс користувача відповідає сучасним вимогам ергономіки інформаційних технологій та забезпечує ефективну взаємодію користувача із системою аналізу тональності текстових відгуків. Використання інтуїтивно зрозумілої структури, графічної візуалізації результатів та сучасних засобів веброзробки дозволило створити зручний програмний інструмент для роботи з текстовими даними та аналізу емоційного забарвлення повідомлень. [23]

4.3 Заходи з охорони праці та безпеки при роботі з програмним забезпеченням

Під час використання сучасних інформаційних технологій важливого значення набуває забезпечення безпечних умов праці користувача та дотримання вимог охорони праці під час роботи з комп’ютерною технікою. Робота із системою аналізу тональності текстових відгуків передбачає тривале

використання персонального комп’ютера, взаємодію із вебзастосунками та обробку великих обсягів текстових даних, що потребує дотримання відповідних правил безпеки та організації робочого процесу.

Одним із основних факторів безпеки є забезпечення електробезпеки під час експлуатації комп’ютерної техніки. Усі технічні засоби, що використовуються для роботи із програмним забезпеченням, повинні відповідати вимогам електробезпеки та мати справну ізоляцію електричних кабелів. Підключення обладнання необхідно здійснювати лише до справних електромереж із наявністю заземлення. Використання несправних електроприладів або пошкоджених кабелів може призводити до виникнення коротких замикань, ураження електричним струмом та пошкодження обладнання.

Під час роботи з комп’ютерною технікою необхідно дотримуватися правил пожежної безпеки. Робоче приміщення повинно бути обладнане засобами пожежогасіння та системою вентиляції. Забороняється розміщувати легкозаймисті матеріали поблизу електронного обладнання, а також перевантажувати електромережу шляхом одночасного підключення великої кількості пристроїв. У разі виникнення несправностей обладнання користувач повинен негайно припинити роботу та відключити пристрій від електромережі.

Особливу увагу необхідно приділяти захисту здоров’я користувача під час тривалої роботи за комп’ютером. Постійне використання монітора може негативно впливати на органи зору, викликати перевтому та зниження концентрації уваги. Для мінімізації негативного впливу рекомендується використовувати монітори із сучасними технологіями захисту зору, правильно налаштовувати яскравість та контрастність екрана, а також дотримуватися режиму праці та відпочинку. Під час тривалої роботи користувач повинен регулярно робити перерви та виконувати вправи для очей і опорно-рухового апарату.

Важливою складовою безпечної роботи з програмним забезпеченням є захист інформації та забезпечення кібербезпеки. Оскільки система аналізу

тональності використовує підключення до мережі Інтернет та взаємодіє із зовнішніми сервісами, зокрема Steam API та сервісами автоматичного перекладу тексту, необхідно забезпечити захист переданих даних та стабільність роботи програмного забезпечення. Для цього рекомендується використовувати сучасні антивірусні засоби, регулярно оновлювати операційну систему та програмне забезпечення, а також застосовувати механізми резервного копіювання даних.

Під час роботи із вебзастосунками необхідно враховувати ризики, пов’язані із використанням зовнішніх мережевих ресурсів. Зокрема, важливо контролювати джерела отримання даних та уникати використання ненадійних вебресурсів. Використання офіційних API та перевірених програмних бібліотек дозволяє підвищити рівень безпеки системи та зменшити ймовірність виникнення програмних помилок або витоку інформації.

У процесі експлуатації програмного забезпечення важливим є також забезпечення стабільності роботи комп’ютерної системи. Для цього рекомендується регулярно контролювати використання оперативної пам’яті, процесорних ресурсів та обсягів дискового простору. Надмірне навантаження на систему може призводити до зниження продуктивності роботи програмного забезпечення та виникнення помилок під час аналізу великих обсягів текстових даних.

