МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Харківський національний університет радіоелектроніки

Факультет комп'ютерних наук
(повна назва)

Кафедра програмної інженерії
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження моделей нейронних мереж типу LSTM
для семантичного та емоційного аналізу природньої мови людини
(тема)

Виконав:

здобувач 2 року навчання

групи ІПЗм-23-2

Юрій КАШНІКОВ
(власне ім'я, ПРІЗВИЩЕ)

Спеціальність 121 – Інженерія програмного забезпечення
(код і повна назва спеціальності)

Тип програми освітньо-наукова

Керівник доц. Ірина АФАНАСЬЄВА
(посада, власне ім'я, ПРІЗВИЩЕ)

Допускається до захисту
Зав. кафедри _____ Кирило СМЕЛЯКОВ
(підпис) (власне ім'я, ПРІЗВИЩЕ)

2025 р.

Харківський національний університет радіоелектроніки

Факультет комп'ютерних наук
Кафедра програмної інженерії
Рівень вищої освіти другий (магістерський)
Спеціальність 121 – Інженерія програмного забезпечення
Тип програми освітньо-наукова
Освітня програма Інженерія програмного забезпечення
(шифр і назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
«___» _____ 2025 р.

ЗАВДАННЯ
на кваліфікаційну роботу

студентові Кашнікову Юрію Костянтиновичу
(прізвище, ім'я, по батькові)

1. Тема роботи «Дослідження моделей нейронних мереж типу LSTM для семантичного та емоційного аналізу природньої мови людини».
   Затверджена наказом по університету від 15 квітня 2025р. № 290 СТ
2. Термін подання студентом роботи до екзаменаційної комісії 18 червня 2025
3. Вихідні дані до роботи текстові корпуси, моделі нейронних мереж, метрики оцінювання.
4. Перелік питань, які потрібно опрацювати в роботі

вступ, аналіз предметної галузі, огляд й аналіз літературних, наукових джерел, постановка задачі, теоретичне дослідження, проведення експерименту, висновки.

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Отримання завдання	15.10.2024	виконано
2	Аналіз предметної галузі і постановка задачі	25.10.2024	виконано
3	Огляд й аналіз літературних, наукових джерел	15.11.2024	виконано
4	Теоретичне дослідження	06.01.2025	виконано
5	Підготовка до апробації результатів дослідження. Публікація матеріалів	05.03.2025	виконано
6	Проведення експерименту	26.03.2025	виконано
7	Підготовка пояснювальної записки	13.04.2025	виконано
8	Підготовка презентації та доповіді	01.05.2025	виконано
9	Перевірка на плагіат	10.06.2025	виконано
10	Нормоконтроль	14.06.2025	виконано
11	Рецензування	14.06.2025	виконано
12	Попередній захист	14.06.2025	виконано
13	Занесення диплома в електронний архів	17.06.2025	виконано
14	Допуск до захисту у зав. кафедри	17.06.2025	виконано

Дата видачі завдання 15 жовтня 2024р.

Студент _____ Юрій КАШНІКОВ
(підпис)

Керівник роботи _____ доц. Ірина АФАНАСЬЄВА
(підпис) (посада, власне ім'я, ПРІЗВИЩЕ)

РЕФЕРАТ / ABSTRACT

Пояснювальна записка містить: 64 с., 2 рис., 1 табл., 14 джерел, 4 додатки.

АНАЛІЗ, ДАТАСЕТ, ДВОСПРЯМОВАНА МОДЕЛЬ, ЕМОЦІЙНИЙ АНАЛІЗ, КЛАСИФІКАЦІЯ, НЕЙРОННА МЕРЕЖА, ОДНОСПРЯМОВАНА МОДЕЛЬ, СЕМАНТИЧНИЙ АНАЛІЗ, ТОКЕНІЗАЦІЯ, ADAM, LSTM, TENSORFLOW.

Об’єктом дослідження є процеси семантичного та емоційного аналізу природної мови, які забезпечують машинне розуміння текстів і їхнього контексту.

Метою роботи є розробка та аналіз ефективності моделей нейронних мереж типу LSTM для семантичного та емоційного аналізу природної мови.

Методами дослідження є аналіз існуючих підходів до семантичного й емоційного аналізу тексту, аналіз існуючих моделей нейронних мереж типу LSTM, створення гібридної моделі з використанням механізму уваги і шарами нормалізації, а також оцінка ефективності моделей на основі метрик точності, повноти та F1-міри.

У результаті роботи було розроблено гібридну модель, протестовано її на наборі даних News Category Dataset та отримано підвищення ефективності семантичного аналізу тексту.

ANALYSIS, DATASET, BIDIRECTIONAL MODEL, EMOTION ANALYSIS, CLASSIFICATION, NEURAL NETWORK, UNIDIRECTIONAL MODEL, SEMANTIC ANALYSIS, TOKENIZATION, ADAM, LSTM, TENSORFLOW.

The object of research is the processes of semantic and emotional analysis of natural language, which provide machine understanding of texts and their context.

The purpose of the work is to develop and analyze the effectiveness of neural network models of the LSTM type for semantic and emotional analysis of natural language.

The research methods include the analysis of existing approaches to semantic and emotional text analysis, the analysis of existing models of neural networks such as LSTM, the creation of a hybrid model using the attention mechanism and normalization layers, as well as the evaluation of the effectiveness of models based on accuracy, completeness and F1-measure metrics.

As a result of the work, a hybrid model was developed, tested on the News Category Dataset, and improved the efficiency of semantic text analysis.

Заява щодо самостійного виконання роботи та можливості її публікації в електронному архіві відкритого доступу ElArKhNURE.

Завідувачу кафедри ПІ
(скорочена назва кафедри)
проф. Кирилу СМЕЛЯКОВУ
(вчене звання, власне ім'я, прізвище)

ЗАЯВА

щодо самостійності виконання кваліфікаційної роботи та можливості її публікації в електронному архіві відкритого доступу ElAr KhNURE

Я,

Кашніков Юрій Костянтинович
(прізвище, ім'я, по батькові)

здобувач вищої освіти на другому (магістерському) рівні вищої освіти академічної групи ІПЗм-23-2

кафедра програмної інженерії,
(повна назва кафедри)

заявляю: моя кваліфікаційна робота на тему

Дослідження моделей нейронних мереж типу LSTM для семантичного та емоційного аналізу природньої мови людини

що буде представлена в екзаменаційну комісію для публічного захисту, виконана самостійно, в ній не містяться елементи плагіату і вона може бути опублікована в репозиторії "ElArKhNURE". Погоджуюся з авторським договором, відповідно до Положення про репозиторій ХНУРЕ "ElArKhNURE". Всі запозичення з друкованих та електронних джерел мають відповідні посилання.

Я ознайомлений (а) з вимогами академічної доброчесності, згідно з якими виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування дисциплінарних заходів.

Дата 09.06.2025 Підпис _____

ЗМІСТ

ВСТУП

У сучасному світі спостерігається стрімке зростання обсягів текстової інформації, яка створюється та споживається людьми в різних сферах життя. Це створює потребу у розробці ефективних інструментів для аналізу природної мови, які здатні обробляти як семантичні, так і емоційні аспекти тексту. Моделі нейронних мереж типу LSTM (Long Short-Term Memory) довели свою ефективність у вирішенні завдань, пов'язаних з аналізом послідовностей даних, зокрема текстів, завдяки здатності враховувати довготривалі залежності. Вивчення та впровадження таких моделей є важливим кроком у розвитку штучного інтелекту, машинного навчання та обробки природної мови.

Актуальність роботи полягає у потребі вдосконалення методів автоматичного аналізу тексту для подолання проблем, пов'язаних із розумінням контексту, виявленням емоційної забарвленості та забезпеченням адаптивності алгоритмів до різноманітних мовних даних.

Метою роботи є розробка та аналіз ефективності моделей нейронних мереж типу LSTM для семантичного та емоційного аналізу природної мови. Для досягнення цієї мети необхідно вирішити такі задачі:

- дослідити існуючі підходи до семантичного та емоційного аналізу текстів;
- розробити архітектуру моделей на основі LSTM, здатних обробляти текстову інформацію;
- провести навчання моделей на вибраних корпусах даних;
- оцінити ефективність моделей за допомогою відповідних метрик;
- визначити перспективи використання розроблених моделей у прикладних задачах.

Об’єктом дослідження є процеси семантичного та емоційного аналізу природної мови, які забезпечують машинне розуміння текстів і їхнього контексту.

Предметом дослідження є моделі нейронних мереж типу LSTM, які застосовуються для аналізу семантичних та емоційних характеристик текстів.

