

# ДОДАТОК А

## Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ

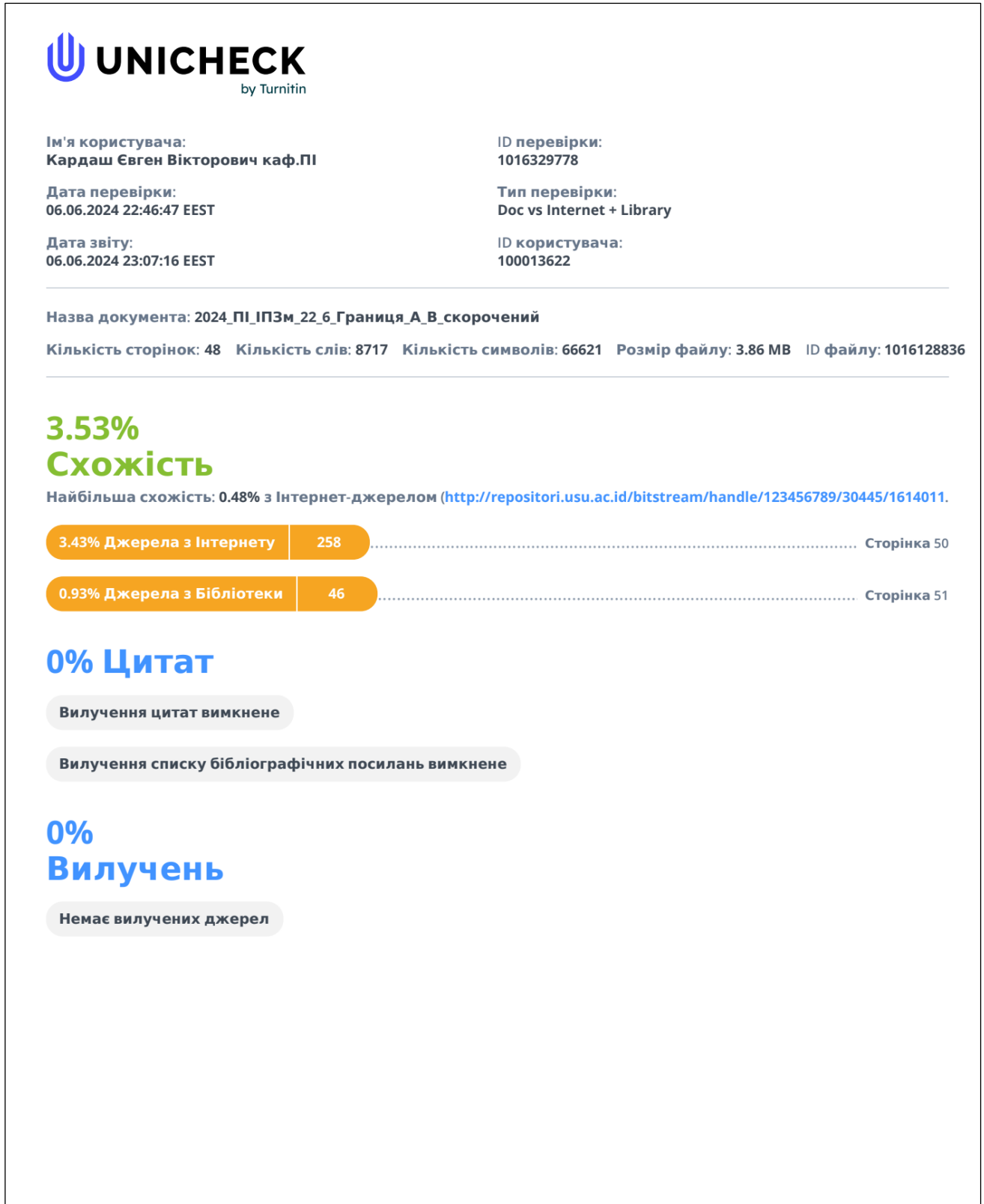


Рисунок А.1 – Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ  
(рисунок виконаний самостійно)

## ДОДАТОК Б

### Слайди презентації



ХАРКІВСЬКИЙ  
НАЦІОНАЛЬНИЙ  
УНІВЕРСИТЕТ  
РАДІОЕЛЕКТРОНИКИ

# Дослідження методів штучного інтелекту для прогнозування динаміки акцій компанії за результатами аналізу корпоративних звітів

Виконав:  
ст. гр. ПЗМ-22-6 Границя А.В.