Під час роботи із системою аналізу тональності необхідно дотримуватися правил безпечного використання програмного забезпечення. Перед запуском програми користувач повинен переконатися у справності обладнання та коректності роботи мережевого підключення. У випадку виникнення помилок або нестабільної роботи системи необхідно завершити виконання програми та перевірити правильність налаштувань програмного середовища. Крім того, важливим є використання ліцензійного програмного забезпечення та офіційних бібліотек, що дозволяє зменшити ризик виникнення вразливостей у системі.

Значну роль у забезпеченні безпеки праці відіграє організація режиму роботи користувача. Тривале виконання монотонних операцій може призводити

до психологічного навантаження та зниження ефективності роботи. Для підтримання працездатності рекомендується чергувати роботу за комп’ютером із короткими перервами, змінювати вид діяльності та підтримувати комфортний мікроклімат у приміщенні.

Під час розробки системи аналізу тональності текстових відгуків також враховувалися вимоги до безпечного інтерфейсу користувача. Зокрема, реалізовано зрозумілу структуру навігації, контрастне відображення інформації та механізми попередження користувача про можливі помилки. Використання інтуїтивно зрозумілого інтерфейсу дозволяє зменшити кількість помилкових дій користувача та підвищує загальний рівень безпеки роботи із системою.

Таким чином, дотримання вимог охорони праці та безпеки під час роботи із програмним забезпеченням є важливою умовою ефективної та стабільної експлуатації системи аналізу тональності текстових відгуків. Виконання правил електробезпеки, забезпечення захисту інформації, дотримання ергономічних вимог та організація безпечного режиму праці дозволяють мінімізувати негативний вплив інформаційних технологій на користувача та забезпечують надійну роботу програмного забезпечення.

Висновки

У результаті виконання бакалаврської роботи було розроблено нейромережеву систему аналізу тональності текстових відгуків, орієнтовану на автоматичне визначення емоційного забарвлення текстових повідомлень користувачів. У ході роботи було проведено аналіз сучасних підходів до обробки природної мови та методів аналізу тональності тексту, розглянуто особливості класичних алгоритмів машинного навчання, рекурентних нейронних мереж, моделей LSTM, BiLSTM та сучасних трансформерних архітектур. На підставі проведеного аналізу обґрунтовано доцільність використання двонапрямленої рекурентної нейронної мережі BiLSTM як основи програмної системи.

У роботі досліджено основні етапи попередньої обробки текстових даних, зокрема очищення тексту, нормалізацію, токенізацію, видалення стоп-слів та

формування числових представлень тексту. Визначено роль preprocessing-процедур у підвищенні точності класифікації та забезпеченні ефективної роботи нейромережевої моделі. Також було розглянуто сучасні методи векторизації текстових даних та особливості використання embedding-представлень у задачах аналізу природної мови.

Практична частина роботи присвячена розробці програмного забезпечення для автоматизованого аналізу текстових відгуків. У процесі реалізації було використано мову програмування Python, бібліотеки PyTorch, TensorFlow, Pandas та Matplotlib, а також фреймворк Streamlit для створення інтерактивного користувацького інтерфейсу. Реалізовано механізм інтеграції із зовнішнім сервісом Steam API, що дозволило автоматизувати отримання реальних користувацьких відгуків із платформи Steam. Додатково впроваджено механізм автоматичного перекладу тексту для підтримки багатомовного аналізу.

Розроблена система забезпечує автоматичне визначення позитивної, нейтральної та негативної тональності текстових повідомлень, формування статистики результатів аналізу, побудову графічних візуалізацій та збереження результатів у форматі CSV. У процесі тестування система продемонструвала стабільну роботу та достатню точність класифікації текстових відгуків, що підтверджує ефективність використання нейромережевих підходів у задачах аналізу природної мови.

