

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**  
Харківський національний університет імені В.Н.Каразіна  
Факультет математики і інформатики  
Кафедра теоретичної та прикладної інформатики

## **Кваліфікаційна робота магістр**

на тему «Методи оцінювання та прогнозування в дослідженні процесів на  
фондовому ринку»

Виконав: студент 2 курсу, групи МФ-61  
спеціальність 122 «Комп'ютерні науки»  
освітньо-наукова програма «Інформатика»

Савченко Р.О.

(прізвище та ініціали)

Керівник Парфенюк Ю.А.

(прізвище та ініціали)

Рецензент

(прізвище та ініціали)

Харків – 2024 року

## ЗМІСТ

1. ВСТУП.....	3
1.1. Формулювання мети роботи, задач та обґрунтування актуальності теми.....	3
1.2. Стислий огляд відомих результатів в області дослідження.....	5
1.3. Відомості про одержані результати та їх новизна.....	6
2. ОСНОВНА ЧАСТИНА.....	8
2.1. Постановка задачі.....	8
2.2. Розвинутий огляд сучасного стану справ в області дослідження.....	10
2.3. Методи дослідження.....	16
2.4. Опис та обґрунтування методів оцінки результатів дослідження.....	25
2.5. Результати прогнозування та їх аналіз.....	33
3. ВИСНОВКИ.....	48
4. СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	49
5. ДОДАТКИ.....	51

## 1. ВСТУП

### 1.1. Формулювання мети роботи, задач та обґрунтування актуальності теми

Мета роботи: Основною метою даної магістерської роботи є розробка, валідація і перевірка ефективності моделей ARIMA, RNN і LSTM для прогнозування рухів трьох міжнародних фондових індексів: BUMIX (Budapest Stock Index), NYSE (New York Stock Exchange Index) і Blue-Chip SBITOR (Slovenia Stock Index). У подальшому, зібуті результати планується застосовувати для прогнозування українських фондових індексів. Зараз, враховуючи складність прогнозування українських індексів через воєнні дії, які вносять значну невизначеність і волатильність, вибір пропонується зосередити на міжнародних індексах, що дозволить застосувати методи передбачення на стабільніших та передбачуваних ринкових даних.

Моделі будуть побудовані на історичних даних за період з 2014 по 2022 рік та перевірені на даних з 2022 по 2024 роки. Оцінка результатів прогнозування виконується за допомогою наступних метрик:

- MAE (Середня абсолютна помилка)
- MSE (Середньоквадратична помилка)
- RMSE (Корінь середньоквадратичної помилки)
- MAPE (Середня абсолютна відсоткова помилка)
- SMAPE (Симетрична середня абсолютна відсоткова помилка)

Цей підхід забезпечує глибокий аналіз ефективності прогнозних моделей на основі реальних історичних даних, що дозволяє виявити їхню придатність і потенціал для адаптації до умов українського фондового ринку.

З метою досягнення поставленої мети в роботі визначено наступні задачі:

Теоретичний огляд існуючих методів: Проаналізувати літературу і вже існуючі підходи, що використовують ARIMA, RNN і LSTM моделі для прогнозування фондових індексів.

Збір та підготовка даних: Зібрати історичні дані для міжнародних фондових індексів BUMIX, NYSE, та SBIPOP за період з 2014 по 2024 рік. Підготувати ці дані для аналізу.

Розробка моделей ARIMA, RNN і LSTM: Сконструювати та налаштувати моделі ARIMA, RNN, і LSTM, використовуючи дані за період з 2014 по 2022 роки. Оцінити параметри моделей для оптимальної точності прогнозування.

Тестування моделей на нових даних: Перевірити моделі на актуальних даних з 2022 по 2024 роки для оцінки їх здатності до прогнозування в динамічних умовах.

Оцінка ефективності моделей: Використати метрики MAE, MSE, RMSE, MAPE, та SMAPE для оцінки точності та надійності прогнозів, які генерують моделі. Аналізувати результати для визначення найефективніших підходів.

Аналіз потенціалу застосування моделей для українського ринку: На основі отриманих результатів оцінити можливість адаптації та використання моделей для прогнозування рухів українських фондових індексів у майбутньому.

Ці задачі забезпечують комплексний підхід до аналізу і прогнозування фондових індексів, покладаючи основу для майбутнього використання отриманих моделей у контексті українського ринку.

Обрана тема є надзвичайно актуальною через кілька причин:

1) волатильність фондових ринків – значні коливання фондових ринків можуть призводити до великих фінансових втрат для інвесторів. Розробка точних і надійних методів прогнозування допомагає зменшити ризики, забезпечуючи більш обґрунтовані інвестиційні рішення;

2) покращення інвестиційних стратегій – адекватні та ефективні методи прогнозування сприяють розробці стратегій управління портфелем, які дозволяють максимізувати дохід і мінімізувати ризики;

3) технологічний прогрес – новітні досягнення в областях машинного навчання та штучного інтелекту відкривають нові можливості для покращення інструментів аналізу та прогнозування на фондових ринках. Інтеграція цих технологій може суттєво підвищити точність і надійність прогнозів;

4) глобалізація економічних процесів – у контексті глобалізації зростає важливість розуміння і аналізу міжнародних ринків. Дослідження, яке охоплює міжнародні індекси, дозволяє краще зрозуміти глобальні тенденції та їх вплив на внутрішній ринок;

5) специфічні умови українського ринку – враховуючи складність і волатильність українського фондового ринку, особливо в умовах політичної та економічної невизначеності, таке дослідження має важливе практичне значення. Адаптація досвіду з міжнародних ринків може допомогти в розробці ефективних методів оцінки та прогнозування для України.

Ці аспекти підкреслюють важливість та своєчасність обраної теми, яка сприяє не лише науковому розвитку в галузі економіки та фінансів, але й має високий потенціал для практичного застосування в управлінні інвестиційними ресурсами.

## 1.2. Стислий огляд відомих результатів в області дослідження

Дослідження процесів на фондовому ринку використовує різноманітні статистичні та обчислювальні методи для аналізу та прогнозування. Основні методи, які застосовуються в цьому напрямку, включають традиційні статистичні моделі та сучасні підходи машинного навчання:

ARIMA (Авторегресійна інтегрована модель ковзаючого середнього): Це один з найпоширеніших статистичних методів для аналізу часових рядів. Модель ARIMA використовується для аналізу та прогнозування даних, що показують тренди, цикли або сезонні коливання. Її ефективність була підтверджена численними дослідженнями, особливо для короткострокових прогнозів на ринках з високою ліквідністю.

Нейронні мережі (RNN і LSTM): Нейронні мережі, особливо рекурентні (RNN) та довгострокові короткочасні мережі (LSTM), є ефективними у моделюванні часових рядів завдяки їх здатності запам'ятовувати інформацію на протязі довгих періодів часу. Вони показали високу ефективність у

прогнозуванні фінансових ринків, де дані часто мають нестационарні властивості та складні шаблони залежностей.

Гібридні моделі: Сполучення статистичних та машинно-навчальних моделей в одну гібридну систему може підвищити точність прогнозування. Наприклад, інтеграція ARIMA для виявлення трендів і сезонності з LSTM для моделювання складних залежностей між точками даних забезпечує більш точні прогнози.

Різні дослідження демонструють успішне застосування цих моделей на різних географічних ринках і фондових індексах, що підкреслює універсальність та адаптивність методів.

Важливою частиною досліджень є використання метрик як MAE, MSE, RMSE, MAPE, і SMAPE для оцінки точності прогнозів, що дозволяє визначити найбільш ефективні підходи.

Наведене вище не лише підкреслює важливість і релевантність теми, але й забезпечують теоретичну основу для розвитку нових досліджень у цій області, зокрема застосування новітніх технологічних підходів для покращення методів оцінки та прогнозування на фондовому ринку.

### 1.3. Відомості про одержані результати та їх новизна

В рамках цього дослідження було досягнуто наступних результатів, які вносять новизну в аналіз фондових ринків:

Незалежне застосування моделей ARIMA, RNN і LSTM. Кожна з моделей була використана окремо для аналізу і прогнозування фондових індексів, що дозволило оцінити та порівняти їх ефективність в ізольованому середовищі. Це забезпечило чітке розуміння сильних і слабких сторін кожної моделі у контексті фондового ринку.

Моделі були побудовані та перевірені виключно на історичних даних за період з 2014 по 2022 рік. Це підкреслює здатність моделей виявляти і використовувати тренди та закономірності з минулого для прогнозування майбутніх ринкових рухів.

Прогнози, отримані за допомогою кожної моделі, були оцінені за допомогою таких метрик як MAE, MSE, RMSE, MAPE, та SMAPE. Це дало можливість кількісно аналізувати та порівнювати ефективність кожної моделі в залежності від їхньої специфіки та типу даних.

Результати показали, як кожна модель адаптується до особливостей різних міжнародних фондових індексів, надаючи інсайти щодо їхньої потенційної ефективності на інших ринках, включаючи український.

Отримані дані і висновки відкривають шлях для дальших досліджень з використанням цих моделей в умовах високої волатильності, особливо з огляду на можливість їх застосування на українському фондовому ринку.

Ці результати розширюють розуміння потенціалу окремого застосування ARIMA, RNN та LSTM моделей для аналізу фондових ринків, зокрема у межах міжнародного контексту, та надають підґрунтя для майбутнього використання цих методів для більш точного та адаптивного прогнозування.

## 2. ОСНОВНА ЧАСТИНА

### 2.1. Постановка задачі

Основна мета практики полягає у застосуванні існуючих аналітичних інструментів для прогнозування динаміки фондових індексів. Результати практики можуть бути використані, у тому числі, для вдосконалення та оптимізації інвестиційного процесу.

Для отримання результату були поставлені наступні задачі:

1. Огляд методів прогнозування що можуть бути використані для здійснення прогнозів на фондовому ринку.
2. Написання коду для здійснення прогнозування значень фондових індексів за обраними методиками.
3. Проведення обчислювальних експериментів із прогнозування значень фондових індексів, що корелюють з індексами України.
4. Розробка моделей для прогнозування значень фондових індексів України.
5. Аналіз результатів отриманих прогнозів та порівняння результатів отриманих за різними методиками.

Предметом практики були фондові індекси, які представляють собою агреговані показники цін акцій певної групи компаній.

Тема оцінки та прогнозування значень фондового ринку завжди залишається актуальною і має значний практичний та науковий інтерес з цілого ряду причин.

По-перше, фондові ринки відіграють ключову роль у світовій економіці, мобілізуючи капітал для компаній та надаючи інвесторам можливість заробітку на капіталі. Ефективне та точне прогнозування ринкових трендів може значно підвищити прибутковість інвестицій та знизити ризики, що є критично важливим для інвестиційних фондів, приватних інвесторів, і навіть для державних політик.



По-друге, точність у прогнозуванні фондового ринку безпосередньо впливає на стратегію управління портфелями, включаючи розподіл активів, хеджування ризиків та таймінг входження чи виходу з ринку. Інвестори, здатні краще передбачити зміни ринку, можуть оптимізувати свої портфелі та підвищити загальний дохід.

По-третє, з розвитком машинного навчання, штучного інтелекту, великих даних та обчислювальних технологій можливості прогнозування фондових індексів стали значно ширшими. Актуальність теми також підвищується завдяки постійним інноваціям у методах збору та обробки інформації, що дозволяє аналітикам отримувати глибші та більш точні інсайти.

По-четверте, фондові ринки стають все більш взаємопов'язаними в результаті глобалізації, що збільшує комплексність викликів перед прогнозуванням. Це підкреслює необхідність більш сучасних та ефективних підходів до аналізу фінансових ринків.

У п'ятих, фондові ринки відомі своєю волатильністю і непередбачуваністю. Здатність точно оцінити й передбачити ці коливання є важливою як для зменшення потенційних втрат, так і для максимізації прибутку. Така здатність стає особливо цінною у часи фінансових криз, коли традиційні інвестиційні стратегії можуть не працювати.