Науковий керівник:  
доц. кафедри ПІ Кравець Н.С.



Рисунок Б.1 – Слайд 1 (рисунок виконаний самостійно)

## Актуальність Дослідження

- **Технологічний прогрес:** Швидкий розвиток технологій, особливо в області штучного інтелекту, створює нові можливості для аналізу та прогнозування фінансових ринків.
- **Зростаюча складність фінансових ринків:** Сучасні ринки стають все більш складними, що вимагає використання передових технологій для точного прогнозування.
- **Великий обсяг даних:** Обсяги корпоративних звітів постійно зростають, потребуючи автоматизації процесу аналізу для ефективного та швидкого прийняття рішень.



Рисунок Б.2 – Слайд 2 (рисунок виконаний самостійно)

# Аналіз Предметної Області

- **Штучний інтелект:** Включає методи та алгоритми, що дозволяють машинам виконувати завдання, які зазвичай потребують людського інтелекту, такі як обробка природної мови та аналіз великих даних.
- **Фінансові ринки:** Комплексні системи, де відбувається купівля-продаж акцій та інших фінансових інструментів. Важливо розуміти їх динаміку для прийняття обґрунтованих інвестиційних рішень.
- **Аналіз текстових даних:** Корпоративні звіти містять важливу інформацію про фінансовий стан та перспективи компаній. Ефективний аналіз цих текстів може значно підвищити точність прогнозів динаміки акцій.



Рисунок Б.3 – Слайд 3 (рисунок виконаний самостійно)

3

# Постановка задачі

- **Глибокий аналіз методів:** Оцінка сучасних підходів ШІ для прогнозування фінансових ринків, вивчення їхніх переваг та недоліків.
- **Збір та підготовка даних:** Збір великого обсягу корпоративних звітів, їх очищення, нормалізація та перетворення для навчання моделей.
- **Навчання та тестування моделей:** Використання історичних даних для тренування моделей, застосування методів крос-валідації для оцінки якості моделей.
- **Інтерпретація результатів:** Аналіз результатів прогнозування для визначення найбільш ефективних підходів та розуміння факторів, що впливають на точність прогнозів.
- **Подальший розвиток:** Оцінка впливу різних архітектур нейронних мереж на точність прогнозів, визначення напрямків для майбутніх досліджень та вдосконалення моделей.



Рисунок Б.4 – Слайд 4 (рисунок виконаний самостійно)

4

# Аналіз Моделей Машинного Навчання (Рекурентні Нейронні Мережі)

- **Рекурентні Нейронні Мережі**
  - Аналіз послідовностей даних.
  - Зберігають інформацію про попередні етапи.
  - Ідеальні для аналізу текстових звітів з урахуванням контекстуальної інформації.
- **LSTM (Long Short-Term Memory)**
  - Ефективно вирішує проблему зникнення градієнту.
  - Використовує спеціалізовані "ворота" для контролю потоку інформації.
  - Здатна моделювати як довгострокові, так і короткострокові залежності в даних.



Рисунок Б.5 – Слайд 5 (рисунок виконаний самостійно)

5

# Аналіз Моделей Машинного Навчання (NLP Моделі)

- **BERT**
  - Аналізує контекст слова в обох напрямках.
  - Висока ефективність для обробки великих текстових даних.
  - Використовує механізм трансформерів для моделювання залежностей між словами.
- **BigBird**
  - Розширена версія BERT для обробки довгих текстів.
  - Використовує модифіковану архітектуру трансформера для ефективної роботи з довгими послідовностями.
  - Підходить для аналізу великих документів, таких як корпоративні звіти.
- **FinBERT**
  - Спеціалізована для фінансових текстів.
  - Навчена на великому корпусі фінансових даних.
  - Демонструє високу точність у задачах, пов'язаних з аналізом настроїв та класифікацією фінансових текстів.