Отримані результати свідчать про перспективність використання нейромережевих моделей для автоматизації аналізу текстових даних у сучасних інформаційних системах. Розроблене програмне забезпечення може бути використане для аналізу користувацьких відгуків, моніторингу громадської думки, оцінювання якості цифрових продуктів та інших задач, пов’язаних із обробкою текстової інформації. Подальший розвиток системи може бути пов’язаний із використанням трансформерних моделей, розширенням підтримки мов та підвищенням точності класифікації за рахунок використання більших навчальних наборів даних.

Список використаних джерел та літератури

1. Jurafsky D., Martin J. H. *Speech and Language Processing*. — 3rd ed. — Stanford University, 2023.

2. Goodfellow I., Bengio Y., Courville A. *Deep Learning*. — MIT Press, 2016.

3. Goldberg Y. *Neural Network Methods for Natural Language Processing*. — Morgan & Claypool Publishers, 2017.

4. Aggarwal C. C. *Neural Networks and Deep Learning*. — Springer, 2018.

5. Manning C. D., Schütze H. *Foundations of Statistical Natural Language Processing*. — MIT Press, 1999.

6. Vaswani A. et al. Attention Is All You Need // *Advances in Neural Information Processing Systems*. — 2017.

7. Hochreiter S., Schmidhuber J. Long Short-Term Memory // *Neural Computation*. — 1997. — Vol. 9, No. 8.

8. Graves A., Schmidhuber J. Framewise phoneme classification with bidirectional LSTM networks // *Proceedings of the International Joint Conference on Neural Networks*. — 2005.

9. Devlin J. et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // *NAACL-HLT*. — 2019.

10. Young T. et al. Recent Trends in Deep Learning Based Natural Language Processing // *IEEE Computational Intelligence Magazine*. — 2018.

11. Liu B. *Sentiment Analysis and Opinion Mining*. — Morgan & Claypool Publishers, 2012.

12. Cambria E., White B. Jumping NLP Curves: A Review of Natural Language Processing Research // *IEEE Computational Intelligence Magazine*. — 2014.

13. Mikolov T. et al. Efficient Estimation of Word Representations in Vector Space // *arXiv preprint arXiv:1301.3781*. — 2013.

14. Pennington J., Socher R., Manning C. GloVe: Global Vectors for Word Representation // *EMNLP*. — 2014.

15. Brownlee J. *Deep Learning for Natural Language Processing*. — Machine Learning Mastery, 2017.

16. Bird S., Klein E., Loper E. *Natural Language Processing with Python*. — O'Reilly Media, 2009.

17. Chollet F. *Deep Learning with Python*. — Manning Publications, 2021.

18. PyTorch Official Documentation

19. TensorFlow Official Documentation

20. Keras Documentation

21. Pandas Documentation

22. Matplotlib Documentation

23. Streamlit Documentation

24. Steam Web API Documentation

25. Google Translate API Documentation

26. Han J., Kamber M., Pei J. *Data Mining: Concepts and Techniques*. — Morgan Kaufmann, 2011.

27. Géron A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. — O'Reilly Media, 2022.

28. Raschka S., Mirjalili V. *Python Machine Learning*. — Packt Publishing, 2019.

29. Scikit-learn Documentation

30. LeCun Y., Bengio Y., Hinton G. Deep Learning // *Nature*. — 2015. — Vol. 521.

Додатки:

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БУДІВНИЦТВА І АРХІТЕКТУРИ

Розробка нейромережевої системи аналізу тональності текстових відгуків

Виконав: Граф Є. С.

ІСТ-22

Науковий керівник: д.т.н., професор Бородавка Є.В.