І, на решті, тема прогнозування фондових ринків продовжує залучати увагу дослідників у сфері економіки, фінансів та комп'ютерних наук. Це поле дослідження надає можливості для розвитку нових теоретичних моделей та тестування існуючих теорій, а також спонукає до розвитку крос-дисциплінарних досліджень, що залучають інженерію даних, статистику, теорію ігор, і багато іншого.

Таким чином, оцінка та прогнозування фондового ринку не тільки має важливе практичне значення, але й є центральною темою в академічних дослідженнях, що зосереджені на розумінні та керуванні фінансовими ризиками та можливостями.

## 2.2. Розвинутий огляд сучасного стану справ в області дослідження

Вчені в Китаї та за кордоном активно займаються дослідженнями з використання машинного навчання для аналізу акційних ринків. В контексті застосування нейронної мережі зворотного поширення помилок (BPNN) для аналізу акцій, Ву В., Чен В. та Лю Б. [1] використали цей підхід для прогнозування коливань індексу Shanghai Composite, демонструючи високу ефективність цієї моделі на китайському фондовому ринку. Згодом, Тікнор Дж.Л. [2] застосував цей метод для прогнозування динаміки акцій Microsoft і Goldman Sachs, підтверджуючи його корисність. Окрім того, Чжан Д.Х. та Лу С. [3] досягли точності прогнозування акцій 73,29% за допомогою емпіричних тестів за допомогою BPNN. Тей Ф.Е.Х. і Цао, Л.Дж. [4] розглядали застосування опорної векторної регресії (SVR) для прогнозів фондового ринку, демонструючи її переваги. Ран Ю. та Цзян, Х. [5] використали BPNN і SVR для створення моделі прогнозування, яка показала менші помилки та вищу точність прогнозування за допомогою SVR. Кім, К.-Дж. [6] використовував SVM для класифікації щоденних змін на корейському фондовому ринку (KOSPI) і порівняв це з нейронними мережами (NN) та аналізом на основі аргументації (CBR), де SVM показали кращі результати. Проте поверхневі алгоритми машинного навчання зі своєю простою структурою часто стикаються з обмеженими можливостями обробки неструктурованих даних, а також з проблемами локальних оптимумів, перенавчання та повільної конвергенції в реальних умовах.

Щоб подолати виклики, пов'язані з машинним навчанням, дослідники звернулись до методів глибокого навчання для аналізу фондових ринків. Глибоке навчання, розроблене Хінтоном Г.Е. та Салахутдіновим Р.Р. [7], ефективно застосовувалось для моделювання даних часових рядів. В іншому дослідженні Сінгх Р. та Шривастава С. [8], методи машинного навчання та глибокого навчання були використані для аналізу акцій на основі даних NASDAQ від Google. Результати показали, що інтеграція двовимірного аналізу головних компонентів (PCA) з глибокими нейронними мережами (DNN) показала кращі

результати, ніж використання двонаправленого РСА з нейронною мережею з радіальним базисом (RBFNN), що демонструє ефективність певних методологій у підвищенні точності прогнозування.

Краус М. та Фейрригель С. [9] поєднали глибокі нейронні мережі (DNN), дерева з градієнтним бустингом і випадкові ліси для прогнозування майбутньої дохідності акцій індексу S&P 500 за обраний період. З іншого боку, Чонг Е., Хан К., і Парк Ф.К. [10] інтегрували DNN з трьома методами неконтрольованого виділення ознак для прогнозування високочастотних тенденцій на фондовому ринку: аналізом головних компонентів (PCA), автокодером та обмеженою машиною Больцмана. Також Кюи Д. [11] використав мережі глибоких переконань (DBN) для прогнозування майбутніх коливань цін на акції, демонструючи кращу продуктивність у порівнянні з BPNN та RBFNN. На аналогічний лад, Лю К. [12] інтегрував нечітку логіку з мережами глибоких переконань (DBN), розробляючи модель глибокого навчання на основі нечіткої теорії для аналізу акційних ринків. Ця модель показала обнадійливі результати та відкрила нові можливості для подальших досліджень, згідно з результатами експериментів. У свою чергу, Ли Х.М., Ян Л., Сюэ Ф.З. та Чжоу Х.Дж. [13] внесли модифікації до DBN, додавши внутрішню пластичність, що надала системі адаптивності. Завдяки цьому вдалося значно покращити точність прогнозування цін на закриття акцій. Цантекідіс А., Пассаліс Н., Тефас А., Канніайнен Дж., Габбуж М. та Йосіфідіс А. [14] застосували кодер для обробки послідовностей даних, а далі використали конволюційні нейронні мережі (CNN) для аналізу, демонструючи, що CNN більш ефективні для прогнозування динаміки акцій порівняно з іншими методами, такими як багат шарові перцептрони (MLP) і підтримувальні векторні машини (SVM). Сім Х.С., Кім Х.І., Ан Дж.Дж. [15] розробили модель прогнозування цін акцій на основі CNN для індексу S&P500 і порівняли її точність зі штучними нейронними мережами (ШНМ) та підтримувальними векторними машинами (SVR). Експериментальні дані підтвердили, що CNN є оптимальним варіантом для створення моделей прогнозування цін акцій. Також Чен В., Цзян М.Р., Чжан В.Г. та Чень З.С. [16]

запропонували модель прогнозування трендів на фондовому ринку на основі CNN, яка використовує конволюційні характеристики графіків. Їхній підхід продемонстрував перевагу, оскільки було використано шість випадково вибраних китайських акцій для підтвердження ефективності моделі. Також Лука Ді Персіо та Олександр Г. [17] застосували багатошаровий перцептрон (MLP) та згорточні нейронні мережі (CNN) для прогнозування відкриваючих та закриваючих цін індексу S&P500 на наступний день, виявивши, що CNN забезпечують менші помилки у прогнозах порівняно з MLP. Водночас Се, Т.Дж., Сяо, Г.Ф. та Йє У.К. [18] використали вейвлет-перетворення для фільтрації шуму в даних цін акцій, після чого застосували рекурентні нейронні мережі (RNN), оптимізовані алгоритмом штучної бджолоїної колонії, для прогнозування акційних цін в реальному часі. Ратер А.М., Агарвал А., Састрий В.Н. [19] розробили гібридну модель для прогнозування прибутковості акцій, комбінуючи авторегресійні моделі ковзного середнього, моделі експоненційного згладжування та RNN. Ця модель демонструвала вищу точність прогнозування порівняно з використанням окремої RNN. З іншого боку, Цинь Ю.С., Сонг Д.Дж., Ченг Х.Ф., Ченг В., Цзян Г.Ф. та Коттрелл, Г.В. [20] представили дворівневу RNN-модель з механізмом уваги, яка адаптивно визначає релевантні вхідні дані для прогнозування. Експериментальні результати підтвердили, що їхня модель більш ефективна для аналізу даних про акції порівняно з іншими підходами.

Для вирішення широко розповсюджених проблем зі зникненням або вибухом градієнтів, а також з довготривалими залежностями у нейронних мережах, було розроблено нейронну мережу довготривалої короткочасної пам'яті (LSTM), запропоновану Хохрейтером С. та Шмідгубером Дж. [21]. Ця мережа активно застосовується для прогнозування часових рядів завдяки своїй здатності ефективно обробляти довготривалі залежності, зберігаючи інформацію з минулих етапів навчання.

В порівнянні з традиційними RNN, LSTM показує кращі результати у вирішенні таких проблем. Лука Ді Персіо та Олександр Г. [17] провели порівняння ефективності RNN, LSTM та GRU у прогнозуванні цін на акції

Google, де LSTM демонструвала вищу точність у прогнозуванні. Ян К. та Ван К. [22] розширили свої дослідження до 30 світових фондових індексів і розробили модель LSTM для аналізу ефективності прогнозів на короткострокові, середньострокові та довгострокові періоди.

В результаті, LSTM показав вищу точність у порівнянні з економетричними методами, такими як ARIMA, а також з методами машинного навчання SVR і MLP, у всіх випробуваних індексах для різних періодів. Це підтверджує, що глибоке навчання забезпечує високу точність прогнозування у широкому діапазоні застосувань. Згідно з дослідженням Ли М.К. [23], переобладнання є значною проблемою в глибокому навчанні, особливо коли обсяг навчальних даних недостатній у порівнянні зі складністю моделі. У таких випадках модель глибокого навчання може перестати узагальнювати та почати запам'ятовувати конкретні дані, на яких вона навчалася, що веде до поганих результатів при роботі з новими, невідомими даними. Це особливо проблематично у контексті фондового ринку, де історичні дані, доступні для навчання, часто обмежені. Незважаючи на існування різних методів регуляризації, які допомагають зменшити переобладнання, проблема все ще залишається важливим викликом у галузі глибокого навчання. Сучасні дослідження активно зосереджені на розробці більш ефективних стратегій для вирішення цієї критичної проблеми.

Зусилля з прогнозування курсів акцій були зосереджені на використанні методів навчання під наглядом, включаючи нейронні мережі, випадкові ліси та методи регресії, як це видно з роботи Нті І.К., Адекоя А.Ф. та Вейорі Б.А. [24]. Детальний аналіз, проведений Сахаре Н.Н. та Імамбі, С.С. [25], вказав на важливість історичних даних для контрольованих моделей та виявив обмеження, які часто призводять до недостатньо точних прогнозів.

У комплексному дослідженні, проведеному Ху З., Чжао Ю. та Хуші М. [26], були розглянуті різні методи глибокого навчання, включаючи CNN, LSTM, DNN, RNN і RL, які були інтегровані з обробкою природної мови (NLP) та WaveNet. Для аналізу використовувалися дані з валютного ринку Forex, а

ефективність методів оцінювалася за допомогою таких показників, як середня абсолютна відсоткова помилка (MAPE), середньоквадратична помилка (RMSE), середня квадратична помилка (MSE) та коефіцієнт Шарпа.

Отримані результати вказують на зростаючу роль методів, таких як RL (Reinforcement Learning) і DNN (Deep Neural Networks), у дослідженнях прогнозування акцій, що підкреслює їхню зростаючу популярність у сфері фінансового моделювання. Це дослідження охоплювало широкий спектр методів прогнозування, але особливо звернуло увагу на відсутність даних про використання комбінацій різних методів глибокого навчання для аналізу акційних ринків. В дослідженнях, проведених Хіранша М., Гопалакрішнаном Е.А., Меноно В.К., Зоманом К.П., Пателем Р., Чоудхарі В., Саксеною Д. та Сінгхом А.К. [27, 28], були вивчені чотири моделі глибокого навчання – MLP (Multilayer Perceptron), RNN (Recurrent Neural Network), CNN (Convolutional Neural Network) та LSTM (Long Short-Term Memory), використовуючи дані з ринків NYSE (New York Stock Exchange) і NSE (National Stock Exchange).

При окремому аналізі кожна з моделей виявляла тенденції на фондових ринках, забезпечуючи унікальне розуміння взаємозв'язків між двома ринками. Особливо модель на базі CNN продемонструвала високу ефективність у прогнозуванні цін акцій окремих компаній. Однак це дослідження не розглядало можливості гібридних мереж, що залишає відкритим питання про потенціал комбінованих моделей у прогнозуванні акцій. З іншого боку, розвиток у сфері машинного навчання сприяв значним досягненням у розпізнаванні мови, обробці тексту та класифікації зображень, як це підтверджують дослідження Камата У., Лю Дж. та Вітакера Дж. [29]. Манолакис Д., Босовський Н. та Інгл В.К. застосували методи цифрової обробки сигналів до аналізу фондових даних, зосередившись на вивченні часових рядів [30]. Навчання з підкріпленням (RL) стало виділятися як метод, здатний долати обмеження традиційних підходів до навчання під наглядом. Каббані Т. та Думан Е. інтегрували прогнозування цін на фінансові активи з кроками розподілу, демонструючи, що алгоритми RL можуть ухвалювати оптимальні рішення у складних умовах фондового ринку [31]. Тим

часом, хоча методи LSTM були інтенсивно досліджені для прогнозування акцій завдяки їхній здатності ефективно обробляти великі обсяги даних, Могар А. та Хаміче, М. [32] вказали на виклики, пов'язані з необхідністю великої кількості історичних даних та значних обчислювальних ресурсів. Однією з критичних проблем моделей LSTM є їх обмеженість у наданні раціональних інвестиційних рішень, таких як купівля, продаж або утримання акцій, заснованих на прогнозах. Це підкреслили Рен Ю., Ляо Ф. та Гонг Ю. [33]. Однак, як показало дослідження Джина З., Яна Ю. та Лю Ю. [34], існує значний потенціал у поєднанні LSTM з аналізом настроїв, що може забезпечити цінну підтримку інвесторам у прийнятті рішень. Також було досліджено застосування методів опорних векторних машин (SVM) для аналізу часових рядів. Хоча моделі SVM відзначаються високою точністю, вони потребують великих обсягів даних і затратних процесів навчання, як зазначили Паррей І. Р., Хурана С.С., Кумар М. та Альтальбе А.А. [35].