Рисунок Б.6 – Слайд 6 (рисунок виконаний самостійно)

6

# Аналіз Моделей Машинного Навчання (Гібридні Моделі)

- **Гібридні Моделі**
  - Комбінації кількох алгоритмів для покращення точності та ефективності;
  - Використовуються для складних задач аналізу даних, де один алгоритм може бути недостатнім.
- **BERT + LSTM: Як це працює та переваги**
  - **Контекстуалізація з BERT:** Генерує контекстуалізовані векторні уявлення тексту, що забезпечує детальне розуміння контексту.
  - **Сегментація тексту:** Довгі тексти розбиваються на сегменти до 512 токенів для ефективної обробки.
  - **Аналіз послідовностей з LSTM:** Векторні уявлення обробляються LSTM, що враховує довготривалі залежності між словами.
  - **Висока точність:** Поєднання BERT та LSTM забезпечує високу точність прогнозування завдяки детальному аналізу контексту та послідовностей.
  - **Стабільність та універсальність:** Модель адаптується до різних типів текстових даних, забезпечуючи надійні та стабільні результати прогнозування.



Рисунок Б.7 – Слайд 7 (рисунок виконаний самостійно)

7

## Збір даних

- **Джерела даних:** Використання корпоративних звітів форми 10-Q з американських компаній, що подаються до Комісії з цінних паперів та бірж США (SEC).
- **Період дослідження:** Звіти зібрані за період з 1994 по 2024 рік, що забезпечує достатньо даних для тренування та тестування моделей.
- **Сервіси для збору даних:** SEC-API.io – для отримання текстів звітів, Yahoo Finance – для отримання даних про ціни акцій.



Рисунок Б.8 – Слайд 8 (рисунок виконаний самостійно)

8

# Використані інструменти

- **Програмна мова та середовище:** Python, середовище розробки PyCharm.
- **Бібліотеки для машинного навчання:** TensorFlow, PyTorch, transformers.
- **Бібліотеки для обробки даних:** pandas, numpy, matplotlib, seaborn.
- **Інструменти для збору даних:** yfinance, sec\_api.



Рисунок Б.9 – Слайд 9 (рисунок виконаний самостійно)