Київ-2026

Актуальність теми

- У сучасному цифровому середовищі кількість текстових даних постійно зростає.
  Користувачі активно залишають відгуки, коментарі та оцінки на вебсайтах, у соціальних мережах, інтернет-магазинах та онлайн -сервісах.
- Аналіз великих обсягів текстової інформації вручну є складним, тривалим та малоефективним процесом. Саме тому виникає необхідність автоматизації аналізу текстових даних із використанням сучасних методів штучного інтелекту та обробки природної мови (NLP).
- Одним із найбільш актуальних напрямів NLP є аналіз тональності тексту, який дозволяє автоматично визначати емоційне забарвлення відгуків та класифікувати їх як позитивні, негативні або нейтральні.
- Використання нейромережевих моделей, зокрема BiLSTM, дозволяє підвищити точність аналізу та враховувати контекст текстових повідомлень.

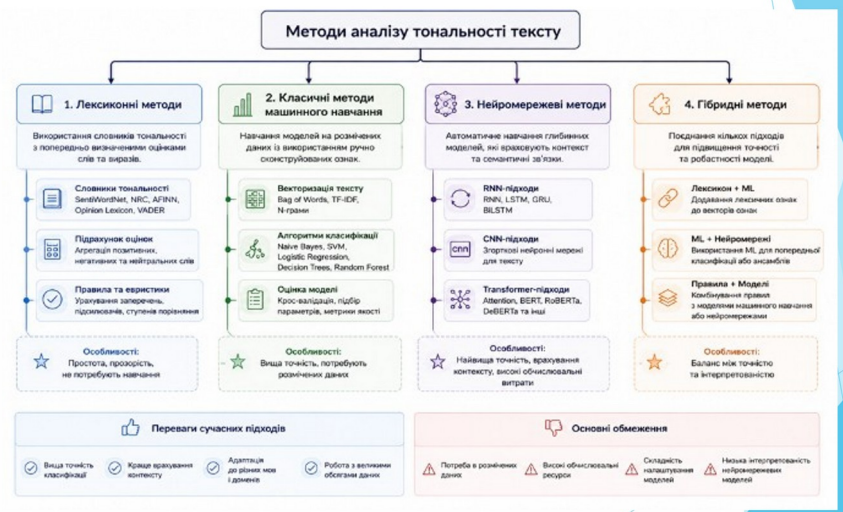
Мета та завдання роботи

- Метою роботи є розробка нейромережевої системи аналізу тональності текстових відгуків із використанням методів обробки природної мови та архітектури BiLSTM.
- **Завдання роботи:**
- провести аналіз сучасних методів аналізу тональності тексту;
- дослідити особливості технологій NLP;
- виконати порівняння нейромережевих архітектур;
- обґрунтувати вибір моделі BiLSTM;
- реалізувати програмну систему аналізу тональності;
- провести навчання та тестування моделі;
- оцінити ефективність роботи системи.

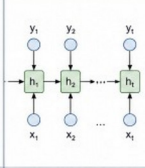
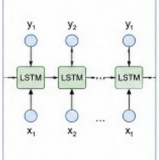
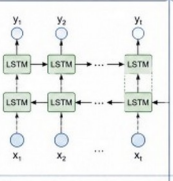
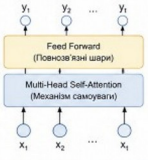
Аналіз предметної області

- Аналіз тональності тексту є одним із напрямів обробки природної мови (NLP), основною задачею якого є автоматичне визначення емоційного забарвлення текстових повідомлень.
- Система аналізу тональності дозволяє класифікувати текстові на позитивні, негативні і нейтральні.
- Основними джерелами текстових даних є соціальні мережі, інтернет-магазини, форуми, онлайн-сервіси, платформи з відгуками користувачів.
- Особливістю предметної області є складність обробки природної мови через наявність контексту, сленгу, скорочень, емоційних конструкцій, неоднозначності тексту.