### 2.3. Методи дослідження

Дослідження фондових індексів використовує передові статистичні та машинні методи для аналізу та прогнозування поведінки ринку. Ці методи еволюціонували, щоб задовольнити потреби в аналізі складних і великих даних, які характеризують сучасні фінансові ринки. Сьогодні у сфері фінансового аналізу і прогнозування фондових індексів існує ряд традиційних та сучасних методів, кожен з яких має свої переваги і недоліки. Далі наведено детальний огляд найпоширеніших із них.

Коли ми говоримо про традиційні методи прогнозування у фінансовому аналізі, перше, що приходить на думку це мовляв ARIMA (AutoRegressive Integrated Moving Average). Вона є однією з найвідоміших і широко використовуваних статистичних моделей для аналізу та прогнозування часових рядів. Ця модель ідеально підходить для аналізу даних, які показують тренди, циклічні коливання або сезонні варіації.

ARIMA є статистичною моделлю, яка аналізує і прогнозує часові ряди, базуючись на ідеї, що майбутні значення в часовому ряду можуть бути передбачені на основі їх минулих значень. Модель включає компоненти авторегресії (AR), інтеграції (I) та ковзного середнього (MA).

1. AR (AutoRegressive) – є основою багатьох моделей часових рядів, включаючи популярну модель ARIMA. Цей компонент використовується для опису залежностей між послідовними спостереженнями, що дозволяє прогнозувати майбутні значення на основі попередніх. Використовує параметр  $p$  (AR( $p$ )) що вказує, скільки попередніх спостережень використовується в моделі. Наприклад, якщо  $p = 1$ , поточне значення часового ряду прогнозується на основі одного попереднього значення.

Для ефективної роботи AR моделей, часовий ряд повинен бути стаціонарним, тобто його статистичні характеристики (середнє, дисперсія) не повинні змінюватись з часом. Якщо ряд не є стаціонарним, його можна перетворити, наприклад, за допомогою взяття перших різниць.



Важливим інструментом при аналізі AR моделей є корелограма, яка дозволяє визначити оптимальне значення  $p$  для моделі, спостерігаючи за затуханням автокореляцій.

AR моделі залишаються популярним інструментом завдяки їхній здатності забезпечувати чітке і теоретично обґрунтоване розуміння динаміки часових рядів у багатьох прикладних областях.

2. I (Integrated) – є ключовим компонентом для обробки нестационарних часових рядів, що мають тренди або сезонні варіації та вказує на необхідність «інтегрування» часового ряду, тобто диференціювання даних один або кілька разів для досягнення стаціонарності. Стаціонарність часового ряду означає, що його статистичні характеристики, такі як середнє і дисперсія, не змінюються з часом, що є важливим для ефективного застосування багатьох статистичних методів аналізу. Інтегрування – це процес усунення трендів та сезонності, що полягає у відніманні попереднього спостереження від поточного.

I(d) компонент ARIMA моделі допомагає стабілізувати середнє значення ряду шляхом видалення трендів і сезонності, що робить його більш передбачуваним і піддається аналізу. Диференціювання допомагає усунути з часових рядів структурні тренди та сезонність, що робить можливим використання AR і MA моделей для аналізу стаціонарних рядів. Це особливо важливо у фінансовій аналітиці, де тренди можуть змішувати прогнози і призводити до неправильних інвестиційних рішень.

Параметр  $d$  в I(d) визначає порядок різниць, необхідний для досягнення стаціонарності часового ряду. Це значення обирається на основі тестування стаціонарності ряду після кожного етапу диференціювання.

При побудові моделі ARIMA необхідно ретельно підібрати параметр  $d$ , враховуючи властивості конкретного часового ряду, щоб забезпечити найкращу можливу основу для аналізу та прогнозування. Це може включати ітеративні процеси тестування і моделювання з різними значеннями  $d$ , щоб знайти ідеальний баланс між видаленням трендів і збереженням інформативності даних.

3. MA (Moving Average) – є фундаментальною частиною багатьох моделей часових рядів і використовується для опису залежностей у часових рядах, де поточні значення залежать від попередніх шумових помилок.

Модель MA порядку  $q$ , зазначена як  $MA(q)$ , передбачає, що поточне значення часового ряду може бути описане через лінійну комбінацію попередніх шумових помилок. MA моделі є корисними у випадках, коли шумові помилки мають значний вплив на динаміку часових рядів. Ці моделі часто застосовуються для «згладжування» часових рядів, видаляючи випадкові коливання та підкреслюючи більш значущі тренди і патерни.

Подібно до AR компоненту, параметри моделі MA можуть бути оцінені через автокореляційну функцію (ACF) і часткову автокореляційну функцію (PACF). Типово, для чистих MA процесів, теоретична ACF швидко знижується до нуля після лагу  $q$ , тоді як PACF може проявляти більш складні шаблони.

MA моделі припускають, що шумові помилки є незалежними і однаково розподіленими з нульовим середнім та постійною дисперсією, що може не завжди бути реалістичним для реальних фінансових часових рядів. Як і AR моделі, MA не завжди ефективні для обробки нестационарних даних без відповідного диференціювання або інших перетворень.

MA компонент в ARIMA дозволяє краще моделювати та прогнозувати часові ряди, де присутні випадкові коливання та шум, надаючи більш гладкі та стабільні прогнози. Він є важливим інструментом у аналітиці часових рядів, особливо в областях, де дані схильні до випадкових шоків або помилок вимірювання.

Процес побудови моделі ARIMA проходить у кілька кроків:

а) перший крок полягає в ідентифікації відповідних значень для параметрів  $p$ ,  $d$ , і  $q$ . Це може вимагати аналізу автокореляційної функції (ACF) та часткової автокореляційної функції (PACF) для визначення авторегресійних та ковзних середніх компонентів;

б) після ідентифікації параметрів модель оцінюється шляхом максимізації правдоподібності та/або мінімізації інформаційних критеріїв, таких як AIC або BIC.

в) перевірка адекватності побудованої моделі зазвичай включає аналіз залишків на предмет випадковості, стаціонарності та відсутності автокореляції.

ARIMA є міцним інструментом у портфелі фінансового аналітика чи економіста завдяки своїй універсальності та спроможності моделювати широкий спектр часових рядів у різних областях. Вона ефективна для даних з трендами і сезонністю. Проста в імплементації та інтерпретації. Проте обмежена до лінійних відносин. Вимагає стаціонарних даних для точного прогнозування.

Одним із сучасних підходів до передбачення фінансової інформації, у тому числі динаміки фондових індексів, є застосування Recurrent Neural Networks (RNN), що є класом нейронних мереж, що ефективно обробляє послідовні дані. Вони здатні зберігати інформацію про попередні входи, що робить їх ідеальними для задач, де контекст з минулих спостережень важливий для розуміння поточного чи майбутнього стану.

Ядро RNN полягає у здатності зберігати інформацію в «прихованих станах», що по суті є пам'яттю мережі. Кожен прихований стан містить інформацію, засновану не лише на поточному входному сигналі, а й на попередніх входах. Це дозволяє мережі враховувати попередній контекст при обробці нових даних.

Вхідний шар приймає вхідні дані і передає їх у вигляді активацій до наступного шару – зазвичай, до прихованого шару. У RNN, кожен вхід обробляється один за одним, із послідовною передачею активацій.

Після обробки вхідних даних через приховані шари, інформація передається до вихідного шару, який генерує кінцевий результат. Це може бути класифікація, прогнозування наступного елементу в послідовності тощо.

#### Повторюваність та Петлі

У відміну від традиційних нейронних мереж, де інформація рухається в одному напрямку (вперед), RNN має циклічні зв'язки, що дозволяють інформації

передаватися назад по мережі. Це зворотнє зв'язування забезпечує зворотній потік інформації і є основою для зберігання пам'яті про попередні входи.

Архітектура RNN значно вплинула на область машинного навчання, особливо у сферах, де критично важливо розуміння контексту та послідовності, таких як обробка природної мови та прогнозування часових рядів. Їх здатність інтегрувати минулу інформацію для вирішення поточних завдань робить їх незамінними в багатьох сучасних застосуваннях.

Унікальна архітектура дозволяє RNN ефективно обробляти послідовні дані, роблячи їх ідеальними для широкого спектра застосувань у різних областях, зокрема у прогнозуванні часових рядів, обробці природної мови, а також у музиці та біоінформатиці. В моделях прогнозування фінансових ринків, таких як ціни на акції або обмінні курси, RNN використовуються для аналізу історичних даних та передбачення майбутніх тенденцій.

Хоча RNN надзвичайно потужні у роботі з послідовностями даних, вона стикається з декількома важливими технічними проблемами, які можуть ускладнити її використання та ефективність. Основні проблеми включають зникнення або вибух градієнтів, обмежену пам'ять, важкість у навчанні, та проблеми із залежностями на великі дистанції.

Коли градієнти дуже малі, вони ефективно «зникають» під час навчання, що робить важким або неможливим оновлення ваг у глибших шарах мережі. Це призводить до того, що ранні дані в послідовності мають дуже малий або ніякий вплив на вихід, що робить RNN неефективними для навчання на довгих послідовностях.

Протилежною проблемою є вибух градієнтів, коли градієнти стають настільки великими, що призводять до нестабільності навчального процесу. Це може викликати значні коливання ваг, що робить нейронну мережу нестабільною і може призвести до числових переповнень.

Однією з основних теоретичних обмежень RNN є їх здатність зберігати інформацію лише з недавнього минулого. При обробці дуже довгих

послідовностей, важливість ранніх даних може бути «забута» мережею, що призводить до втрати контексту та точності у прогнозах або рішеннях.

RNN може бути складною для ефективного навчання через високу чутливість до налаштувань гіперпараметрів, таких як швидкість навчання, розмір шарів, і тип оптимізатора. Неправильне налаштування може легко призвести до піднавчання або перенавчання.

Стандартна RNN часто бореться з виявленням залежностей у даних, які відокремлені великими інтервалами часу. Ця обмежена здатність моделювати довготривалі залежності обмежує її ефективність в деяких застосуваннях, наприклад, у деяких типах аналізу послідовностей або комплексних завданнях прогнозування.

LSTM (Long Short-Term Memory) і GRU (Gated Recurrent Units) були розроблені як відповідь на деякі з цих проблем. Вони включають механізми воріт, які дозволяють контролювати потік інформації всередині мережі, зберігаючи важливу інформацію на довші періоди і знижуючи вплив зникнення або вибуху градієнтів.

Незважаючи на заявленні недоліки, RNN залишається потужним інструментом у глибинному навчанні, особливо коли вона використовується в поєднанні з сучасними поліпшеннями, що допомагають подолати властиві їй обмеження.

Як вже було зазначено вище, LSTM (Long Short-Term Memory) є розширенням традиційної RNN. Її було спеціально розроблено для вирішення проблеми «зникаючого» або «вибухаючого» градієнтів, яка часто виникає при роботі з базовою RNN. LSTM здатна ефективно обробляти довготривалі залежності в даних, що робить її особливо корисною для складних завдань, які вимагають розуміння контексту з великих послідовностей даних.