Результати виконаної роботи можуть бути застосовані у різноманітних сферах, забезпечуючи інноваційні рішення для сучасних викликів, пов’язаних із взаємодією людини та інформаційних систем. Однією з ключових областей є створення інтелектуальних чат-ботів та систем автоматизованої підтримки клієнтів, які можуть не лише розуміти зміст текстових запитів, але й враховувати емоційний стан користувача. Це дозволить покращити якість обслуговування, підвищити лояльність клієнтів та знизити навантаження на людський персонал.

Дослідження також можуть бути використані в аналізі соціальних медіа, зокрема для відстеження громадської думки, виявлення трендів, оцінки настроїв і емоцій користувачів у масштабі великих даних. Такий підхід може сприяти ефективному управлінню репутацією брендів, виявленню кризових ситуацій та розробці маркетингових стратегій.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

Предметна область, пов'язана з семантичним та емоційним аналізом природної мови, є однією з найбільш перспективних областей досліджень у сфері штучного інтелекту та машинного навчання. Ця галузь спрямована на створення систем, здатних інтерпретувати, аналізувати та відповідати на текстову інформацію з урахуванням її значення, контексту та емоційного забарвлення. Розробка таких систем є важливою складовою сучасних технологій обробки природної мови, які знаходять застосування у широкому спектрі задач — від автоматизації бізнес-процесів до підвищення якості взаємодії між людиною і комп’ютером.

1.1 Тенденції та перспективи

Сучасний розвиток технологій обробки природної мови характеризується активним впровадженням методів глибокого навчання, зокрема рекурентних нейронних мереж, таких як LSTM[1]. Головними тенденціями є покращення якості аналізу текстів через багатомовні моделі, такі як трансформери, інтеграція контексту в реальному часі та оптимізація обчислювальних ресурсів. Перспективи розвитку галузі включають створення більш інтелектуальних систем, здатних до самонавчання та адаптації, що значно розширить сферу їх застосування. Особливу увагу приділяють морально-етичним аспектам використання цих технологій, таким як забезпечення приватності даних і мінімізація упередженості в алгоритмах. Очікується, що у найближчі роки технології LSTM та їхні вдосконалені версії стануть ще більш важливими у вирішенні складних завдань обробки природної мови.

Іншою важливою тенденцією є інтеграція штучного інтелекту в різноманітні галузі, такі як медицина, освіта, фінанси та маркетинг. Наприклад, моделі LSTM активно використовуються для аналізу емоцій у відгуках пацієнтів, розробки персоналізованих освітніх програм, прогнозування поведінки клієнтів у банківському секторі та оптимізації маркетингових кампаній. Важливим напрямом залишається вдосконалення систем синхронного перекладу, що базуються на

нейронних мережах, з метою забезпечення більш природної взаємодії між носіями різних мов. Також, нейронні мають здатність адаптуватися та навчатися на нових шаблонах, що робить їх безцінними у боротьбі з постійно розвиваючимися методами шахрайства[2].

Крім того, перспективним є розвиток технологій, які враховують культурні та соціальні особливості текстів, забезпечуючи їх більш точне розуміння. Зокрема, у контексті багатомовних моделей вивчається проблема врахування специфічної граматики та ідіоматичних виразів, що дозволить створювати універсальні системи аналізу природної мови. У сукупності ці тенденції свідчать про те, що дослідження в галузі нейронних мереж типу LSTM мають величезний потенціал для покращення якості роботи інформаційних систем і розширення їх можливостей.

1.2 Огляд існуючих підходів

У сучасній обробці природної мови існує широкий спектр підходів, що використовуються для аналізу текстових даних. Традиційні методи, такі як підрахунок частоти слів та статистичний аналіз, мали обмежені можливості у розумінні контексту тексту. Перехід до машинного навчання дозволив створити більш складні моделі, наприклад методи на основі Bag-of-Words (BoW)[3] та TF-IDF, які забезпечували базовий рівень семантичного аналізу.

Згодом із розвитком глибокого навчання стали популярними нейронні мережі, зокрема рекурентні (RNN) та їхні модифікації, такі як LSTM і GRU. Вони продемонстрували здатність ефективно враховувати довготривалі залежності у текстових даних, що стало проривом у розумінні послідовностей. Одночасно, моделі на основі конволюційних нейронних мереж (CNN) виявилися корисними для виявлення локальних патернів у тексті.

Останніми роками велику увагу привернули трансформери, зокрема моделі типу BERT і GPT, які використовують механізми самоприділення уваги для аналізу тексту. Вони показали надзвичайно високі результати у завданнях класифікації, генерації тексту та машинного перекладу. Порівняно з LSTM, трансформери є менш залежними від послідовності, але вимагають значних обчислювальних ресурсів.

У цілому, кожен із підходів має свої переваги та недоліки. Традиційні методи є швидкими та простими, але обмежені в аналізі контексту. Нейронні мережі типу LSTM забезпечують високу точність для обробки послідовностей, тоді як трансформери є найбільш універсальними, проте обчислювально затратними. Вибір методу залежить від специфіки задачі, доступних ресурсів та вимог до точності й швидкодії.

1.3 Обмеження існуючих рішень

Незважаючи на значні досягнення в галузі обробки природної мови, існуючі рішення мають ряд обмежень, які ускладнюють їх повсюдне застосування. Однією з головних проблем є висока обчислювальна складність сучасних моделей, особливо трансформерів. Це ускладнює їх використання на пристроях із обмеженими ресурсами, таких як мобільні телефони або вбудовані системи.

Іншою значною проблемою є залежність від великих обсягів даних для навчання. Більшість сучасних моделей вимагають величезних обсягів текстової інформації для досягнення високої точності. Це створює труднощі у роботі з рідкісними мовами або специфічними галузевими текстами, де обсяги даних є обмеженими.

Ще одним викликом є проблема узагальнення. Багато моделей демонструють високу ефективність на тестових наборах даних, але можуть виявлятися менш точними у реальних сценаріях або при роботі з текстами, що відрізняються від тих, на яких вони навчалися.

Крім того, важливим є питання інтерпретованості моделей. Більшість сучасних нейронних мереж функціонують як "чорні ящики", що ускладнює аналіз їхньої роботи та прийняття рішень на основі їхніх результатів. Це особливо критично у сферах, де потрібна прозорість, таких як медицина або фінанси.

Нарешті, етичні аспекти також залишаються актуальними. Багато моделей можуть демонструвати упередження, якщо вони навчаються на даних, що містять стереотипи або нерівності. Це може призводити до соціальних проблем і

дискримінації, якщо такі моделі застосовуються у чутливих сферах, наприклад, у системах прийняття рішень.

Таким чином, подолання цих обмежень є важливим завданням для дослідників і розробників, що працюють у галузі обробки природної мови.

1.4 Масштаб проблеми

Масштаб проблеми аналізу природної мови є надзвичайно значним, оскільки текстова інформація є основним джерелом комунікації та знань у сучасному суспільстві. Зростання обсягів даних, що створюються щодня, спричиняє необхідність розробки ефективних алгоритмів для їх обробки. Щохвилини у світі надходять мільйони текстових повідомлень, постів у соціальних мережах, статей, електронних листів та інших форм текстової інформації, що робить ручний аналіз цих даних неможливим.

Крім того, велика частина цієї інформації є неструктурованою, що ускладнює її інтерпретацію та використання традиційними алгоритмами. Відсутність універсальних підходів до аналізу текстів різними мовами, з урахуванням їхньої семантики, синтаксису та культурних особливостей, створює додаткові виклики. Наприклад, тексти можуть містити ідіоми, емоційно забарвлені вислови або двозначності, які складно обробляти автоматично[4].

Окрім цього, важливо враховувати різноманітність сфер застосування таких технологій, від автоматизації бізнес-процесів до медичної діагностики та аналізу соціальних трендів. У кожній із цих сфер є свої специфічні вимоги до точності, швидкодії та адаптивності моделей, що ускладнює створення універсальних рішень.

Таким чином, масштаб проблеми аналізу природної мови підкреслює необхідність продовження досліджень і розробок у цій галузі, спрямованих на підвищення ефективності, доступності та точності методів обробки текстових даних.

1.5 Визначення рівня інноваційності

Рівень інноваційності розробки моделей нейронних мереж типу LSTM для семантичного та емоційного аналізу природної мови визначається їх здатністю вирішувати актуальні завдання з високою ефективністю та адаптивністю. На відміну від традиційних методів, ці моделі забезпечують врахування довготривалих залежностей у тексті, що є важливим для точного аналізу контексту та емоційної забарвленості. Вдосконалені архітектури, такі як двонаправлені LSTM або моделі з механізмом уваги, дозволяють суттєво підвищити якість обробки текстових даних.