Методи аналізу тональності



Архітектури нейромереж для аналізу

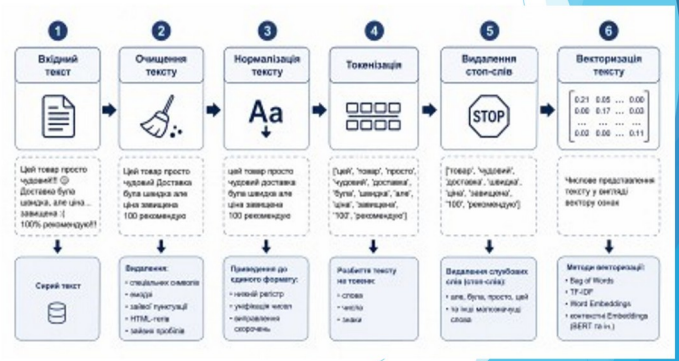
Архітектура	RNN	LSTM	BiLSTM	Transformer
Схема архітектури				
Використання контексту	Лише попередній контекст (зліва направо)	Попередній контекст (зліва направо)	Попередній та наступний контекст (зліва направо і справа наліво)	Весь контекст одночасно (глобальні залежності)
Робота з довгими залежностями	Низька	Середня – висока	Висока	Дуже висока
Переваги	• Проста архітектура • Невеликі вимоги до ресурсів	• Враховує довгострокові залежності • Менше проблем із згасанням градієнта	• Враховує двосторонній контекст • Краща якість класифікації	• Паралельна обробка • Найкраще розуміння контексту • Висока точність
Недоліки	• Проблема згасання градієнта • Складно працює з довгими послідовностями	• Більше параметрів • Більша обчислювальна складність	• Ще більше параметрів • Потребує більше пам'яті та часу навчання	• Дуже великі вимоги до ресурсів • Потребує великих обсягів даних
Обчислювальна складність	Низька	Середня	Висока	Дуже висока
Точність у задачах аналізу тональності	Низька – середня	Середня – висока	Висока	Найвища

Попередня обробка тексту

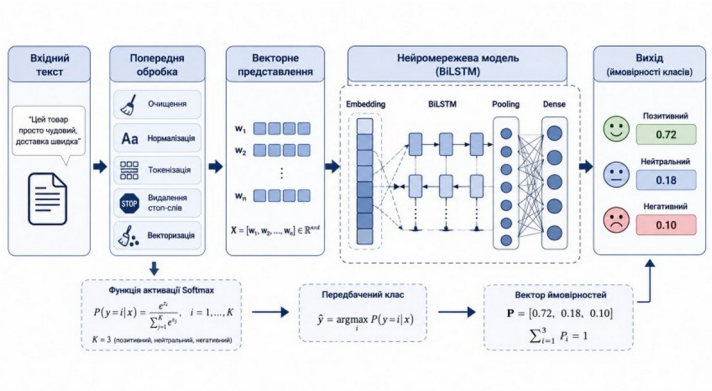
Перед подачею текстових даних до нейромережі виконується етап попередньої обробки, який дозволяє підвищити якість навчання моделі та зменшити кількість шуму в даних.

- **Основні етапи обробки:**

1. Очищення тексту від спеціальних символів;
2. Перетворення тексту до нижнього регістру;
3. Токенізація тексту;
4. Видалення stop-слів;
5. Векторизація текстових даних.



Структура нейромережевої системи



Технології реалізації системи

Технологія	Призначення
Python	Основна мова програмування
PyTorch	Реалізація нейромережі
Pandas	Обробка даних
NumPy	Робота з масивами даних
Matplotlib	Візуалізація результатів
Streamlit	Реалізація вебінтерфейсу

Інтерфейс користувача системи

Для взаємодії користувача із системою було реалізовано вебінтерфейс, який дозволяє вводити текстові відгуки та отримувати результат аналізу тональності в режимі реального часу.

- **Основні можливості інтерфейсу:**
  - введення текстового повідомлення;
  - автоматичний аналіз тональності;
  - визначення класу тексту;
  - відображення результату класифікації;
  - простий та зручний інтерфейс.

Навчання та тестування моделі

Для навчання моделі аналізу тональності було використано набір текстових відгуків із попередньо визначеними класами тональності.