LSTM мережа відрізняється від RNN наявністю так званих «воріт» (gates), які регулюють потік інформації. Ці ворота допомагають моделі вирішити, коли оновлювати внутрішній стан, коли скидати інформацію, яка більше не потрібна,

і коли виводити інформацію на наступні шари. Основні компоненти воріт у LSTM включають:

Forget Gate, що визначає, яка частина інформації з попереднього стану повинна бути «забута» або відкинута. Це робиться за допомогою сигмоїдної функції, яка виводить значення між 0 (забути все) та 1 (зберегти все).

Input Gate – вирішує, яка нова інформація повинна бути додана до внутрішнього стану. Це включає в себе дві частини: сигмоїдну функцію, яка визначає важливість оновлюваної інформації, та тангенс гіперболічний, який створює нові кандидатські вектори для додавання до стану.

Output Gate – вирішує, яка частина внутрішнього стану повинна бути передана на вихід. Подібно до input gate, включає сигмоїдну функцію для визначення важливості вихідних значень і тангенс гіперболічний для регуляції вихідних значень на основі поточного стану.

LSTM широко застосовуються в багатьох галузях, зокрема для прогнозування часових рядів: від фінансових ринків до погодних умов, LSTM можуть ефективно прогнозувати майбутні значення на основі довгих історичних даних.

LSTM є потужним різновидом RNN. Вона пропонує ряд значущих переваг, особливо коли мова йде про роботу з довготривалими залежностями в послідовностях даних:

а) завдяки унікальній архітектурі з воротами, LSTM може зберігати інформацію на тривалі періоди часу. Це дозволяє їй ефективно вирішувати проблеми, які вимагають залучення знань про раніше відбулі події, такі як контекст у текстах або патерни у фінансових часових рядах.

б) LSTM здатна розпізнавати та моделювати складні залежності в даних. Це робить її ідеальною для застосувань, де потрібно розуміти контекст або залежності, які простягаються на значні відрізки часу;

в) LSTM здатна уникати проблему зникнення градієнтів. Механізми воріт допомагають підтримувати стабільність градієнтів протягом тривалого навчання, що сприяє кращому засвоєнню даних;

г) ворота в LSTM дозволяють моделі вибірково оновлювати пам'ять, забувати непотрібну інформацію та зберігати тільки корисні дані. Це підвищує ефективність та адаптивність моделі до різних типів даних та задач.

д) наявність воріт також допомагає контролювати ризик перенавчання, оскільки модель може регулювати кількість інформації, що проходить через мережу, тим самим уникнувши перевантаження моделі занадто багатою деталізацією з навчальних даних.

LSTM є однією з найпопулярніших та найбільш ефективних архітектур в глибокому навчанні для роботи з послідовностями, їхні переваги роблять їх вибором для багатьох важливих застосувань в індустрії і дослідженнях.

Розробку та застосування зазначених моделей у рамках дослідження було здійснено за допомогою мови програмування Python. Вона є однією з найпопулярніших мов програмування у світі аналізу даних і машинного навчання, зокрема для розробки моделей, у тому числі таких як ARIMA, RNN та LSTM, з декількох ключових причин, що перелічені нижче.

Python підтримується обширним набором високоякісних бібліотек для наукових обчислень, статистичного аналізу та машинного навчання. Далі наведено кілька ключових бібліотек, які підтримують розробку ARIMA, RNN і LSTM моделей:

NumPy і SciPy – для загальних наукових обчислень.

Pandas – забезпечує високопродуктивні, гнучкі структури даних для маніпуляцій з даними та аналізу.

Statsmodels – особливо корисна для статистичного моделювання та включає реалізацію ARIMA.

Scikit-learn – для простих методів машинного навчання.

TensorFlow і Keras – забезпечують інтуїтивні API для побудови і тренування нейронних мереж, включаючи RNN та LSTM.

PyTorch – інша популярна бібліотека для глибокого навчання, яка є вибором для досліджень і розробки в галузі штучного інтелекту.

Python має одну з найбільших і найактивніших спільнот розробників у світі. Велика кількість навчальних матеріалів, відкритих проєктів, форумів та конференцій дозволяє легко знайти відповіді на питання, здобути нові знання та отримати підтримку в рішенні технічних проблем.

Python відомий своїм чистим та читабельним синтаксисом, який робить програмування зрозумілим та доступним для новачків. Це сприяє швидкому прототипуванню і експериментуванню, що є критично важливим у дослідженнях і розробці моделей машинного навчання.

Python легко інтегрується з більшістю баз даних, веб-додатків, інструментів візуалізації даних і багатьма іншими технологіями. Це робить його ідеальним вибором для вбудовування статистичного моделювання та машинного навчання в ширші додатки та системи.

Python є універсальною мовою, яка може бути використана не тільки для статистичного аналізу та машинного навчання, але й для розробки веб-додатків, скриптів автоматизації, системних утиліт та багато іншого. Ця універсальність робить його ідеальним вибором для компаній, які хочуть стандартизувати мову програмування на різних етапах розробки продукту.

Використання Python для розробки ARIMA, RNN, та LSTM моделей забезпечує могутні інструменти, значну підтримку спільноти, та гнучкість у застосуванні, роблячи його одним з найкращих виборів для науковців, інженерів, та розробників, які працюють у галузі аналізу даних та штучного інтелекту.



## 2.4. Опис та обґрунтування методів оцінки результатів дослідження

Для оцінки результатів моделювання запропоновано використати наступні метрики:

- середня абсолютна помилка;
- середньоквадратична помилка;
- коренева середньоквадратична помилка;
- середня абсолютна відсоткова помилка;
- симетрична середня абсолютна відсоткова помилка.

Середня абсолютна помилка (Mean Absolute Error – MAE) – є однією з основних метрик для оцінки точності прогнозів у статистиці та машинному навчанні, особливо у задачах регресії. MAE вимірює середню абсолютну відстань між фактичними значеннями та прогнозами, наданими моделлю. Якщо при прогнозі отримано значення MAE на рівні 50 000, це означає, що в середньому прогнозні результати відрізняються від фактичних на це значення – 50 000.

Формула для розрахунку:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|$$

де

$Y_i$  – фактичні значення;

$\hat{Y}_i$  – прогнозовані значення;

$n$  – кількість спостережень.

Плюси MAE:

а) Робастність до викидів. Відмінною особливістю MAE є те, що вона менш чутлива до викидів у порівнянні з MSE. Великі або малі помилки

впливають на результати оцінки лінійно, що забезпечує більш стабільну оцінку помилок.

б) Легкість інтерпретації. MAE вимірює середню помилку в тих самих одиницях, що й фактичні дані, що робить результати легко інтерпретованими.

Мінуси MAE:

а) Менша чутливість до великих помилок. Через лінійний характер оцінки помилок, MAE може недооцінювати вплив дуже великих помилок, що може бути критичним в деяких застосуваннях, де великі помилки мають більші наслідки.

б) Відсутність гладкості. Оскільки абсолютне значення є недиференційованим в точці 0, оптимізація моделей, що використовують MAE як функцію втрат, може бути ускладнена у порівнянні з MSE, що має гладкі властивості.

MAE широко використовується в різних областях, де важлива проста і зрозуміла метрика оцінки точності моделі, особливо коли необхідно мінімізувати вплив аномальних даних. Застосування MAE допомагає забезпечити стійкість оцінки моделей до викидів, хоча це і може призводити до недооцінки великих помилок у деяких сценаріях.

Середньоквадратична помилка (Mean Squared Error – MSE) – є однією з найпопулярніших метрик для оцінювання якості моделей у статистиці та машинному навчанні, особливо у задачах регресії. Вона вимірює середню квадратичну відстань між фактичними значеннями та прогнозованими значеннями, які видає модель. Якщо, при оцінці прогнозних результатів, величина MSE становить 100, це свідчить про те, що в середньому прогноз відрізняється від фактичних даних приблизно на 10 одиниць (у більшу або меншу сторону).

Розраховується за формулою:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

де

$Y_i$  – фактичні значення;

$\hat{Y}_i$  – прогнозовані значення;

$n$  – кількість спостережень.

Плюси MSE:

а) Чутливість до великих помилок. Завдяки квадратичній компоненті, MSE сильно покарає модель за великі помилки, що робить цю метрику корисною, коли важливо мінімізувати великі відхилення.

б) Диференційованість. MSE є гладкою функцією, що її легко оптимізувати за допомогою числових методів, таких як градієнтний спуск.

с) Легкість інтерпретації. Вона вимірює середню квадратичну помилку, що забезпечує зручне порівняння між різними моделями та датасетами.

Мінуси MSE:

а) Чутливість до викидів. Квадратична компонента робить MSE особливо чутливою до викидів (аномально великих або малих значень), що може призвести до того, що оцінка моделі буде зосереджена на декількох непрезентативних даних.

б) Масштаб залежність. Значення MSE залежить від масштабу цільових значень, тому порівнювати MSE між задачами з різними масштабами даних може бути не коректно.

с) Нелінійність відповідей. Великі відхилення від середнього підвищують помилку більше, ніж малі, що може створювати проблеми у випадках, коли малі помилки є критично важливими.

MSE застосовується в різноманітних областях, включаючи прогнозування в економіці, аналіз часових рядів у метеорології, оцінювання моделей у фінансах та багатьох інших. Це дозволяє отримати вагоме уявлення про ефективність моделей, але потрібно застосовувати з обережністю, особливо при наявності викидів в даних.

Коренева середньоквадратична помилка (Root Mean Squared Error – RMSE) є однією з базових метрик для оцінки точності прогнозів у статистиці та машинному навчанні, особливо у задачах регресії. RMSE вимірює стандартне відхилення помилок прогнозу, тобто середнє квадратичне відхилення між фактичними значеннями та прогнозами, наданими моделлю. Як і у випадку MSE, отримана величина RMSE на рівні 10 одиниць, свідчить про середнє відхилення прогнозу від фактичних результатів на  $\pm 10$  одиниць.

Формула для RMSE:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}$$

де

$Y_i$  – фактичні значення;

$\hat{Y}_i$  – прогнозовані значення;

$n$  – кількість спостережень.

Плюси застосування RMSE:

- a) Чутливість до великих помилок. Як і MSE, RMSE карає більш суворо великі помилки (вони впливають на результат квадратично), що робить цю метрику корисною в застосуваннях, де важливо мінімізувати великі відхилення.
- b) Легкість інтерпретації. Оскільки RMSE виражається в тих самих одиницях, що й цільова змінна, її легше інтерпретувати порівняно з MSE.

с) Диференційованість. RMSE є гладкою функцією, що дозволяє використовувати ефективні методи оптимізації, такі як градієнтний спуск.

Мінуси застосування RMSE:

а) Чутливість до викидів. Як і MSE, RMSE дуже чутлива до викидів, що може призвести до непропорційного впливу окремих аномально великих помилок на загальну оцінку моделі.

б) Залежність від масштабу. Значення RMSE залежить від масштабу цільових значень, тому порівнювати RMSE між задачами з різними масштабами даних може бути не коректно.

с) Складність в порівнянні з MAE. RMSE, будучи більш чутливою до великих помилок, може бути не такою робастною, як MAE, в деяких контекстах, де викиди мають мале значення.

RMSE використовується в широкому спектрі областей, включаючи фінансове прогнозування, метеорологію, енергетику та багато інших. Ця метрика корисна для оцінки та порівняння моделей, особливо коли потрібно зосередитися на мінімізації великих помилок і водночас потрібна чітка інтерпретабельність результатів.

Середня абсолютна відсоткова помилка (Mean Absolute Percentage Error – MAPE) — це метрика, що вимірює точність прогнозів моделі у відсотках. MAPE показує, наскільки в середньому прогнози відхиляються від фактичних значень у відсотковому виразі, і часто використовується у фінансах, продажах, економіці та інших областях, де необхідно оцінювати точність прогнозів. Якщо MAPE для прогнозування становить 10%, це означає, що в середньому прогнози відрізняються від фактів на зазначені 10%.