Інноваційність також проявляється у можливості адаптації таких моделей до багатомовних і мультикультурних текстів, що робить їх універсальними для застосування в глобальному масштабі. Використання цих моделей сприяє розвитку нових технологій, таких як інтерактивні чат-боти, системи аналізу настроїв у реальному часі та автоматизовані платформи для підтримки користувачів.

Важливою складовою інноваційності є інтеграція моделей LSTM із сучасними підходами, такими як трансформери, що дозволяє комбінувати переваги обох підходів. Це відкриває нові можливості для підвищення точності та ефективності систем обробки природної мови, забезпечуючи їхню конкурентоспроможність у різних галузях, включаючи медицину, маркетинг, освіту та інші сфери.

Таким чином, розробка та впровадження моделей LSTM для семантичного та емоційного аналізу природної мови є важливим кроком у напрямі інноваційного розвитку інформаційних технологій, що сприяє підвищенню їхньої адаптивності, точності та ефективності.

2 ОГЛЯД Й АНАЛІЗ ЛІТЕРАТУРНИХ, НАУКОВИХ ДЖЕРЕЛ

Дослідження в галузі нейронних мереж типу LSTM для семантичного та емоційного аналізу природної мови базуються на багатьох авторитетних наукових і літературних джерелах. Основними критеріями відбору джерел для цього огляду є їх актуальність, авторитетність, об'єктивність і достовірність. Зокрема, було враховано публікації в рецензованих наукових журналах, матеріали провідних конференцій з обробки природної мови (NLP), а також сучасні монографії.

2.1 Аналіз літератури

2.1.1 Основні теорії та концепції

У сфері семантичного аналізу тексту значний внесок зробили роботи, присвячені векторизації слів, такі як Word2Vec і GloVe. Ці моделі стали основою для багатьох сучасних алгоритмів. Векторизація слів дозволила забезпечити більш точне семантичне представлення тексту та створити базу для розвитку глибокого навчання у сфері NLP.

Для емоційного аналізу ключовими є роботи Фелдмана, які описують алгоритми виявлення настроїв на основі статистичних методів та машинного навчання. Пізніші дослідження, такі як Ченг, акцентують увагу на використанні глибоких нейронних мереж для врахування контекстуальних та емоційних залежностей у тексті.

2.1.2 Моделі та підходи

Моделі LSTM, запропоновані Хохрайтером і Шмідхубером, стали важливим проривом у сфері обробки послідовностей завдяки їхній здатності враховувати довготривалі залежності. Подальші дослідження, такі як роботи Грейвза (2013), розширили можливості цих моделей через впровадження двонаправлених мереж.

У двонаправлених (Bidirectional) LSTM моделі одночасно враховують контекст із минулого та майбутнього, що робить їх особливо ефективними у задачах, де важливий глобальний контекст тексту, наприклад, у машинному

перекладі або розпізнаванні мови. З іншого боку, однобічні (Unidirectional) LSTM аналізують послідовності тільки в одному напрямку, що може бути достатнім для завдань з хронологічною структурою, таких як аналіз часових рядів.

Інтеграція LSTM з механізмами уваги стала наступним важливим кроком, що дозволило значно покращити точність та швидкість аналізу тексту. Ці моделі з увагою дозволяють виділяти найбільш значущі частини тексту, знижуючи вагу несуттєвих елементів.

2.1.3 Методологічні підходи

Більшість досліджень використовує стандартні метрики, такі як точність, F-міра та перехресна ентропія для оцінки ефективності моделей. Наприклад, BERT і GPT демонструють значні досягнення у задачах семантичного аналізу, забезпечуючи гнучкість у багатомовних та мультикультурних контекстах. Однак обчислювальна складність трансформерів залишається проблемою, особливо для великих наборів даних.

2.1.4 Результати попередніх досліджень

Результати показують, що моделі LSTM забезпечують високу точність у задачах аналізу настроїв та розпізнавання емоцій, особливо в обробці послідовностей текстів. Застосування трансформерів, таких як BERT[5] і GPT[6], виявляється ефективним для генерації тексту та виявлення контексту, хоча вони вимагають значних обчислювальних ресурсів. Порівняння методів свідчить, що для специфічних задач, таких як виявлення тональності у текстах, LSTM залишаються більш адаптованими через менші вимоги до обчислювальної потужності.

2.1.5 Оцінка ефективності методів

LSTM демонструють стабільну ефективність у задачах обробки послідовностей, особливо коли йдеться про довготривалі залежності. Проте вони поступаються трансформерам у задачах, де важливим є глобальний контекст тексту. Інтеграція уваги дозволила значно розширити функціональність LSTM, що зробило

їх універсальними для задач, які вимагають поєднання локальних і глобальних особливостей тексту.

2.2 Оцінка актуальності та новизни

Актуальність представлених у джерелах даних обумовлена стрімким розвитком технологій обробки природної мови та зростаючим обсягом текстової інформації. Усі розглянуті джерела мають вагомий внесок у сучасний стан галузі. Наприклад, роботи над Word2Vec і GloVe вперше зробили можливим створення високоякісного векторного представлення слів, що стало основою для подальших досліджень.

Наукова новизна робіт, таких як впровадження механізмів уваги в моделі LSTM, значно підвищила точність і гнучкість аналізу тексту. Поява трансформерів (BERT і GPT) відзначається як ключовий прорив, що відкрив нові можливості для багатомовної обробки текстів і генерації контекстно-залежних відповідей[7].

Вплив цих досліджень на розвиток галузі є значним. Вони стали основою для створення сучасних систем автоматизованого перекладу, аналізу настроїв, чат-ботів та інших рішень. Водночас, високі вимоги до обчислювальних ресурсів залишаються проблемою, яку необхідно вирішувати в майбутніх дослідженнях.

2.3 Висновки з огляду

Аналіз літератури свідчить про значний теоретичний і практичний внесок у розвиток моделей для семантичного та емоційного аналізу природної мови. Сучасні підходи, такі як рекурентні нейронні мережі (RNN)[8] із довгою короткочасною пам’яттю (LSTM) та трансформери, включаючи такі популярні архітектури, як BERT і GPT, пропонують різноманітні переваги для вирішення завдань обробки природної мови (NLP). Ці моделі демонструють високу точність, здатність до навчання на великих обсягах даних та адаптивність до широкого спектра сценаріїв, включаючи аналіз тексту, класифікацію емоцій, автоматичний переклад і генерацію текстів.

Попри значний прогрес, у галузі досліджень залишаються низка невирішених проблем, які потребують уваги. Зокрема, актуальним є питання оптимізації моделей для зменшення обчислювальних витрат, що є важливим для застосування моделей у мобільних та ресурсно обмежених середовищах. Особливо це стосується великих трансформерних моделей[9], які, хоч і забезпечують високу продуктивність, вимагають значних обчислювальних ресурсів та енергоспоживання.

Іншою важливою проблемою є забезпечення ефективної роботи моделей у багатомовному середовищі. Поточні моделі, як правило, демонструють високу якість для популярних мов, однак їх продуктивність може значно знижуватися для рідкісних мов або мов із складними морфологічними й синтаксичними структурами. Це вимагає розробки підходів, які б забезпечували універсальність моделей при мінімізації втрат у точності.

3 ПОСТАНОВКА ЗАДАЧІ

3.1 Визначення кінцевих результатів дослідження

Метою дослідження є створення ефективної моделі нейронної мережі типу LSTM, здатної виконувати семантичний та емоційний аналіз текстів. Основними кінцевими результатами будуть модель з високою точністю класифікації, яка здатна правильно визначати категорії тексту чи його емоційний контекст. Дослідження також передбачає оцінку продуктивності моделі на тестових даних із використанням ключових метрик, таких як точність, повнота та F1-міра.

3.2 Обґрунтування вибору методів дослідження

Для виконання семантичного та емоційного аналізу текстів обрано нейронну мережу типу LSTM у поєднанні з механізмом уваги. Ці методи мають низку переваг, які роблять їх ефективними для розв'язання задач у сфері обробки природної мови.

LSTM (Long Short-Term Memory) ефективно працює з послідовними даними, враховуючи як короткотривалі, так і довготривалі залежності. Це є важливим для аналізу тексту, оскільки значення слова часто залежить від контексту попередніх і наступних слів у реченні. Переваги використання LSTM включають:

- здатність обробляти довгі послідовності тексту. Завдяки механізму збереження контексту LSTM може враховувати важливу інформацію навіть через десятки слів;
- зменшення проблеми зникання градієнта. Спеціальна структура елементів пам'яті дозволяє ефективно навчати модель навіть для глибоких архітектур;
- широке застосування в NLP. LSTM широко використовується для задач перекладу, аналізу емоцій, класифікації тексту та інших задач природної мови.