9

## Основні Метрики Для Оцінки Ефективності Моделей

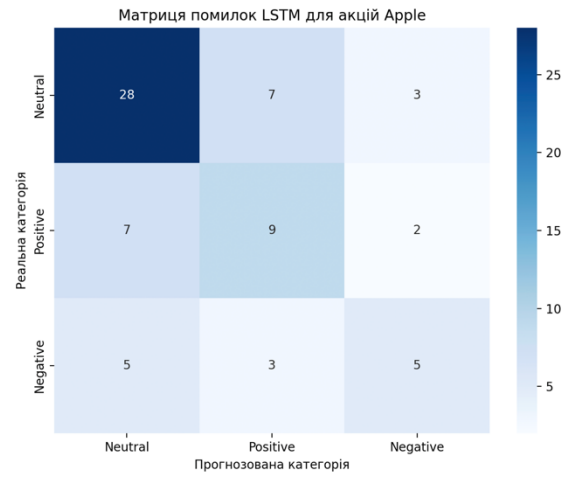
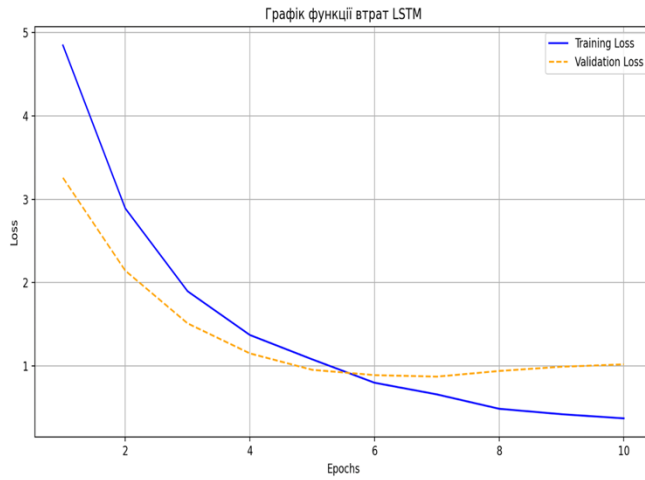
- **Accuracy:** Визначає частку правильних передбачень серед усіх зроблених передбачень. Корисний показник, коли кількість позитивних і негативних класів приблизно однакова.
- **Precision:** Вимірює частку правильно передбачених позитивних випадків серед усіх передбачених позитивних випадків. Важливий для уникнення помилкових спрацьовувань.
- **Recall:** Визначає здатність моделі виявляти всі релевантні позитивні випадки. Вимірює частку правильно ідентифікованих позитивних випадків серед усіх фактичних позитивних випадків.
- **F1-Score:** Комбінований показник, що поєднує точність передбачень та повноту. Забезпечує збалансовану оцінку ефективності моделі, особливо при дисбалансі між класами.
- **Час Прогнозування:** Вимірює, скільки часу потрібно моделі для здійснення передбачення. Важливий для оцінки продуктивності та придатності моделей для застосувань у реальному часі.
- **Час Навчання Моделі:** Вимірює, скільки часу потрібно для навчання моделі. Важливий для оцінки загальної ефективності моделі та ресурсів, необхідних для її навчання.



Рисунок Б.10 – Слайд 10 (рисунок виконаний самостійно)

10

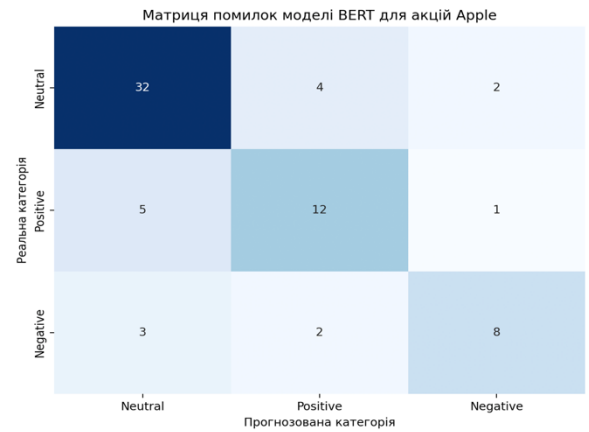
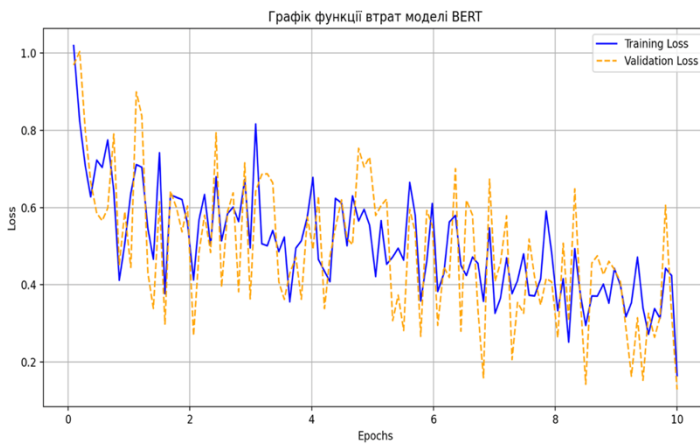
# Результати навчання LSTM