Формула MAPE:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{|Y_i - \hat{Y}_i|}{|Y_i|}$$

де

$Y_i$  – фактичні значення;

$\hat{Y}_i$  – прогнозовані значення;

$n$  – кількість спостережень.

Плюси застосування MAPE:

a) Інтуїтивна зрозумілість: MAPE вимірює помилки у відсотковому виразі, що робить цю метрику легко зрозумілою та інтерпретованою для нефахівців.

b) Незалежність від масштабу. Відсоткова помилка дозволяє порівнювати моделі та їх прогнози між різними датасетами, незалежно від масштабу даних.

Мінуси застосування MAPE:

a) Проблеми з нульовими або дуже малими фактичними значеннями. Коли вони дуже малі (близькі до нуля), ділення на це значення може призвести до нескінченно великих результатів MAPE або до необґрунтованих спотворень.

b) Асиметрія. Оскільки MAPE використовує абсолютні значення відхилень, прогнози, які перевищують фактичні значення, і прогнози, які їх недооцінюють, мають однаковий вплив на середнє значення помилки, незалежно від контексту, в якому важливіші надмірні або недостатні оцінки.

MAPE широко застосовується у багатьох областях, де важливо оцінити відсоткові помилки між прогнозованими та фактичними даними, але потрібно бути уважним до специфіки даних, особливо при наявності малих значень. В таких випадках можуть бути використані модифікації MAPE, такі як wMAPE (важена MAPE), яка уникає деяких недоліків стандартної версії.

Симетрична середня абсолютна відсоткова помилка (Symmetric Mean Absolute Percentage Error – SMAPE) – це варіація звичайної метрики MAPE, розроблена для виправлення деяких її недоліків, зокрема асиметрії та проблем, пов'язаних із нульовими або малими значеннями в деномінаторі. SMAPE є більш

збалансованою метрикою, яка враховує величину як фактичних, так і прогнозованих значень для мінімізації впливу екстремальних значень. Якщо SMAPE при оцінці прогнозування становить 15%, це означає, що в середньому прогнозні значення відрізняються від фактичних на 15% відносно середнього арифметичного прогнозованих та фактичних результатів.

SMAPE розраховується за наступною формулою:

$$SMAPE = \frac{100}{n} \sum_{i=1}^n \frac{|Y_i - \hat{Y}_i|}{(|Y_i| + |\hat{Y}_i|) / 2}$$

де

$Y_i$  – фактичні значення;

$\hat{Y}_i$  – прогнозовані значення;

$n$  – кількість спостережень.

Плюси застосування SMAPE:

а) Симетрія. На відміну від MAPE, SMAPE однаково покарає як надмірні, так і недостатні прогнози, завдяки симетричному врахуванню фактичних та прогнозованих значень.

б) Менша чутливість до екстремальних значень. Включення суми абсолютних значень у знаменнику зменшує ефект дуже малих або великих значень, роблячи метрику стабільнішою та надійнішою.

с) Легкість інтерпретації. SMAPE вимірює помилки у відсотковому виразі, що робить результати зрозумілими та легко порівнюваними.

Мінуси застосування SMAPE:

а) Проблеми з нульовими значеннями. Хоча SMAPE і менш чутлива до малих значень порівняно з MAPE, проблеми все ще можуть виникнути, коли обидва значення  $Y_i$  та  $\hat{Y}_i$  нульові, оскільки це призводить до невизначеності у формулі.

б) Обмежена диференційованість. Як і MAPE, SMAPE може мати точки, де вона не диференційована, що ускладнює оптимізацію моделей за допомогою градієнтного спуску.

SMAPE часто використовується у багатьох галузях, особливо в економіці та фінансах, для забезпечення більш справедливого порівняння між моделями, що генерують прогнози для даних різного масштабу. Ця метрика забезпечує більш збалансовану оцінку точності прогнозів.

Вибір метрики залежить від контексту задачі, характеристик даних та вимог до інтерпретації. MAE та RMSE часто використовуються для задач регресії, MAPE та SMAPE - для прогнозування часових рядів та задач, де потрібно порівнювати похибки між різними наборами даних. Важливо також враховувати чутливість метрик до викидів та нульових/близьких до нуля значень.



## 2.5. Результати прогнозування та їх аналіз

Результати моделювання значень фондових індексів та їх оцінка представлені на наступних малюнках та таблицях.

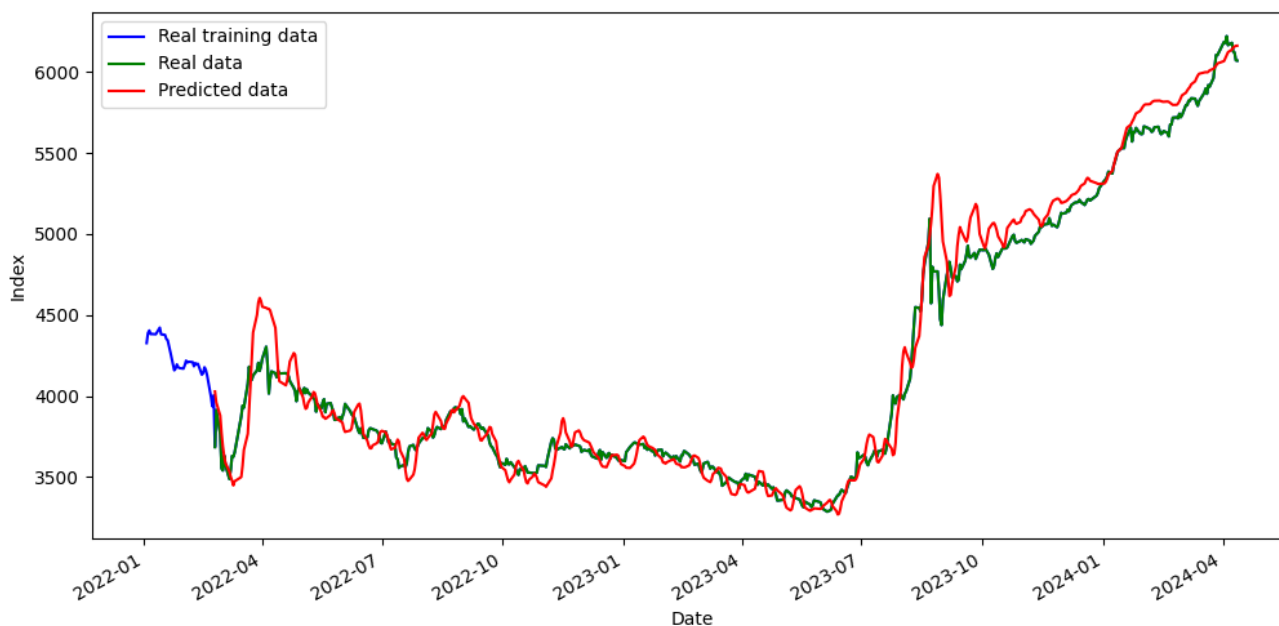
Малюнок 1.

Прогноз значень індексу BUMIX на 3 дні за допомогою моделі ARIMA



Малюнок 2.

Прогноз значень індексу BUMIX на 3 дні за допомогою RNN мережі



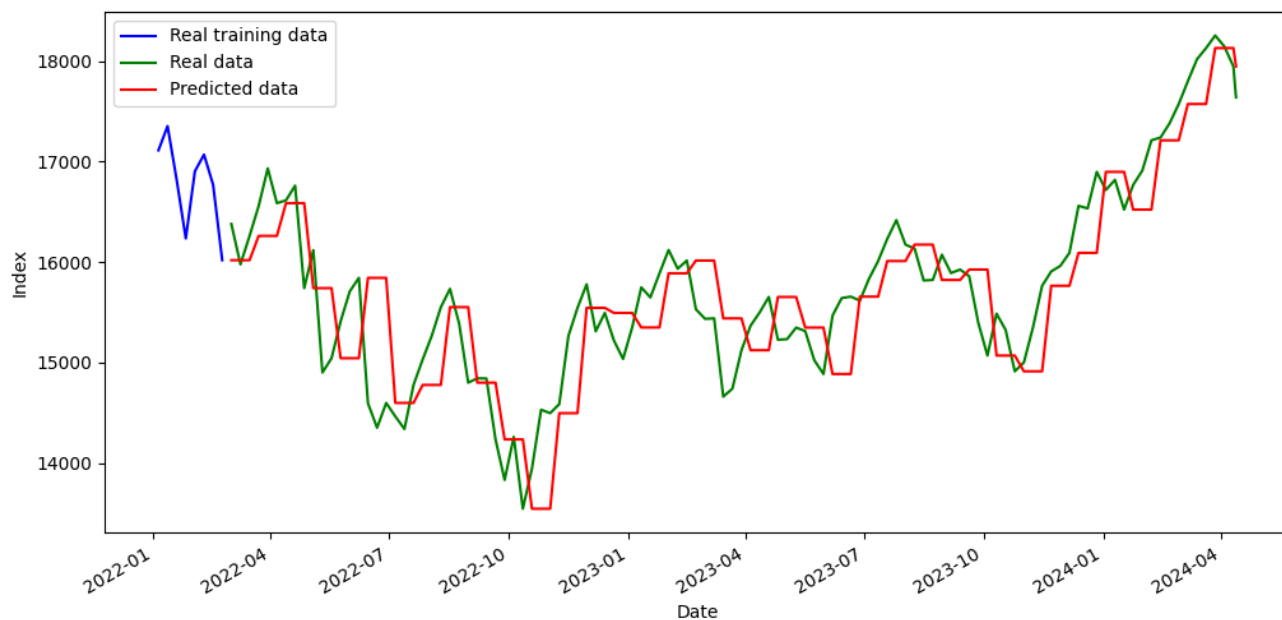
Малюнок 3.

Прогноз значень індексу BUMIX на 3 дні за допомогою LSTM мережі



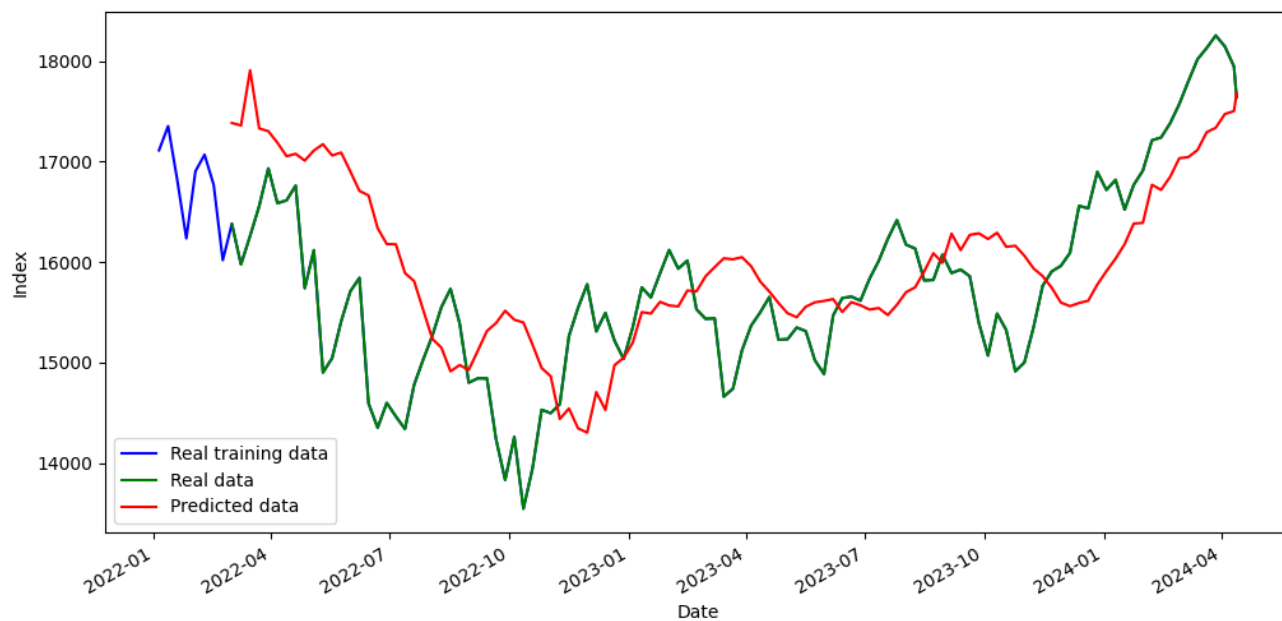
Малюнок 4.