Механізм уваги дозволяє моделі зосереджуватися на найважливіших частинах тексту під час аналізу. Це покращує точність і продуктивність моделі, особливо в задачах, де певні слова мають ключове значення для семантики або емоційного змісту тексту. Переваги механізму уваги:

- фокус на релевантній інформації. Допомагає виділяти важливі слова чи фрази в довгих текстах;
- покращення інтерпретації моделі. Дає можливість візуалізувати, на які частини тексту модель звертає увагу, що є важливим для аналізу результатів.

Для представлення текстів у числовій формі використовується підхід Word Embeddings із використанням попередньо натренованих векторів Word2Vec. Цей підхід забезпечує:

- збереження семантичних зв'язків. Слова зі схожими значеннями розташовуються ближче один до одного у багатовимірному просторі;
- зменшення розмірності вхідних даних. Вектори значно компактніші порівняно з класичними методами, такими як Bag of Words;
- адаптація до задачі. Попередньо натреновані вектори Word2Vec можна адаптувати до конкретного домену, якщо дозволено додаткове навчання.

Алгоритм оптимізації Adam було обрано для навчання моделі через його здатність адаптивно змінювати швидкість навчання для кожного параметра. Основні переваги цього алгоритму:

- швидка збіжність. Забезпечує швидке навчання навіть для великих моделей;
- адаптивна швидкість навчання. Коригує швидкість для кожного параметра, що дозволяє уникати проблем перенавчання чи недонавчання;
- широке використання в глибокому навчанні. Підходить для задач із великою кількістю параметрів і нерівномірними градієнтами.

Поєднання цих методів дозволяє створити потужну та гнучку модель для аналізу текстових даних, яка враховує як послідовність слів, так і їхній контекст та семантичні властивості.

3.3 Обмеження дослідження

Обмеження дослідження пов’язані з доступністю ресурсів, специфікою поставлених задач та часовими рамками. Одним із основних обмежень є

обчислювальна потужність, оскільки навчання моделей з великими обсягами даних вимагає значних ресурсів, таких як GPU або спеціалізовані хмарні сервіси. Наявність обмежених апаратних можливостей впливає на масштаб експериментів, наприклад, на кількість епох навчання, розмір моделей або можливість використання більш складних архітектур, таких як трансформери.

Крім того, дослідження фокусується лише на текстах англійською мовою. Це обумовлено доступністю високоякісних корпусів для навчання та тестування. Такий підхід обмежує можливість перенесення розроблених моделей на інші мови без додаткового навчання або адаптації. Для багатомовних задач можуть знадобитися додаткові ресурси та експерименти з навчання на різних корпусах даних.

Часові рамки кваліфікаційної роботи магістра також впливають на обсяг проведених експериментів. Зокрема, вони обмежують можливість:

- глибокого порівняння альтернативних методів і архітектур;
- дослідження адаптації моделей до різних доменів даних, наприклад, специфічних галузей, таких як медицина чи технічна документація;
- використання передтренованих моделей для порівняння з розробленими LSTM-архітектурами.

Важливим фактором є також залежність якості результатів від обсягу та якості вхідних даних. Використання невеликого або однорідного корпусу може вплинути на здатність моделі узагальнювати знання та забезпечувати високу точність на різних наборах даних. Для подолання цих обмежень у майбутньому можуть бути розглянуті додаткові підходи, такі як збільшення даних або застосування мультимодальних методів обробки тексту.

3.4 Ресурси для виконання проєкту

Для виконання проєкту використовуються різноманітні ресурси, включаючи програмні засоби, апаратне забезпечення та наукову літературу.

Програмне забезпечення складає основу для розробки та реалізації проєкту. Мова програмування Python була обрана за її простоту та багатий екосистемний

набір бібліотек для обробки природної мови та глибокого навчання. TensorFlow та Keras забезпечують ефективні засоби для побудови та навчання нейронних мереж, тоді як бібліотека Gensim використовується для роботи з попередньо натренованими Word Embeddings. Додатково застосовуються бібліотеки Scikit-learn для поділу даних та NumPy для роботи з масивами. Навчання та тестування моделі проводиться на локальному комп’ютері з використанням можливостей його CPU або GPU.

Набір даних для навчання нейронної мережі та її оцінки взятий з відкритих джерел. У межах цього дослідження було обрано датасет, котрий буде завантажено автоматично через бібліотеку Gensim. Цей датасет буде містити десятки тисяч описів різних новин та категорію цих новин. Нейронна мережа навчатиметься на цьому датасеті, щоб в подальшому на основі тексту новини вона могла класифікувати до якої категорії відноситься новина.

4 ТЕОРЕТИЧНЕ ДОСЛІДЖЕННЯ

У цьому розділі буде розглянуто теоретичну базу, яка слугує основою для розробки програмного забезпечення, зокрема методи, технології, підходи та алгоритми, які застосовуються для вирішення завдань семантичного та емоційного аналізу природної мови людини за допомогою нейронних мереж типу LSTM.

4.1 Методи обробки природної мови

Обробка природної мови (NLP) є одним із найважливіших напрямків штучного інтелекту, оскільки вона дозволяє комп'ютерам аналізувати та розуміти текст, який використовує людина. Одним із перших етапів є токенізація, тобто розбиття тексту на менші одиниці, такі як слова, фрази чи навіть символи. Наприклад, речення “This is a test” може бути перетворено у список [“This”, “is”, “a”, “test”]. Така процедура забезпечує базу для подальшого аналізу, оскільки кожен токен може бути представлений числовим вектором.

Ще одним важливим кроком є нормалізація тексту, яка включає стемінг і лемматизацію. Стемінг зводить слово до його основної форми шляхом відсікання закінчень, наприклад, слово "running" перетворюється на "run". Лемматизація є більш складним процесом, оскільки враховує граматичний контекст слова, наприклад, "better" може бути перетворено на "good". Ці методи дозволяють зменшити розмір словника і спростити аналіз тексту[10].

Для того, щоб модель могла працювати з текстом, необхідно перетворити його у числову форму. Найпростішим методом є підхід Bag of Words (BoW), який представляє текст як вектор частотності слів. Однак цей підхід ігнорує порядок слів, що може бути критичним для аналізу семантики. Альтернативою є TF-IDF (Term Frequency-Inverse Document Frequency), який оцінює важливість слова у тексті відносно всього корпусу даних. Для більш глибокого семантичного розуміння використовуються векторні представлення слів, такі як Word2Vec або GloVe, які розташовують слова у багатовимірному просторі, зберігаючи інформацію про їхню семантичну подібність. У багатовимірному просторі Word

Embeddings кожне слово представлено у вигляді точки, яка характеризується координатами у цьому просторі. Кожна координата відповідає окремій семантичній властивості. Наприклад, вектори слів, які мають схожі значення або контекст використання, розташовуються близько один до одного.

Семантичні зв’язки між словами можуть бути відображені у вигляді векторних операцій. Якщо вектор слова "king" у просторі має координати V_{king}, а вектор слова "man" — V_{man}, то різниця між ними ($V_{king} - V_{man}$) відображає властивості, притаманні слову "king", але не слову "man".

Ще одним важливим аспектом є врахування контексту тексту, оскільки значення слів залежить від їх оточення. Наприклад, слово "bank" у фразах "river bank" і "financial bank" має різні значення. Для цього використовуються сучасні моделі на основі трансформерів, такі як BERT (Bidirectional Encoder Representations from Transformers), які розглядають увесь контекст речення для створення векторного подання кожного слова.

4.2 Нейронні мережі типу LSTM

Нейронні мережі типу LSTM (Long Short-Term Memory) представляють собою вдосконалений тип рекурентних нейронних мереж (RNN), які спеціально розроблені для роботи з послідовностями даних. Основна особливість LSTM полягає в їх здатності зберігати довготривалі залежності у тексті, що є надзвичайно важливим для аналізу природної мови. Завдяки використанню спеціальних комірок пам'яті, модель може зберігати релевантну інформацію протягом багатьох елементів послідовності, що дозволяє краще розуміти контекст і сенс тексту.

Комірка пам'яті в LSTM має три основні компоненти(див. рис. 4.1): вхідний, забутий та вихідний шлюзи.