11

Рисунок Б.11 – Слайд 11 (рисунок виконаний самостійно)

# Результати навчання BERT



12

Рисунок Б.12 – Слайд 12 (рисунок виконаний самостійно)

# Результати навчання FinBERT

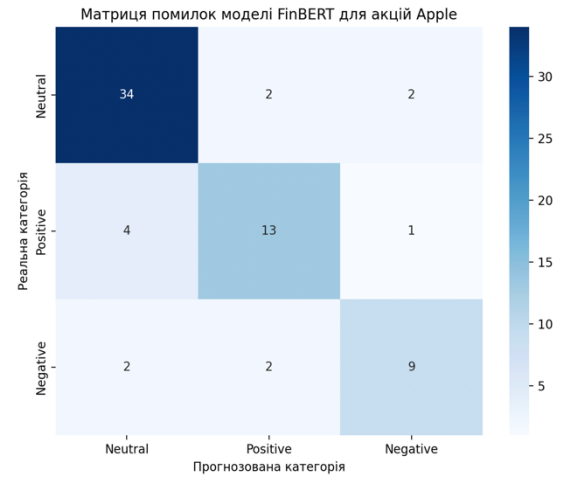
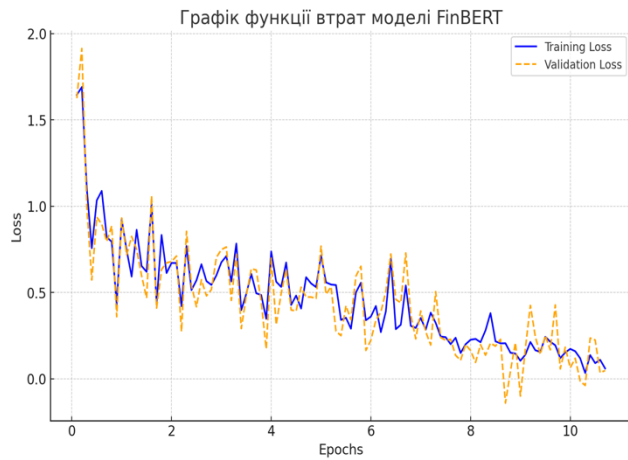


Рисунок Б.13 – Слайд 13 (рисунок виконаний самостійно)

13

# Результати навчання BigBird

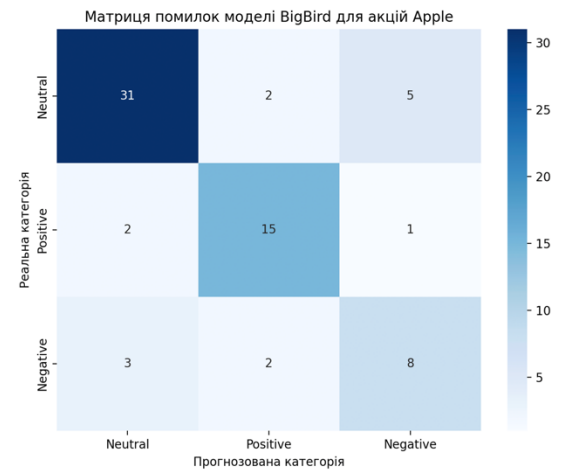
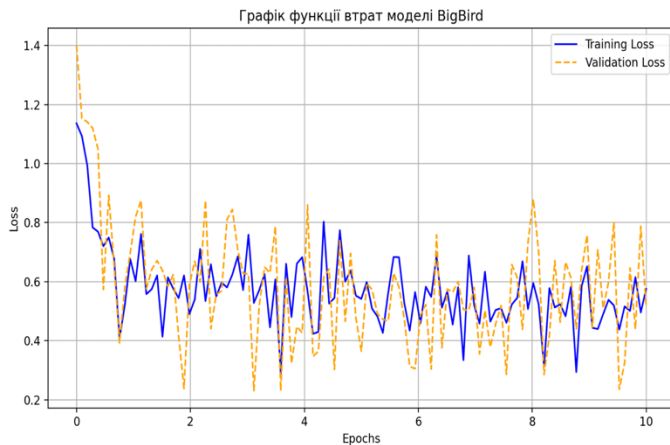