Прогноз значень індексу NYSE на 3 тижні за допомогою моделі ARIMA



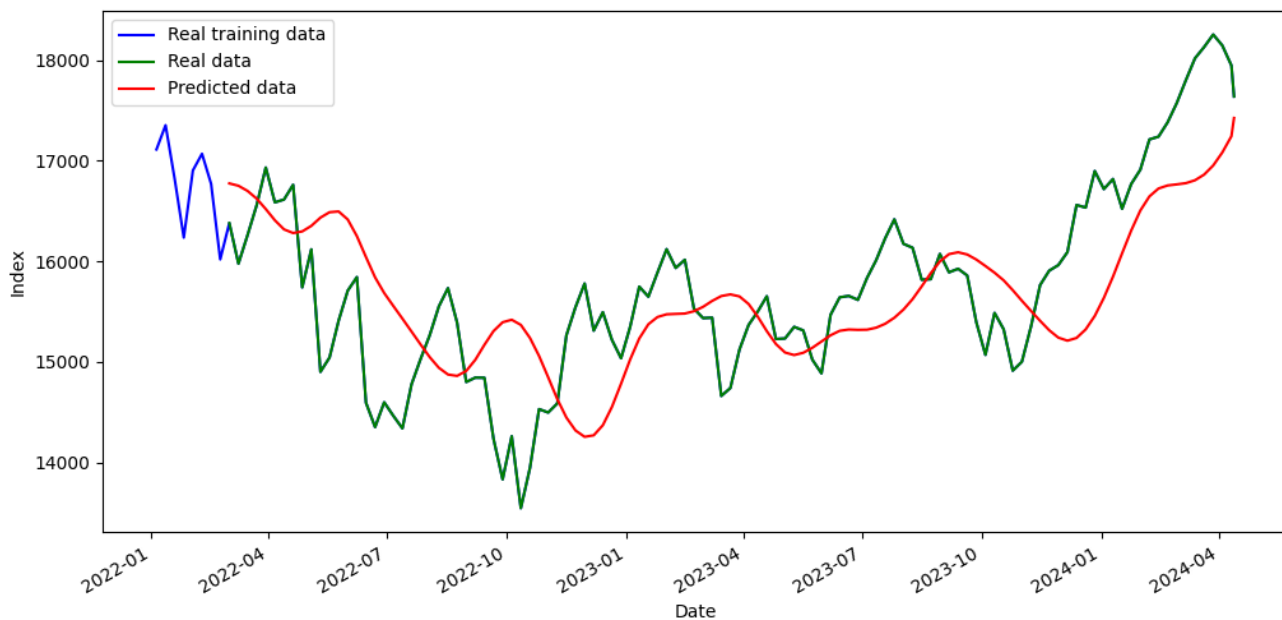
Малюнок 5.

Прогноз значень індексу NYSE на 3 тижні за допомогою RNN мережі



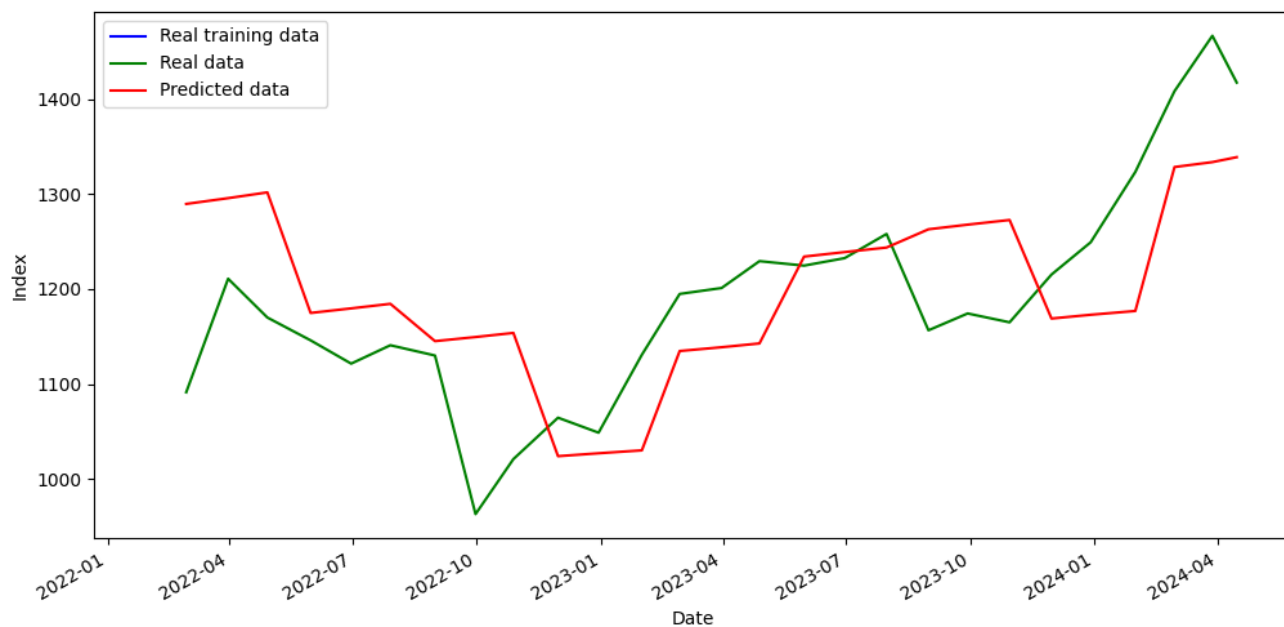
Малюнок 6.

Прогноз значень індексу NYSE на 3 тижні за допомогою LSTM мережі



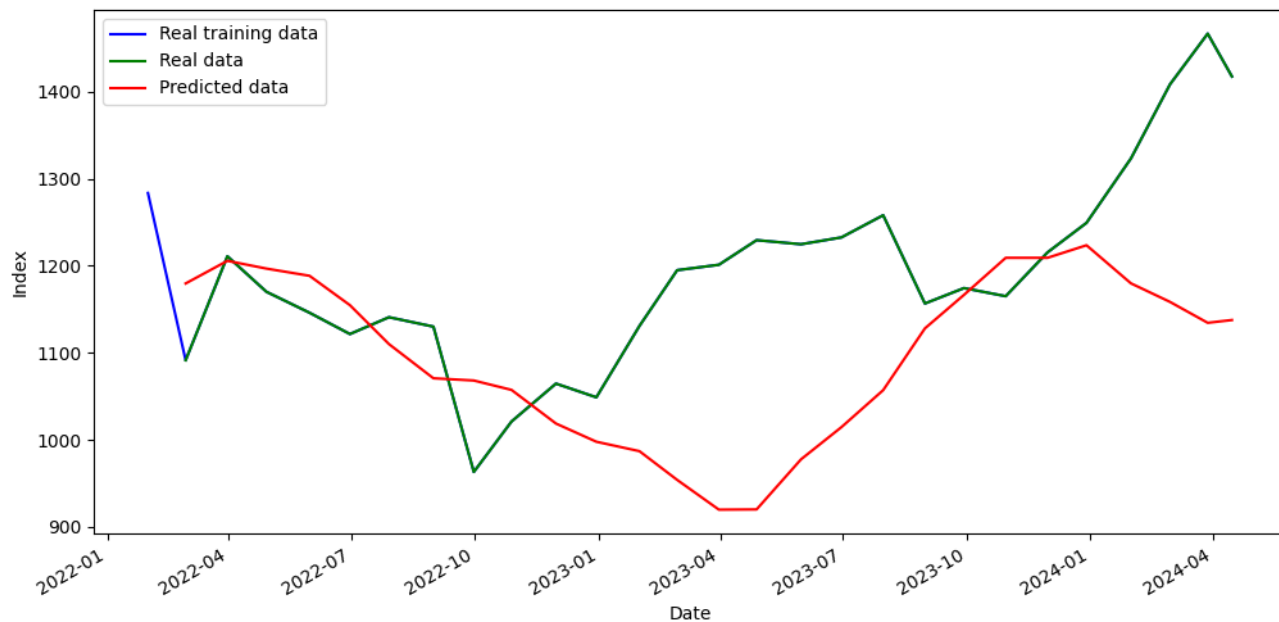
Малюнок 7.

Прогноз значень індексу SBIPOR на 3 місяці за допомогою моделі ARIMA



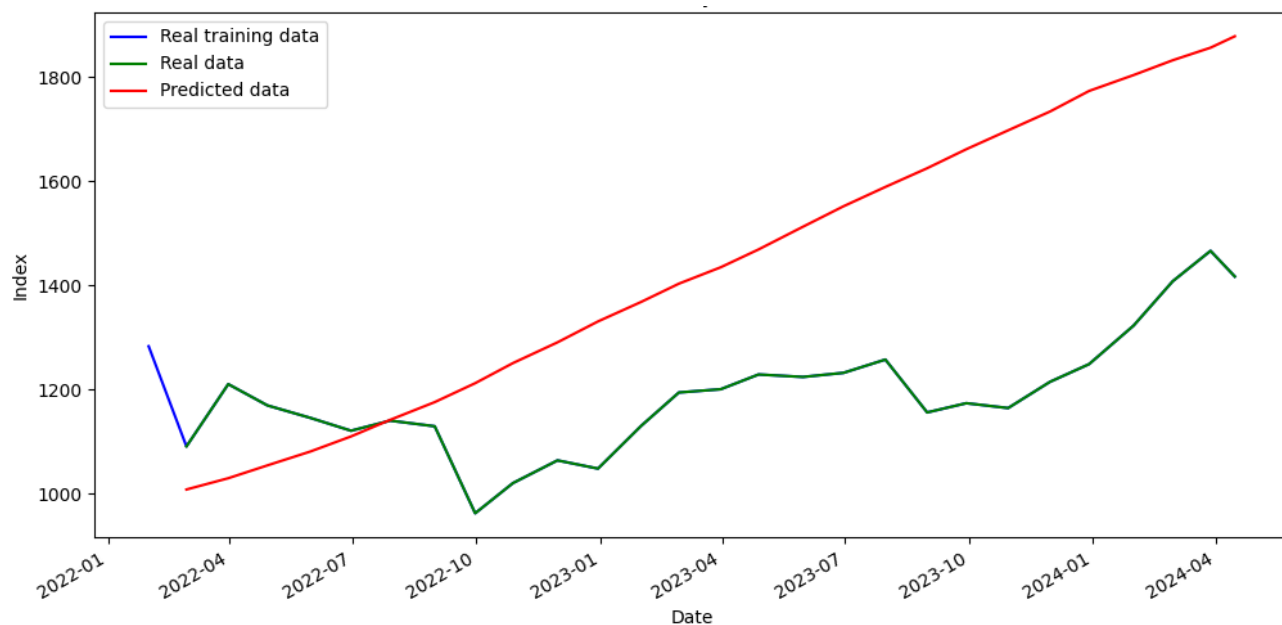
Малюнок 8.