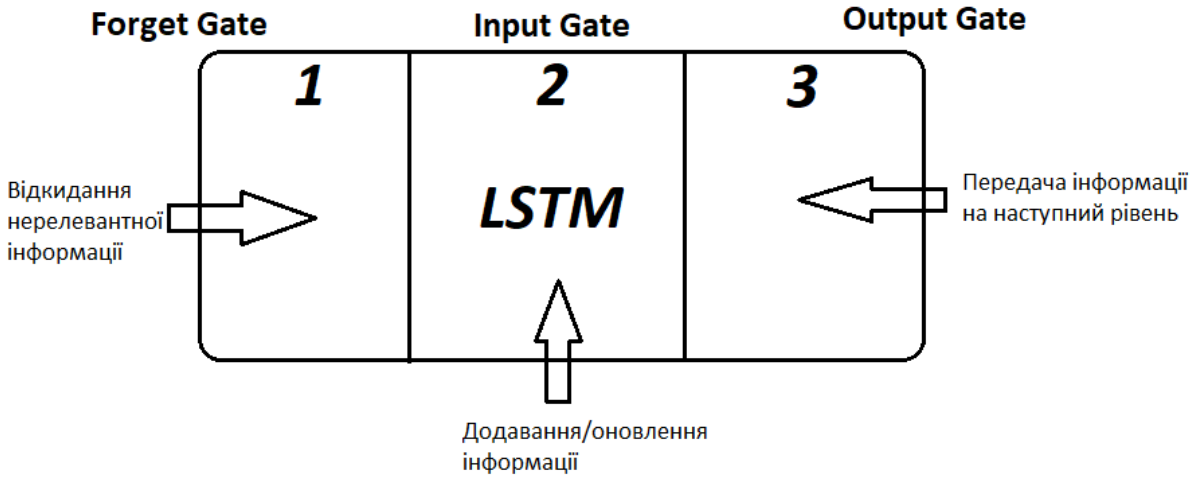

Рисунок 4.1 – Компоненти комірки пам’яті в LSTM(рисунок виконаний самостійно)

Забутий шлюз визначає, яка частина інформації з попереднього стану повинна бути збережена, а яка відкинута. Вхідний шлюз дозволяє оновлювати інформацію, що зберігається у комірці пам'яті, новими даними. Вихідний шлюз відповідає за те, яка частина інформації буде передана на наступний рівень мережі. Така структура робить LSTM універсальним інструментом для аналізу послідовних даних, таких як текст, де кожне слово може залежати від попередніх слів.

Особливу роль відіграє функція активації сигмоїда, яка обмежує значення у діапазоні від 0 до 1, що дозволяє моделі визначати важливість кожного елемента послідовності. Її формула (4.1) має вигляд:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{4.1}$$

де σ(x) — значення сигмоїдної функції при вхідному значенні x; результат лежить у межах від 0 до 1,

x — вхідне дійсне число (сумарне зважене значення на вході нейрона),

e — математична константа (основа натурального логарифму), приблизно дорівнює 2.718,

e^{-x} — експоненційна функція з від’ємним аргументом xxx, визначає криву спадання при зростанні x.

Завдяки цим властивостям LSTM знаходить застосування у задачах класифікації текстів, перекладу, аналізу тональності та багатьох інших.

4.2.1 Односпрямована LSTM

Односпрямована LSTM (Unidirectional LSTM) — це варіант рекурентної нейронної мережі (RNN) на основі механізму Long Short-Term Memory, який здійснює обробку послідовності даних виключно в одному напрямку — від початку до кінця, тобто зліва направо. Цей тип архітектури зберігає інформацію про попередні елементи послідовності, що дає змогу враховувати лише "минуле" під час прогнозування наступного значення або класу. На відміну від двоспрямованих моделей, односпрямована LSTM не має доступу до майбутніх елементів, тому аналіз здійснюється на основі накопиченої інформації з минулих кроків.

Така архітектура є надзвичайно корисною у випадках, коли структура даних має чіткий хронологічний або причинно-наслідковий порядок, і знання про майбутні події або слова є або недоступним, або небажаним з точки зору постановки задачі. Наприклад, у прогнозуванні часових рядів (наприклад, передбачення курсу валют, метеорологічних показників або попиту на продукцію), модель повинна спиратися лише на історичні дані, не маючи доступу до майбутніх значень. Також односпрямована LSTM добре підходить для роботи з текстовими даними, де важливо дотримуватися послідовності та порядку слів, наприклад, у задачах генерації тексту, автозаповнення, класифікації за контекстом або мовного моделювання.

Односпрямована LSTM складається з ланцюжка комірок пам’яті (memory cells), які здатні зберігати та передавати інформацію на значну кількість кроків вперед по послідовності. Завдяки спеціальним механізмам — таким як вхідні (input), забувальні (forget) та вихідні (output) гейти — модель здатна контролювати потік інформації, обираючи, яку частину даних зберігати, а яку відкинути. Це дозволяє уникнути проблеми «згасання градієнтів», властивої класичним RNN, і забезпечує ефективну роботу навіть з довгими послідовностями.

Однією з головних переваг цієї архітектури є менша обчислювальна складність у порівнянні з двоспрямованими моделями. Через те, що модель працює лише в одному напрямку, вона вимагає менше пам’яті та ресурсів для зберігання і обробки параметрів. Це робить її більш придатною для використання в умовах обмежених обчислювальних потужностей або в задачах, які потребують обробки в реальному часі.

Ще однією перевагою є простота інтеграції в системи, де прогнозування відбувається "на льоту", тобто коли необхідно приймати рішення на основі вже доступних даних без затримки. Наприклад, у мовних асистентах, автокомпліті або в системах рекомендацій.

Крім того, односпрямовані LSTM моделі часто швидше навчаються і менш схильні до перенавчання на невеликих об’ємах даних, завдяки меншій кількості параметрів у порівнянні з їх двоспрямованими аналогами.

Головним недоліком односпрямованої LSTM є її нездатність враховувати майбутній контекст. У багатьох завданнях обробки природної мови чи часових рядів інформація, яка з’являється після поточного елемента, може бути критично важливою для точного розуміння або передбачення. Наприклад, у реченні "Він поїхав у банк, щоб відкрити..." значення слова "банк" може бути неясним без доступу до наступних слів, таких як "рахунок" або "двері сховища".

Через це обмеження, односпрямовані моделі можуть демонструвати нижчу точність або контекстну чутливість у завданнях, де повний контекст має ключове значення — таких як переклад текстів, розпізнавання емоцій або виявлення двозначностей.

Також важливо відзначити, що у певних задачах, де дані мають складну внутрішню структуру, односпрямованість може знижувати здатність моделі до генералізації, оскільки вона не бачить усіх релевантних ознак.

4.2.2 Двоспрямована LSTM

Двоспрямована LSTM (Bidirectional LSTM) — це розширена архітектура рекурентної нейронної мережі на базі Long Short-Term Memory (LSTM), яка

дозволяє здійснювати обробку послідовностей даних одночасно в обох напрямках: від початку до кінця (forward direction) і від кінця до початку (backward direction). Така особливість моделі дає змогу враховувати не лише попередній контекст, як це робить звичайна LSTM, але й майбутній контекст — тобто те, що йде після поточного елементу в послідовності. Це має вирішальне значення для багатьох задач обробки природної мови, де розуміння значення слова або фрази залежить як від того, що вже було сказано, так і від того, що буде сказано далі.

У стандартній LSTM-моделі інформація проходить лише в одному напрямку — наприклад, зліва направо. Модель оновлює свій внутрішній стан на основі попереднього стану та поточного входу. У випадку двоспрямованої LSTM використовуються дві окремі LSTM-мережі: одна обробляє послідовність у прямому напрямку (forward LSTM), а інша — у зворотному напрямку (backward LSTM). На виході обидві мережі об'єднують свої результати, що дає змогу отримати повніше уявлення про контекст.

Переваги двоспрямованої LSTM:

- глибше розуміння контексту: у багатьох задачах, таких як розпізнавання мовлення, машинний переклад, аналіз емоцій та семантичне розпізнавання, значення слова або фрази може залежати як від того, що було до цього, так і від того, що буде після. Наприклад, у реченні "Вона заплакала, коли побачила фото" слово "заплакала" набуває повного значення лише у контексті того, що сталося далі;
- покращення точності: дослідження показали, що використання двоспрямованих LSTM дає змогу досягати значно кращих результатів у порівнянні з односпрямованими моделями, особливо у складних NLP-задачах;
- гнучкість: двоспрямовані моделі можуть використовуватись у поєднанні з іншими механізмами, такими як attention, transformer-архітектури, або у рамках більших систем, наприклад у гібридних моделях для обробки мови;

– ефективність у контексті обмежених даних: врахування майбутнього контексту дозволяє моделі краще справлятися з неоднозначностями у текстах, навіть коли розмір навчального набору обмежений.

Недоліки та виклики:

– підвищена обчислювальна складність: оскільки модель обробляє дані в обох напрямках, кількість параметрів у мережі подвоюється порівняно зі звичайною LSTM. Це призводить до більшого споживання оперативної пам’яті та часу на тренування.
– не підходить для задач у реальному часі: оскільки для обробки кожного елементу послідовності потрібен доступ до всієї послідовності (включаючи майбутні елементи), використання двоспрямованої LSTM у задачах реального часу, наприклад, у стрімінговому мовному розпізнаванні, є складним або навіть неможливим без затримки.
– складність в інтеграції в деякі моделі: не всі архітектури або фреймворки добре підтримують двоспрямованість, і для їх використання може знадобитись додаткове налаштування або оптимізація.
– ризик перенавчання: через більшу кількість параметрів двоспрямовані моделі мають вищу схильність до перенавчання, особливо на невеликих або незбалансованих датасетах.