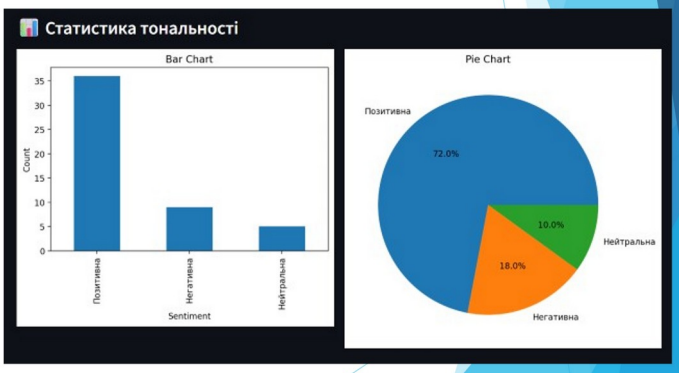
Під час навчання використовувалися:
•train/test розподіл даних;
•embedding-представлення тексту;
•функція втрат CrossEntropyLoss;
•оптимізатор Adam;
•декілька епох навчання моделі.

```
Epoch 1/15 | Loss: 2381.2663 | Accuracy: 0.7488
Epoch 2/15 | Loss: 1716.0339 | Accuracy: 0.7897
Epoch 3/15 | Loss: 1465.8687 | Accuracy: 0.8050
Epoch 4/15 | Loss: 1309.2226 | Accuracy: 0.8140
Epoch 5/15 | Loss: 1186.4007 | Accuracy: 0.8185
Epoch 6/15 | Loss: 1084.2055 | Accuracy: 0.8198
Epoch 7/15 | Loss: 994.4402 | Accuracy: 0.8212
Epoch 8/15 | Loss: 907.6033 | Accuracy: 0.8235
Epoch 9/15 | Loss: 830.0783 | Accuracy: 0.8242
Epoch 10/15 | Loss: 758.4257 | Accuracy: 0.8165
Epoch 11/15 | Loss: 684.1401 | Accuracy: 0.8268
Epoch 12/15 | Loss: 618.8808 | Accuracy: 0.8261
Epoch 13/15 | Loss: 556.7404 | Accuracy: 0.8280
Epoch 14/15 | Loss: 497.9587 | Accuracy: 0.8282
Epoch 15/15 | Loss: 454.0823 | Accuracy: 0.8309
Model training completed.
```

Результати роботи системи

- Після навчання нейромережевої моделі було проведено тестування системи на текстових відгуках різної тональності.

Результати показали:
•коректне визначення позитивних відгуків;
•успішну класифікацію негативних повідомлень;
•можливість роботи з нейтральними текстами;
•стабільну роботу моделі на тестових даних.



Переваги розробленої системи

- Розроблена нейромережева система аналізу тональності має ряд переваг, які забезпечують ефективність її використання для обробки текстових даних.
- **Основні переваги:**
  - автоматизація аналізу текстових відгуків;
  - висока швидкість обробки даних;
  - врахування контексту повідомлень;
  - підтримка різних типів текстів;
  - можливість масштабування системи;
  - зручний вебінтерфейс користувача;
  - можливість подальшого вдосконалення моделі.

Висновки

- У результаті виконання дипломної роботи було досліджено сучасні методи аналізу тональності тексту та особливості застосування технологій NLP для обробки текстових даних.
- У ході роботи:
  - проведено аналіз існуючих підходів до Sentiment Analysis;
  - досліджено архітектури нейронних мереж для задач NLP;
  - обґрунтовано вибір моделі **BiLSTM**;
  - реалізовано програмну систему аналізу тональності;
  - створено вебінтерфейс користувача;
  - проведено навчання та тестування моделі.
- Результати роботи підтвердили ефективність використання нейромережевих методів для автоматичного визначення тональності текстових відгуків.

Дякую за увагу