Прогноз значень індексу SBIPOR на 3 місяці за допомогою RNN мережі



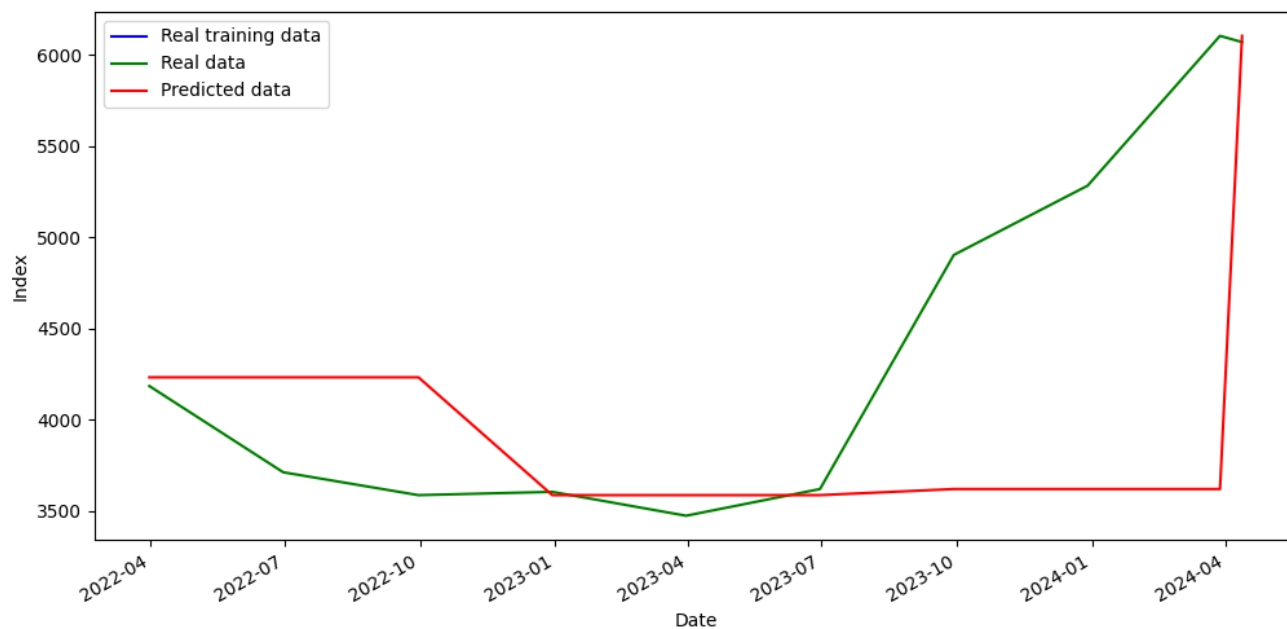
Малюнок 9.

Прогноз значень індексу SBIPOR на 3 місяці за допомогою LSTM мережі



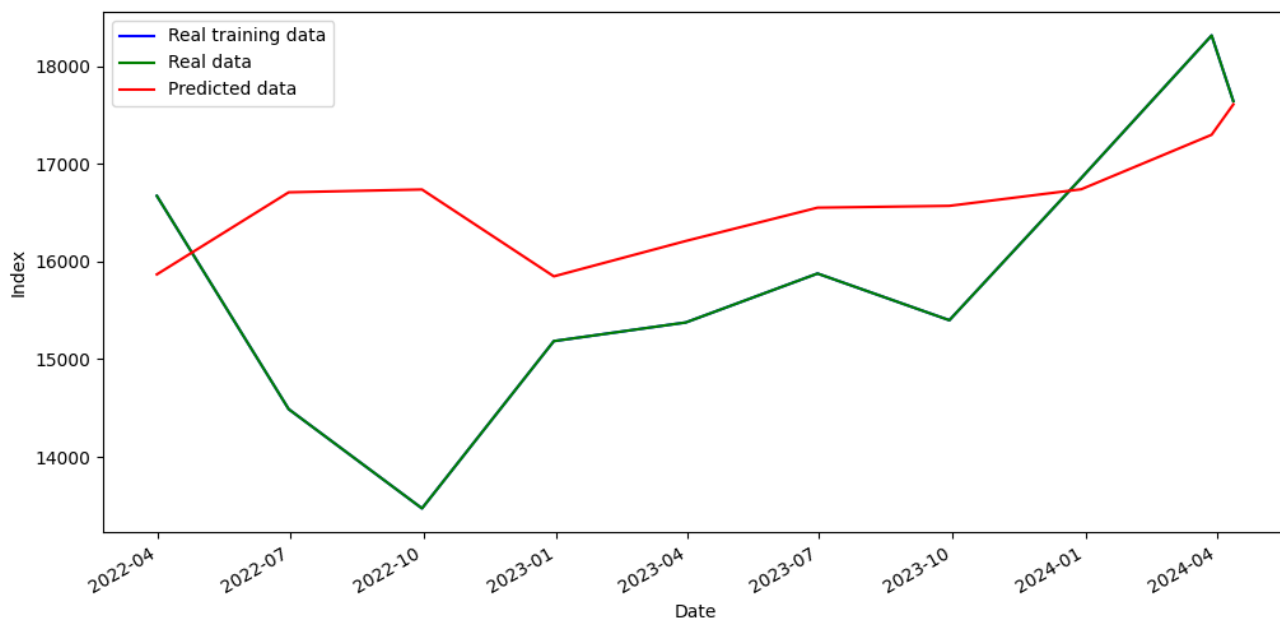
Малюнок 10.

Прогноз значень індексу BUMIX на 3 квартали за допомогою моделі ARIMA



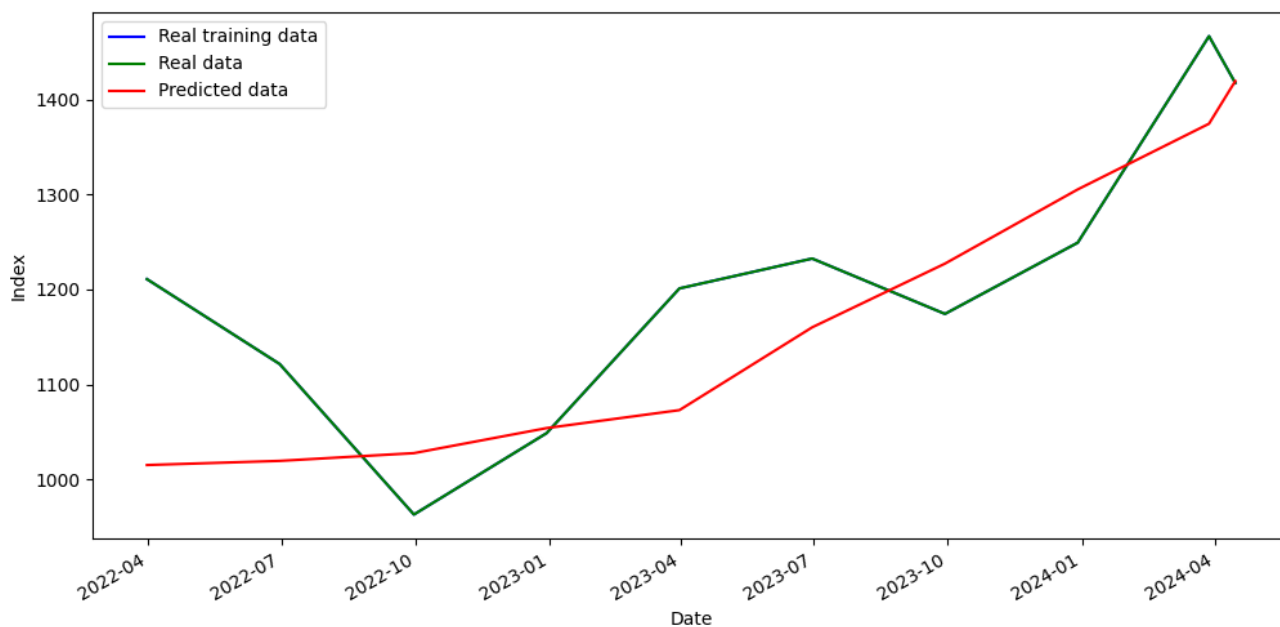
Малюнок 11.

Прогноз значень індексу NYSE на 3 квартали за допомогою RNN мережі



Малюнок 12.

Прогноз значень індексу SBIPOP на 3 квартали за допомогою LSTM мережі



Як можна побачити з наведених графіків, за певних умов, усі обрані методи оцінки та прогнозування майбутніх значень фондових індексів дають досить гарні результати. Оцінки усіх методів для прогнозів на різні періоди часу, що здійсненні за допомогою обраних метрик, наведено у таблицях нижче.

Таблиця 1.

Результати прогнозування значень індексу BUMIX на періоди від 1 до 5 днів

Оцінка	Модель	Період для передбачення (днів)				
		1	2	3	4	5
MAE	ARIMA	30,29	37,44	43,67	48,38	55,13
	RNN	96,85	99,50	96,66	111,37	135,13
	LSTM	68,79	82,47	89,59	103,91	108,34
MSE	ARIMA	2 680,31	3 760,25	5 147,68	5 682,62	8 702,67
	RNN	25 374,01	25 797,70	19 392,79	32 901,00	45 166,82
	LSTM	11 191,39	14 165,72	18 002,39	24 159,72	26 779,83
RMSE	ARIMA	51,77	61,32	71,75	75,38	93,29
	RNN	159,29	160,62	139,26	181,39	212,52
	LSTM	105,79	119,02	134,17	155,43	163,65
MAPE	ARIMA	0,01	0,01	0,01	0,01	0,01
	RNN	0,02	0,02	0,02	0,03	0,03
	LSTM	0,02	0,02	0,02	0,02	0,03
SMAPE	ARIMA	0,73	0,90	1,05	1,18	1,32
	RNN	2,31	2,39	2,26	2,62	3,23
	LSTM	1,68	2,04	2,19	2,51	2,60

Таблиця 2.

Результати прогнозування значень індексу BUMIX на періоди від 1 до 5 тижнів

Оцінка	Модель	Період для передбачення (тижнів)				
		1	2	3	4	5
MAE	ARIMA	73,16	89,31	116,03	118,06	136,14
	RNN	214,85	231,71	451,24	300,30	369,09
	LSTM	356,19	415,16	458,83	514,01	561,05
MSE	ARIMA	11 750,49	16 508,18	25 600,02	26 156,41	37 570,94
	RNN	70 614,33	89 047,83	305 079,04	140 009,74	223 679,57
	LSTM	189 483,37	241 720,23	311 234,21	367 683,55	493 075,97

Таблиця 2.  
(продовження)

Оцінка	Модель	Період для передбачення (тижнів)				
		1	2	3	4	5
RMSE	ARIMA	108,40	128,48	160,00	161,73	193,83
	RNN	265,73	298,41	552,34	374,18	472,95
	LSTM	435,30	491,65	557,88	606,37	702,19
MAPE	ARIMA	0,02	0,02	0,03	0,03	0,03
	RNN	0,05	0,06	0,12	0,07	0,09
	LSTM	0,09	0,10	0,11	0,13	0,14
SMAPE	ARIMA	1,78	2,16	2,81	2,84	3,30
	RNN	5,24	5,88	10,81	7,08	8,67
	LSTM	8,69	10,02	11,10	12,38	13,28

Таблиця 3.