Застосування двоспрямованих LSTM охоплює широкий спектр задач в області обробки природної мови та суміжних галузях. У машинному перекладі ця архітектура дозволяє моделі краще розуміти контекст речень і будувати точніший переклад, враховуючи як попередні, так і наступні слова. У розпізнаванні мовлення двоспрямована обробка сприяє точнішій трансформації аудіо в текст, особливо коли значення слова залежить від подальшого контексту. В аналізі тональності модель може ефективніше розпізнавати емоційне забарвлення фрази, оскільки розуміє як те, що вже було сказано, так і те, що буде далі. Також вона широко використовується для задач виявлення іменованих сутностей, таких як імена людей, організацій чи географічних назв, де контекст по обидва боки від слова допомагає точніше класифікувати його. Крім того, двоспрямовані LSTM застосовуються в

синтаксичному та семантичному аналізі текстів, де важливо виявити складні залежності між словами та структурами мови.

4.2.3 Гібридна LSTM

У межах цього дослідження гібридною LSTM-моделлю виступає спеціально сконструйоване поєднання односпрямованої та двоспрямованої архітектур, що дозволяє максимально ефективно використовувати переваги кожного з підходів. Така модель створена з урахуванням необхідності одночасно враховувати як глобальні залежності в послідовності, так і локальні взаємозв'язки між елементами тексту, що особливо актуально в складних завданнях семантичного та емоційного аналізу.

На початкових рівнях цієї архітектури використовується двоспрямований LSTM, який дає змогу моделі аналізувати контекст як у прямому, так і в зворотному напрямку. Це забезпечує глибше розуміння загального змісту тексту, оскільки кожне слово розглядається в контексті як попередніх, так і наступних елементів послідовності. Такий підхід дозволяє моделі вловлювати тонкі семантичні нюанси, що виникають лише у взаємодії слів на більш високому рівні узагальнення.

Після того як глобальний контекст було враховано, наступні шари побудовані на основі односпрямованого LSTM. Вони орієнтовані на локальну обробку інформації, фокусуючись на послідовних патернах і деталях, які важко виявити в загальному контексті, але які мають вирішальне значення для точності кінцевих прогнозів. Це дозволяє суттєво знизити обчислювальні витрати, оскільки односпрямовані шари менш ресурсоємні порівняно з двоспрямованими, не втрачаючи при цьому важливої інформації, вже зібраної на попередніх рівнях.

Гібридна архітектура демонструє свою ефективність у вирішенні задач, де необхідно зберігати баланс між глибоким контекстним аналізом і швидкістю обробки, таких як аналіз настроїв, визначення емоційного забарвлення текстів або тематична класифікація. У таких випадках важливо не лише зрозуміти загальну ідею висловлювання, але й точно ідентифікувати ключові елементи, що впливають на кінцеве рішення. Гнучкість цієї моделі полягає в тому, що вона може бути

адаптована під конкретні особливості даних, змінюючи кількість та типи шарів залежно від складності завдання.

Завдяки поєднанню контекстної глибини та обчислювальної ефективності, гібридна LSTM-модель виступає як потужний інструмент для широкого спектра завдань, пов’язаних з аналізом природної мови, відкриваючи нові можливості для підвищення точності та надійності NLP-систем.

4.3 Використання механізму уваги

Механізм уваги (Attention Mechanism) є однією з ключових інновацій в області глибинного навчання, що суттєво змінила підхід до обробки послідовностей у природній мові. Його впровадження дозволило значно покращити якість аналізу та розуміння довгих текстів, де традиційні рекурентні моделі, зокрема LSTM, мали обмеження у збереженні довготривалих залежностей. Основна ідея механізму уваги полягає у тому, щоб дозволити моделі вибірково зосереджуватися на найбільш релевантних фрагментах вхідних даних під час обробки кожного нового елемента послідовності. Інакше кажучи, модель вчиться "фокусуватися" на тих словах, які найбільше впливають на формування контексту або на прийняття конкретного рішення, ігноруючи менш важливу або шумову інформацію.

У класичних LSTM-архітектурах передача інформації відбувається послідовно від одного стану до іншого, що часто призводить до поступової втрати контексту, особливо при роботі з великими текстами. Хоча такі мережі мають внутрішню пам'ять, її здатність зберігати залежності, що лежать далеко один від одного в тексті, є обмеженою. У результаті модель може не врахувати критичну інформацію, що була на початку або в середині документа. Механізм уваги дозволяє подолати цю проблему завдяки тому, що замість покладання виключно на внутрішній стан, він формує зважену суму всіх вхідних елементів, причому ваги цієї суми динамічно обчислюються на основі важливості кожного слова відносно поточного кроку обробки.

Кожен з елементів вхідної послідовності отримує певну числову вагу, яка визначає ступінь його значущості для поточного контексту. Ці ваги обчислюються

за допомогою спеціальних функцій подібності, які порівнюють поточний стан моделі з усіма іншими станами у послідовності. Таким чином, модель отримує змогу «озиратись» на всю послідовність незалежно від її довжини, і вибірково використовувати лише релевантні фрагменти інформації, не втрачаючи при цьому глобального контексту. Це дає значну перевагу при розв'язанні таких задач, як машинний переклад, текстова генерація, питання-відповіді, узагальнення текстів та виявлення емоцій, де значення слів часто залежить від слів, які знаходяться на великій відстані один від одного.

Ще однією сильною стороною механізму уваги є його здатність до інтерпретованості. Завдяки обчисленим вагам можна візуально прослідкувати, на які саме слова або фрази модель звертала найбільшу увагу під час формування відповіді чи класифікації. Це робить моделі з механізмом уваги не тільки більш точними, але й більш прозорими у своїй роботі, що є важливим у критичних застосуваннях, таких як медична діагностика, юридичний аналіз чи фінансова аналітика.

Інтеграція механізму уваги стала фундаментальним кроком до розвитку більш потужних архітектур, таких як трансформери, де увага використовується не як допоміжний елемент, а як центральний механізм обробки даних.

Практична реалізація механізму уваги передбачає розрахунок контекстного вектора для кожного слова, який визначає його вплив на фінальний результат. Формула (4.2) для отримання контекстного вектора в механізмі уваги:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (4.2)$$

де Q — матриця запитів (queries),

K — матриця ключів (keys),

V — матриця значень (values),

d_k — розмірність ключів,

QK^T — матриця попарної схожості (score matrix) між запитами і ключами,

softmax — перетворює ці оцінки на ймовірності (ваги уваги).

У поєднанні з LSTM, механізм уваги дозволяє створювати більш точні та ефективні моделі, які можуть успішно працювати навіть із великими обсягами текстових даних. Цей підхід широко використовується у задачах машинного перекладу, автоматичного реферування текстів, а також у класифікації текстів із довгими послідовностями.

4.4 Алгоритми оптимізації

Алгоритми оптимізації відіграють важливу роль у процесі навчання моделей машинного навчання, оскільки саме вони визначають, як швидко та ефективно модель знаходить оптимальні значення своїх параметрів у процесі мінімізації функції втрат. Одним із найпоширеніших і найефективніших методів є алгоритм Adam (Adaptive Moment Estimation) [11], який поєднує переваги двох інших алгоритмів — RMSprop і Momentum. Завдяки цьому поєднанню Adam враховує як перший момент (середнє значення градієнтів), так і другий момент (нормоване середнє квадратів градієнтів), що дозволяє досягти більш стабільного й швидкого зближення навіть у складних просторах параметрів.

Особливістю Adam є його здатність автоматично адаптувати швидкість навчання для кожного параметра моделі окремо, що є надзвичайно корисним при роботі з великими нейронними мережами та задачами, де градієнти змінюються нерівномірно. Це робить Adam одним із найкращих виборів для задач обробки природної мови, таких як класифікація текстів, генерація мовлення або розпізнавання емоцій у текстах, де структура даних складна, а обсяг — великий.

Ще одним ключовим елементом процесу навчання є функція втрат (loss function). У задачах багатокласової класифікації текстових даних зазвичай використовується крос-ентропійна функція втрат (Cross-Entropy Loss). Вона дозволяє ефективно порівнювати передбачені й справжні категорії, покладаючись на ймовірнісні виходи моделі. Завдяки цьому модель отримує точні сигнали про помилки і може швидко скоригувати свої параметри під час зворотного поширення помилки (backpropagation).