Рисунок Б.14 – Слайд 14 (рисунок виконаний самостійно)

14

# Результати навчання BERT + LSTM

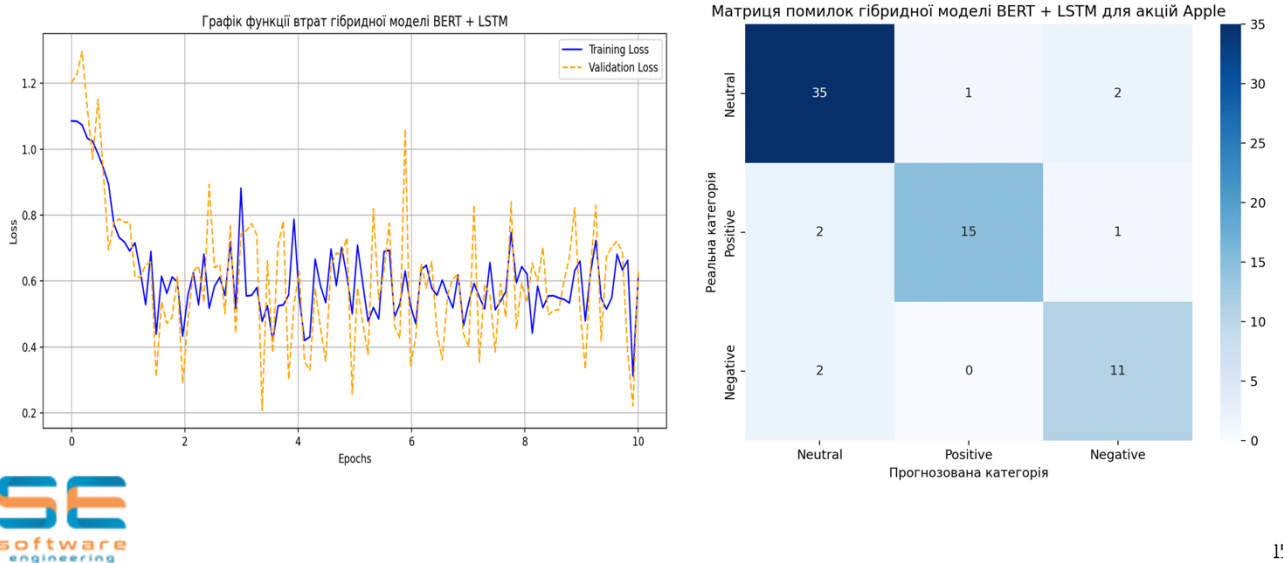


Рисунок Б.15 – Слайд 15 (рисунок виконаний самостійно)

## Отримані результати дослідження

Модель	Час навчання (секунди)	Час прогнозування (секунди)	Accuracy	Precision	Recall	F1-Score
LSTM	406.42	1.09	0.779	0.778	0.678	0.728
BERT	5217.46	2.23	0.81	0.803	0.731	0.762
FinBERT	6108.89	2.37	0.84	0.814	0.81	0.794
BigBird	34317.23	4.37	0.832	0.798	0.81	0.782
BERT + LSTM	5437.23	20.14	0.851	0.84	0.828	0.839

Рисунок Б.16 – Слайд 16 (рисунок виконаний самостійно)

# Підсумки

- **Ефективність Моделей:**

- Гібридна модель BERT + LSTM продемонструвала найвищу точність прогнозування.
- FinBERT, спеціально розроблена для обробки фінансових текстів, показала найвищу точність серед усіх NLP моделей.
- BERT та BigBird продемонстрували значні результати у обробці великих обсягів тексту.
- LSTM показала найнижчі результати.