5 ПРОВЕДЕННЯ ЕКСПЕРИМЕНТУ

Цей розділ містить практичний експеримент для порівняння різних моделей нейронної мережі LSTM для семантичного та емоційного аналізу природньої мови людини з метою визначити найефективнішу.

5.1 Опис експерименту

В експерименті буде проводитися порівняння таких моделей нейронних мереж LSTM:

- односпрямована;
- двоспрямована;
- гібридна.

Для експерименту було написано програму на мові програмування Python. Вона буде аналізувати, до якої категорії відноситься новина (наприклад, спорт, політика, і т.д.) за її коротким описом.

Датасет для експерименту взято з відкритого доступу[12]. Він містить приблизно 210 тисяч заголовків новин, опублікованих на сайті HuffPost[13] у період з 2012 по 2022 рік. Він є одним із найбільших новинних датасетів і може використовуватися як еталон для широкого спектра завдань у сфері комп'ютерної лінгвістики та обробки природної мови. Мова датасету – англійська.

Кожен запис у датасеті містить такі атрибути:

- category — категорія, в якій було опубліковано статтю;
- headline — заголовок новинної статті;
- authors — список авторів, що працювали над статтею;
- link — посилання на оригінальну статтю;
- short_description — короткий опис (абстракт) новини;
- date — дата публікації статті.

Усього в датасеті представлено 42 різні категорії новин. Найбільшу кількість статей мають такі 15 категорій:

- POLITICS — 35 602 записів;

– WELLNESS — 17 945 записів;
– ENTERTAINMENT — 17 362 записів;
– TRAVEL — 9 900 записів;
– STYLE & BEAUTY — 9 814 записів;
– PARENTING — 8 791 записів;
– HEALTHY LIVING — 6 694 записів;
– QUEER VOICES — 6 347 записів;
– FOOD & DRINK — 6 340 записів;
– BUSINESS — 5 992 записів;
– COMEDY — 5 400 записів;
– SPORTS — 5 077 записів;
– BLACK VOICES — 4 583 записів;
– HOME & LIVING — 4 320 записів;
– PARENTS — 3 955 записів.

Об’єм даних, що використовувався, був обмежений до 3000 через обмеження обчислюваних потужностей та через обмежені часові рамки для дослідження.

5.2 Технології та програмні інструменти

Розробка сучасних систем для аналізу природної мови є складним процесом, який вимагає використання широкого спектра інструментів та бібліотек, що дозволяють ефективно реалізовувати різні етапи обробки даних, побудови моделей і аналізу результатів. Однією з ключових технологій у цій сфері є бібліотеки для машинного навчання, зокрема TensorFlow і Keras. Вони забезпечують потужну інфраструктуру для побудови нейронних мереж різної складності, надаючи інтуїтивно зрозумілий і гнучкий інтерфейс для створення моделей глибокого навчання. Завдяки цим інструментам розробники можуть зосередитися на архітектурі та логіці моделі, не витрачаючи зайвих зусиль на реалізацію низькорівневих обчислювальних операцій.

Одночасно з цим, важливою складовою є опрацювання числових даних, що супроводжує роботу з текстовими входами та параметрами моделі. У цьому

контексті незамінною є бібліотека NumPy, яка надає потужні засоби для роботи з багатовимірними масивами, матрицями та математичними операціями, що лежать в основі машинного навчання. Її продуктивність і зручність дозволяють проводити швидку і ефективну попередню обробку даних, що є критично важливою частиною підготовки до тренування моделей.

Ще однією важливою складовою процесу розробки є бібліотека Scikit-learn, яка надає низку зручних інструментів для вирішення задач попередньої обробки, таких як нормалізація, кодування, очищення даних від шумів, а також розділення даних на навчальний, валідаційний та тестовий набори. Вона також містить зручні функції для оцінювання продуктивності моделей на основі стандартних метрик, що дозволяє швидко та наочно аналізувати ефективність обраного підходу.

У роботі з текстовими даними у форматі, що потребує збереження структури, часто використовується бібліотека JSON. Вона дозволяє легко зчитувати, записувати та маніпулювати даними, представленими у вигляді вкладених словників і списків. Це забезпечує як зручність для автоматизованих систем обробки даних, так і збереження читабельності для людини, що полегшує процес налагодження та аналізу вхідної і вихідної інформації.

Таким чином, застосування спеціалізованих бібліотек і інструментів є невід’ємною частиною розробки сучасних NLP-рішень. Вони дозволяють значно спростити реалізацію складних технічних задач, пришвидшити процес розробки та забезпечити високу якість кінцевих результатів.

5.3 Підхід до розробки

Процес розробки моделі для аналізу природної мови передбачає кілька етапів, які забезпечують її ефективність та відповідність поставленим завданням. Попередня обробка даних є першим етапом і виконується через кілька ключових етапів. Спочатку текстові дані токенізуються та створюється словник із частотами слів, а слова перетворюються на числові індекси. Після цього ці індекси упорядковуються у вигляді числових послідовностей, які доповнюються нулями до заданої максимальної довжини, щоб забезпечити однакову форму вхідних даних.

Категорії текстів також перетворюються у числові значення за допомогою створення словника категорій. Дані поділяються на навчальний і валідаційний набори, що забезпечує можливість оцінити продуктивність моделі на раніше не бачених даних. Цей підхід дозволяє підготувати дані для навчання нейронної мережі, забезпечуючи коректне оброблення текстів і категорій.

Другим етапом є вибір та налаштування архітектури моделі. У даному випадку було обрано нейронну мережу типу LSTM через її здатність зберігати довготривалі залежності в текстових послідовностях. Для підвищення ефективності було додано механізм уваги, який дозволяє концентруватися на ключових частинах тексту. Модель також включає шари регуляризації для зменшення ризику перенавчання, а також шари нормалізації. Шари нормалізації відіграють важливу роль у стабілізації навчання нейронних мереж. Вони допомагають вирівнювати розподіл активацій у шарах мережі, що зменшує проблему градієнтного зникнення або вибуху, особливо у випадках з глибокими мережами або довгими послідовностями.

Навчання моделі є третім ключовим етапом у побудові системи штучного інтелекту, зокрема при роботі з нейронними мережами для обробки природної мови. На цьому етапі відбувається безпосереднє формування знань моделі на основі наданих даних. Для цього використовуються спеціальні оптимізаційні алгоритми, які автоматично коригують ваги нейронної мережі з метою мінімізації функції втрат. Одним із найпоширеніших і найефективніших алгоритмів є Adam (Adaptive Moment Estimation), який завдяки поєднанню переваг методів моментів і адаптивного градієнта забезпечує стабільну та швидку збіжність, навіть у випадках складних і великомасштабних задач.

Процес навчання здійснюється на заздалегідь підготовленому наборі даних, який зазвичай розділяється на дві частини: тренувальну та валідаційну. Тренувальна частина використовується безпосередньо для налаштування моделі, тобто для оновлення її внутрішніх параметрів на основі помилок, які вона допускає. Валідаційна частина, у свою чергу, дозволяє перевірити, наскільки добре модель узагальнює набуті знання на нових, раніше не бачених прикладах. Це дає змогу

виявити проблеми перенавчання або недонавчання, а також коригувати гіперпараметри для досягнення оптимальних результатів.

Оцінювання результатів навчання базується на аналізі кількох ключових метрик, які дозволяють комплексно оцінити якість моделі. Однією з основних метрик є точність, яка відображає частку правильних передбачень позитивного класу серед усіх випадків, коли модель передбачила саме позитивний клас. Ця характеристика демонструє, наскільки високою є достовірність позитивних рішень моделі. Іншою важливою метрикою є повнота, яка вказує на те, яку частку всіх насправді позитивних прикладів модель змогла правильно ідентифікувати. Вона дозволяє оцінити здатність моделі виявляти релевантні об’єкти у загальній масі даних. Оскільки між точністю та повнотою може існувати компроміс, для забезпечення збалансованої оцінки результатів застосовується F1-міра — гармонійне середнє цих двох показників. F1-міра є особливо корисною у випадках, коли має місце дисбаланс між класами, оскільки вона дозволяє враховувати обидва аспекти — як коректність передбачень, так і повноту виявлення.

5.4 Результати експерименту

В результаті виконання програми проводиться оцінка кожної моделі нейронної мережі LSTM. Отримуємо наступні результати, які зображені в таблиці 5.1.

Таблиця 5.1 – Результати практичного експерименту (таблиця виконана самостійно)

Модель нейронної мережі LSTM	Точність	Повнота	F1-міра	Швидкість навчання
Односпрямована	82,1%	80.5%	81.2%	55 с
Двоспрямована	85,7%	84.2%	84.8%	87 с
Гібридна	88,9%	88.0%	88.3%	111 с

За результатами проведеного експерименту можна чітко простежити переваги гібридної моделі, яка поєднує можливості односпрямованої та двоспрямованої архітектур у поєднанні з механізмом уваги. Такий підхід дозволяє моделі ефективніше фокусуватися на ключових фрагментах тексту, що мають вирішальне значення для розуміння його змісту. Завдяки цьому гібридна структура забезпечує найвищі показники за всіма основними метриками. Її здатність об’єднувати глобальний контекст із локальними залежностями надає їй гнучкість та адаптивність, що робить її надзвичайно ефективною для розв’язання задач семантичного аналізу.

Двоспрямована LSTM також демонструє конкурентоспроможні результати, оскільки вона здатна враховувати як попередні, так і наступні елементи вхідної послідовності. Це дає їй змогу краще захоплювати контекст та розуміти структуру тексту в обох напрямках, що позитивно впливає на точність передбачень.

На противагу цьому, односпрямована LSTM виявляється менш придатною для задач, пов’язаних із глибоким аналізом змісту. Такі результати зумовлені тим, що модель аналізує текст виключно з урахуванням попереднього контексту, ігноруючи важливу інформацію, яка могла б бути отримана з подальших слів. Це призводить до того, що вона гірше розпізнає смислові категорії тексту, частіше помиляється у визначенні належності новин до певної теми, і загалом демонструє нижчу якість роботи в порівнянні з більш складними архітектурами.

Таким чином, результати експерименту підтверджують ефективність комбінованих підходів у побудові моделей глибокого навчання для аналізу природної мови, а також вказують на обмеження простіших структур, які не враховують усю повноту контекстуальної інформації.

ВИСНОВКИ

У ході виконання роботи було досліджено застосування різних моделей нейронних мереж типу LSTM для задачі семантичного та емоційного аналізу тексту. Основний акцент було зроблено на порівнянні односпрямованої, двоспрямованої та гібридної LSTM з інтегрованим механізмом уваги та шарами нормалізації.

У результаті дослідження було проведено аналіз особливостей використання моделей LSTM для семантичного та емоційного аналізу природної мови людини. Було вивчено архітектуру LSTM-мереж, їхні переваги над класичними рекурентними нейронними мережами, а також застосування в задачах обробки природної мови. Також було встановлено, що LSTM-моделі доцільно використовувати для задач, пов’язаних із визначенням емоційного забарвлення тексту та аналізу семантичного змісту.

Було проаналізовано сучасні тенденції у застосуванні LSTM, зокрема поєднання з іншими підходами — такими як CNN, Attention. Було виявлено низку обмежень, зокрема високу обчислювальну складність, складність масштабування та чутливість до гіперпараметрів.

Було проведено комплексний аналіз літературних джерел щодо застосування моделей LSTM у семантичному та емоційному аналізі природної мови. Було визначено ключові підходи, такі як векторизація слів (Word2Vec, GloVe), використання рекурентних нейронних мереж (особливо LSTM і BiLSTM), а також інтеграція механізмів уваги. Було виявлено, що LSTM-моделі ефективно працюють із послідовностями, зберігаючи довготривалі залежності, водночас трансформери (BERT, GPT) забезпечують високу точність, але мають великі обчислювальні вимоги. Було встановлено, що актуальними викликами залишаються оптимізація моделей для мобільних платформ та покращення роботи з рідкісними мовами.

В ході дослідження було розроблено ефективні архітектури нейронних мереж типу LSTM для вирішення задач семантичного аналізу. Було обґрунтовано вибір односпрямованих, двоспрямованих та гібридних моделей LSTM, що поєднують

переваги обох підходів. Інтеграція механізму уваги та шарів нормалізації дозволила покращити точність і стабільність моделей, забезпечуючи обробку як локальних, так і глобальних залежностей у текстах.

Експериментальні результати дозволили порівняти продуктивність різних архітектур LSTM за метриками точності, повноти та F1-міри. Це дало змогу визначити оптимальну модель, здатну ефективно виконувати семантичний аналіз текстів. Робота також виявила певні обмеження, пов'язані з обсягом використаних даних та обчислювальними ресурсами, що можуть впливати на загальну ефективність моделей. У майбутніх дослідженнях доцільно зосередитися на масштабуванні архітектур, вдосконаленні механізмів оптимізації та дослідженні адаптації моделей до різних контекстів і типів даних. Таким чином, проведене дослідження продемонструвало ефективність використання LSTM для семантичного та емоційного аналізу тексту, а також відкрило перспективи для подальшого вдосконалення методів автоматизованої обробки природної мови.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Long Short-Term Memory Networks [Електронний ресурс] // *ScienceDirect*. – 2021. – Режим доступу: https://www.sciencedirect.com/topics/computer-science/long-short-term-memory-networks (дата звернення: 27.04.2025).

2. Афанасьєва І. В., Скримінський Н. О. Дослідження методів нейронних мереж для виявлення шахрайства // *Поліграфічні, мультимедійні та web-технології*: тези доп. IX Міжнар. наук.-техн. конф. – 2024. – Т. 1. – С. 132–133.

3. A Gentle Introduction to the Bag-of-Words Model [Електронний ресурс] // *MachineLearningMastery*. – 2019. – Режим доступу: https://machinelearningmastery.com/gentle-introduction-bag-words-model/ (дата звернення: 30.04.2025).

4. Назаренко Д. С., Афанасьєва І. В., Голян Н. В. Нейромережевий підхід для емоційного розпізнавання тексту // *Біоніка інтелекту*. – 2019. – Т. 1, № 92. – С. 9–14.

5. BERT language model [Електронний ресурс] // *TechTarget*. – 2022. – Режим доступу: https://www.techtarget.com/searchenterpriseai/definition/BERT-language-model (дата звернення: 01.05.2025).

6. What is GPT? Everything you need to know [Електронний ресурс] // *Zapier*. – 2024. – Режим доступу: https://zapier.com/blog/what-is-gpt/ (дата звернення: 05.05.2025).

7. Назаренко Д. С., Афанасьєва І. В., Голян Н. В. Investigation of the Deep Learning Approaches to Classify Emotions in Texts // *CEUR Workshop Proceedings*. – 2021. – Vol. 2870. – С. 206–224.

8. Introduction to Recurrent Neural Networks [Електронний ресурс] // *GeeksforGeeks*. – 2024. – Режим доступу: https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/ (дата звернення: 07.05.2025).

9. Transformer Neural Networks: A Step-by-Step Breakdown [Електронний ресурс] // *Built In*. – 2024. – Режим доступу: https://builtin.com/artificial-intelligence/transformer-neural-network (дата звернення: 09.05.2025).

10. What Are Stemming and Lemmatization? [Електронний ресурс] // *IBM*. – 2023. – Режим доступу: https://www.ibm.com/think/topics/stemming-lemmatization (дата звернення: 20.05.2025).

11. Jamhuri M. Understanding the Adam Optimization Algorithm: A Deep Dive into the Formulas [Електронний ресурс] // *Medium*. – 2023. – Режим доступу: https://jamhuri.medium.com/understanding-the-adam-optimization-algorithm-a-deep-dive-into-the-formulas-3ac5fc5b7cd3 (дата звернення: 23.05.2025).

12. Misra R. News Category Dataset [Електронний ресурс] // *Kaggle*. – Режим доступу: https://www.kaggle.com/datasets/rmisra/news-category-dataset (дата звернення: 24.05.2025).

13. HuffPost [Електронний ресурс] // *HuffPost*. – Режим доступу: https://www.huffpost.com/ (дата звернення: 24.05.2025).

14. Kashnikov Y. 2025_M_PI_IPZm-23-2_Kashnikov_Y_K [Електронний ресурс] // *GitHub*. – Режим доступу: https://github.com/dybasser/2025_M_PI_IPZm-23-2_Kashnikov_Y_K (дата звернення: 05.06.2025).

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ ЗА НАУКОВИМИ НАПРЯМАМИ КЕРІВНИКА ТА НАУКОВЦІВ КАФЕДРИ ПРОГРАМНОЇ ІНЖЕНЕРІЇ

2. Афанасьєва І. В., Скримінський Н. О. Дослідження методів нейронних мереж для виявлення шахрайства // *Поліграфічні, мультимедійні та web-технології*: тези доп. IX Міжнар. наук.-техн. конф. – 2024. – Т. 1. – С. 132–133.

4. Назаренко Д. С., Афанасьєва І. В., Голян Н. В. Нейромережевий підхід для емоційного розпізнавання тексту // *Біоніка інтелекту*. – 2019. – Т. 1, № 92. – С. 9–14.

7. Назаренко Д. С., Афанасьєва І. В., Голян Н. В. Investigation of the Deep Learning Approaches to Classify Emotions in Texts // *CEUR Workshop Proceedings*. – 2021. – Vol. 2870. – С. 206–224.