- **Практична Цінність:**

- Моделі можуть бути використані для прийняття обґрунтованих інвестиційних рішень.
- Застосування для автоматизації фінансової аналітики та зниження впливу людського чинника.
- Використання в наукових дослідженнях для розробки нових методів прогнозування.

- **Перспективи Подальших Досліджень:**

- Розробка нових гібридних архітектур для покращення точності прогнозів.
- Інтеграція додаткових джерел даних для підвищення ефективності моделей.
- Оптимізація моделей для підвищення ефективності та зменшення часу навчання і прогнозування.



Рисунок Б.17 – Слайд 17 (рисунок виконаний самостійно)

## ДОДАТОК В

### Апробація результатів роботи

УДК 004.8

DOI: <https://doi.org/10.30837/IYF.IIS.2024.440>

#### **ОЦІНКА ЕФЕКТИВНОСТІ ВИКОРИСТАННЯ МЕТОДІВ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ АНАЛІЗУ ФІНАНСОВИХ ЗВІТІВ ТА ПРОГНОЗУВАННЯ ДИНАМІКИ АКЦІЙ**

Границя А. В.

Науковий керівник – к.т.н., доц. Кравець Н. С.

Харківський національний університет радіоелектроніки, каф. ПІ

м. Харків, Україна

e-mail: [andrii.hranysia@nure.ua](mailto:andrii.hranysia@nure.ua)

This research investigates artificial intelligence techniques for forecasting stock market trends through corporate report analysis. It emphasizes the utilization of neural networks and machine learning algorithms, particularly focusing on natural language processing (NLP) and recurrent neural networks, for sentiment analysis and trend prediction. The study aims to enhance investment strategies by improving the accuracy and speed of market change forecasts, demonstrating AI's significant potential in financial analysis and decision-making processes.

Світовий досвід використання технологій штучного інтелекту для аналізу фінансових ринків демонструє значний потенціал таких методів у підвищенні якості та швидкості прогнозування ринкових змін. Текстовий аналіз корпоративних звітів з використанням нейронних мереж та алгоритмів машинного навчання відкриває нові можливості для інвесторів та аналітиків.

Автоматичне визначення настрою у текстах корпоративних звітів, підрахунок ключових показників, а також класифікація інформації на основі її потенційного впливу на ринкові ціни акцій, може дозволити швидко відслідковувати потенційні тенденції зростання або зниження вартості цінних паперів.

Особлива увага приділяється методам обробки природної мови (NLP) та рекурентним нейронним мережам, які дозволяють виявляти закономірності у текстових даних і прогнозувати ринкові зміни. NLP моделі оснащені здатністю розуміння та інтерпретації людської мови в її природній формі, що дозволяє виявляти суттєві для ринку сигнали [1]. З іншого боку, рекурентні нейронні мережі спеціалізуються на роботі з послідовними даними, ідеально підходячи для аналізу текстів, де контекст і послідовність слів відіграють вирішальну роль. Завдяки своїй структурі, рекурентні мережі ефективно зберігають інформацію про попередній контекст, що дозволяє більш точно аналізувати зміст тексту та передбачати можливі тенденції та зміни на ринку [2]. Такий підхід забезпечує комплексний погляд на дані, враховуючи як лінгвістичні так і контекстуальні особливості тексту, що значно підвищує точність прогнозів та ефективність аналітичних моделей в області фінансових ринків.

Рисунок В.1 – Скан тез доповіді з наукової конференції «Інформаційні інтелектуальні системи» (рисунок виконаний самостійно)

В рамках дослідження, особливу увагу приділено використанню моделі BERT (Bidirectional Encoder Representations from Transformers), яка є передовою технологією у сфері обробки природної мови. BERT використовує техніку трансформерів, забезпечуючи глибше розуміння контексту в тексті шляхом аналізу слова в усіх можливих комбінаціях його використання перед та після нього [3]. Ця здатність аналізувати текст у двох напрямках дозволяє BERT точніше інтерпретувати сенс слова в контексті, що значно підвищує якість розпізнавання, класифікації та генерації тексту. У дослідженні, BERT застосовується для глибокого аналізу текстів корпоративних звітів з метою визначення настрою, ключових фактів та тенденцій, які можуть вплинути на ринкову динаміку акцій компаній. Завдяки своїй високій ефективності в розпізнаванні нюансів мови, BERT дозволяє виявляти складні закономірності у великих обсягах текстової інформації, тим самим забезпечуючи більш точне та оперативне прогнозування змін на фінансових ринках. Ця модель стає ключовим інструментом у розробці передових інвестиційних стратегій, дозволяючи аналітикам швидко оцінювати потенціал акцій на основі аналізу корпоративних звітів. З метою оптимізації процесу прогнозування динаміки акцій, необхідно провести порівняльний аналіз ефективності різних моделей машинного навчання у рамках дослідження. Крім того, буде досліджено ефективність комбінування NLP моделей з RNN, що має потенціал для покращення точності і глибини аналізу текстових даних. Такий підхід спрямований на забезпечення більш точного прогнозування шляхом виявлення складних закономірностей у великих обсягах неструктурованих даних.

На основі детального аналізу сучасних технологій штучного інтелекту та їх застосування в аналізі текстових даних, були встановлені конкретні критерії для оцінки доцільності використання ШІ, а саме, ефективність виявлення ключових трендів та аналіз настрою. Отже, на основі аналізу даних та оцінки доступних методів обробки природної мови, було виявлено підходи, що дозволяють покращити точність прогнозування динаміки фінансових показників. Також практичне значення роботи полягає у впровадженні розроблених методик у стратегії інвестиційних компаній для оптимізації процесів прийняття рішень.

Список використаних джерел:

1. Smelyakov K., Karachevtsev D., Kulemza D., Samoilenko Y., Patlan O., Chupryna A. (2020). Effectiveness of Preprocessing Algorithms for Natural Language Processing Applications. URL: <https://doi.org/10.1109/PICST51311.2020.9467919> (дата звернення: 08.03.2024).
2. Cho K., van Merriënboer B., Bahdanau D., Bengio Y (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. URL: <https://arxiv.org/abs/1406.1078> (дата звернення: 05.03.2024).
3. BERT: Pre-training of Deep Bidirectional Transformers. URL: <https://arxiv.org/abs/1810.04805> (дата звернення: 03.03.2024).

Рисунок В.2 – Скан тез доповіді з наукової конференції «Інформаційні інтелектуальні системи» (рисунок виконаний самостійно)

## ДОДАТОК Г

Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ 3008: 2015

Експертний висновок результатів перевірки кваліфікаційної роботи

студент  
(посада)

програмної інженерії  
(кафедра)

ІІЗМ-22-6  
(група)

Границя Андрій Володимирович

(прізвище, ім'я, по батькові)

Зауваження

Пункт ДСТУ 3008-2015	Зміст пункту	Сторінка кваліфікаційної роботи
1	2	3
	<b>7.1 Загальні положення</b>	
	<b>7.3 Нумерація сторінок звіту</b>	
	<b>7.4 Нумерація розділів, підрозділів, пунктів, підпунктів</b>	
	<b>7.5 Рисунки</b>	
	<b>7.6 Таблиці</b>	
	<b>7.7 Переліки</b>	
	<b>7.8 Примітки</b>	
	<b>7.9 Виноски</b>	
	<b>7.10 Формули та рівняння</b>	
	<b>7.11 Посилання</b>	
	<b>7.13 Список авторів</b>	
	<b>7.14 Скорочення та умовні позначки</b>	
	<b>7.15 Додатки</b>	

Експерт зауважень немає  
\_\_\_\_\_ (підпис)

Олена ОЛІЙНИК  
(прізвище, ініціали)

09.06.2024

Рисунок Г.1 – Експертний висновок результатів перевірки кваліфікаційної роботи  
(рисунок виконаний самостійно)