Результати прогнозування значень індексу BUMIX на періоди від 1 до 5 місяців

Оцінка	Модель	Період для передбачення (місяців)				
		1	2	3	4	5
MAE	ARIMA	183,52	210,81	276,49	289,91	313,36
	RNN	597,97	558,81	705,32	1 200,58	887,91
	LSTM	1 034,08	1 058,73	1 098,42	1 095,42	1 167,24
MSE	ARIMA	54 861,04	84 717,45	133 311,63	201 782,34	222 162,28
	RNN	554 835,04	524 643,51	579 835,63	1 891 672,75	931 851,82
	LSTM	1 251 765,03	1 349 198,45	1 468 681,83	1 459 422,30	1 675 225,67
RMSE	ARIMA	234,22	291,06	365,12	449,20	471,34
	RNN	744,87	724,32	761,47	1 375,38	965,32
	LSTM	1 118,82	1 161,55	1 211,89	1 208,07	1 294,31
MAPE	ARIMA	0,04	0,05	0,06	0,07	0,07
	RNN	0,14	0,13	0,17	0,32	0,23
	LSTM	0,26	0,27	0,29	0,28	0,31
SMAPE	ARIMA	4,37	5,04	6,47	7,04	7,45
	RNN	14,11	13,33	16,42	26,34	20,23
	LSTM	23,19	23,66	24,41	24,36	25,71



Таблиця 4

Результати прогнозування значень індексу BUMIX на періоди від 1 до 5 кварталів

Оцінка	Модель	Період для передбачення (кварталів)				
		1	2	3	4	5
MAE	ARIMA	345,95	550,14	684,31	657,31	1 121,25
	RNN	1 253,76	1 226,06	1 475,41	1 210,15	1 263,05
	LSTM	944,23	993,87	1 072,41	1 115,72	1 153,78
MSE	ARIMA	274 624,15	602 310,29	1 128 857,78	689 816,52	2 065 766,34
	RNN	2 288 098,30	1 975 563,66	2 897 238,50	1 933 829,81	2 119 760,60
	LSTM	1 076 645,41	1 185 015,31	1 414 516,52	1 558 219,96	1 703 153,06
RMSE	ARIMA	524,05	776,09	1 062,48	830,55	1 437,28
	RNN	1 512,65	1 405,55	1 702,13	1 390,62	1 455,94
	LSTM	1 037,62	1 088,58	1 189,33	1 248,29	1 305,05
MAPE	ARIMA	0,07	0,11	0,14	0,14	0,23
	RNN	0,33	0,31	0,39	0,31	0,33
	LSTM	0,23	0,25	0,27	0,28	0,30
SMAPE	ARIMA	7,74	12,05	15,45	14,75	25,54
	RNN	26,63	26,19	30,46	25,82	26,78
	LSTM	20,88	21,89	23,41	24,22	24,92

Таблиця 5.

Результати прогнозування значень індексу NYSE на періоди від 1 до 5 днів

Оцінка	Модель	Період для передбачення (днів)				
		1	2	3	4	5
MAE	ARIMA	123,82	144,04	162,04	176,80	209,79
	RNN	178,16	219,57	272,70	292,01	337,90
	LSTM	182,77	219,72	251,27	285,12	310,49
MSE	ARIMA	26 218,19	35 345,86	47 171,40	55 378,53	74 830,80
	RNN	55 786,10	86 674,99	130 883,40	143 128,95	194 121,39
	LSTM	58 996,27	87 304,33	108 917,98	133 838,64	156 712,02
RMSE	ARIMA	161,92	188,00	217,19	235,33	273,55
	RNN	236,19	294,41	361,78	378,32	440,59
	LSTM	242,89	295,47	330,03	365,84	395,87
MAPE	ARIMA	0,01	0,01	0,01	0,01	0,01
	RNN	0,01	0,01	0,02	0,02	0,02
	LSTM	0,01	0,01	0,02	0,02	0,02

Таблиця 5.

(продовження)

Оцінка	Модель	Період для передбачення (днів)				
		1	2	3	4	5
SMAPE	ARIMA	0,80	0,93	1,05	1,14	1,36
	RNN	1,15	1,42	1,76	1,88	2,17
	LSTM	1,18	1,42	1,62	1,83	2,00

Таблиця 6.

Результати прогнозування значень індексу NYSE на періоди від 1 до 5 тижнів

Оцінка	Модель	Період для передбачення (тижнів)				
		1	2	3	4	5
MAE	ARIMA	269,05	340,06	397,93	443,76	545,15
	RNN	616,85	550,21	698,98	1 130,79	772,85
	LSTM	624,07	631,97	625,98	676,17	723,99
MSE	ARIMA	121 450,04	182 665,75	246 220,20	319 398,63	432 959,75
	RNN	509 883,92	439 804,48	758 076,09	2 295 238,34	911 458,82
	LSTM	558 258,74	581 823,77	577 721,85	659 404,05	780 056,48
RMSE	ARIMA	348,50	427,39	496,21	565,15	658,00
	RNN	714,06	663,18	870,68	1 515,00	954,70
	LSTM	747,17	762,77	760,08	812,04	883,21
MAPE	ARIMA	0,02	0,02	0,03	0,03	0,04
	RNN	0,04	0,04	0,05	0,07	0,05
	LSTM	0,04	0,04	0,04	0,04	0,05
SMAPE	ARIMA	1,73	2,19	2,57	2,85	3,49
	RNN	3,96	3,49	4,42	6,98	4,91
	LSTM	3,99	4,05	4,00	4,31	4,59

Таблиця 7.

Результати прогнозування значень індексу NYSE на періоди від 1 до 5 місяців

Оцінка	Модель	Період для передбачення (місяців)				
		1	2	3	4	5
MAE	ARIMA	680,29	827,86	861,69	1 086,58	723,17
	RNN	927,48	1 486,79	1 784,90	597,47	965,45
	LSTM	1 654,81	1 451,54	1 119,73	1 083,44	1 091,72

Таблиця 7.

(продовження)

Оцінка	Модель	Період для передбачення (місяців)				
		1	2	3	4	5
MSE	ARIMA	592 791,43	948 740,46	1 019 431,51	1 567 102,31	851 740,05
	RNN	1 253 302,99	3 005 624,41	4 327 505,02	589 121,29	1 480 884,33
	LSTM	3 777 278,85	2 794 125,21	1 640 286,95	1 600 034,89	1 620 041,82
RMSE	ARIMA	769,93	974,03	1 009,67	1 251,84	922,90
	RNN	1 119,51	1 733,67	2 080,27	767,54	1 216,92
	LSTM	1 943,52	1 671,56	1 280,74	1 264,92	1 272,81
MAPE	ARIMA	0,04	0,05	0,05	0,07	0,05
	RNN	0,06	0,09	0,11	0,04	0,06
	LSTM	0,11	0,09	0,07	0,07	0,07
SMAPE	ARIMA	4,37	5,31	5,48	7,05	4,60
	RNN	5,92	9,70	11,79	3,82	6,07
	LSTM	10,11	8,94	7,01	6,79	6,82

Таблиця 8.

Результати прогнозування значень індексу NYSE на періоди від 1 до 5 кварталів

Оцінка	Модель	Період для передбачення (кварталів)				
		1	2	3	4	5
MAE	ARIMA	1 016,15	946,46	1 744,53	1 385,20	1 783,48
	RNN	816,61	1 052,74	1 079,14	1 897,46	1 877,05
	LSTM	1 243,21	899,85	747,59	873,71	708,20
MSE	ARIMA	1 418 893,52	1 336 966,10	4 094 094,37	3 105 021,95	4 435 071,43
	RNN	1 027 405,25	2 134 336,03	2 023 368,93	5 617 710,30	5 601 065,05
	LSTM	1 963 602,66	1 204 332,45	974 515,90	1 196 873,25	914 841,38
RMSE	ARIMA	1 191,17	1 156,27	2 023,39	1 762,11	2 105,96
	RNN	1 013,61	1 460,94	1 422,45	2 370,17	2 366,66
	LSTM	1 401,29	1 097,42	987,18	1 094,02	956,47
MAPE	ARIMA	0,06	0,06	0,11	0,09	0,11
	RNN	0,05	0,07	0,07	0,12	0,12
	LSTM	0,08	0,06	0,05	0,06	0,05
SMAPE	ARIMA	6,47	5,94	11,25	8,65	11,09
	RNN	5,17	6,63	6,84	12,00	11,53
	LSTM	7,75	5,66	4,74	5,49	4,47

Таблиця 9.

Результати прогнозування значень індексу SBIPOP на періоди від 1 до 5 днів

Оцінка	Модель	Період для передбачення (днів)				
		1	2	3	4	5
MAE	ARIMA	6,75	8,37	10,70	11,32	12,40
	RNN	13,33	16,89	18,20	18,17	26,98
	LSTM	13,18	15,33	17,63	20,48	21,86
MSE	ARIMA	106,70	165,25	274,57	280,69	366,44
	RNN	378,80	571,67	729,17	748,65	1 478,87
	LSTM	389,02	534,93	721,76	883,63	1 057,29
RMSE	ARIMA	10,33	12,86	16,57	16,75	19,14
	RNN	19,46	23,91	27,00	27,36	38,46
	LSTM	19,72	23,13	26,87	29,73	32,52
MAPE	ARIMA	0,01	0,01	0,01	0,01	0,01
	RNN	0,01	0,01	0,02	0,02	0,02
	LSTM	0,01	0,01	0,02	0,02	0,02
SMAPE	ARIMA	0,58	0,72	0,92	0,97	1,06
	RNN	1,14	1,44	1,57	1,56	2,29
	LSTM	1,14	1,32	1,52	1,74	1,88

Таблиця 10.

Результати прогнозування значень індексу SBIPOP на періоди від 1 до 5 тижнів

Оцінка	Модель	Період для передбачення (тижнів)				
		1	2	3	4	5
MAE	ARIMA	17,34	21,63	27,17	26,51	34,31
	RNN	42,97	101,94	114,65	116,93	123,50
	LSTM	63,88	52,47	54,67	65,19	71,33
MSE	ARIMA	592,61	1 000,68	1 319,40	1 298,12	1 926,73
	RNN	3 275,07	14 732,02	18 146,56	20 199,90	22 389,26
	LSTM	5 619,64	4 319,72	5 120,33	6 201,24	7 450,19
RMSE	ARIMA	24,34	31,63	36,32	36,03	43,89
	RNN	57,23	121,38	134,71	142,13	149,63
	LSTM	74,96	65,72	71,56	78,75	86,31
MAPE	ARIMA	0,01	0,02	0,02	0,02	0,03
	RNN	0,04	0,09	0,10	0,10	0,11
	LSTM	0,05	0,04	0,05	0,06	0,06

Таблиця 10.

(продовження)

Оцінка	Модель	Період для передбачення (тижнів)				
		1	2	3	4	5
SMAPE	ARIMA	1,50	1,86	2,35	2,26	2,93
	RNN	3,68	8,31	9,33	9,52	10,00
	LSTM	5,41	4,45	4,63	5,51	6,00

Таблиця 11

Результати прогнозування значень індексу SBIPOP на періоди від 1 до 5 місяців

Оцінка	Модель	Період для передбачення (місяців)				
		1	2	3	4	5
MAE	ARIMA	51,80	63,37	79,63	82,83	86,35
	RNN	219,78	118,66	121,64	97,42	131,35
	LSTM	412,18	467,28	282,91	297,01	267,51
MSE	ARIMA	4 959,51	6 469,30	8 944,51	9 700,59	11 775,33
	RNN	75 672,75	19 080,88	26 331,73	14 367,28	22 296,16
	LSTM	199 760,15	284 810,87	106 450,58	127 012,89	107 820,45
RMSE	ARIMA	70,42	80,43	94,58	98,49	108,51
	RNN	275,09	138,13	162,27	119,86	149,32
	LSTM	446,95	533,68	326,27	356,39	328,36
MAPE	ARIMA	0,04	0,05	0,07	0,07	0,07
	RNN	0,20	0,10	0,10	0,09	0,11
	LSTM	0,35	0,39	0,23	0,24	0,22
SMAPE	ARIMA	4,41	5,34	6,68	7,04	7,28
	RNN	16,99	10,17	10,62	8,10	10,78
	LSTM	28,85	31,13	20,48	21,13	19,23

Таблиця 12.

Результати прогнозування значень індексу SBIPOP на періоди від 1 до 5 кварталів

Оцінка	Модель	Період для передбачення (кварталів)				
		1	2	3	4	5
MAE	ARIMA	98,37	123,40	146,87	176,32	143,22
	RNN	194,41	93,38	160,49	151,05	126,65
	LSTM	94,56	76,28	77,13	84,35	96,66

Таблиця 12.  
(продовження)

Оцінка	Модель	Період для передбачення (кварталів)				
		1	2	3	4	5
MSE	ARIMA	12 616,07	18 508,27	31 547,87	37 203,71	31 737,20
	RNN	51 291,08	13 464,14	32 596,63	37 031,17	26 285,15
	LSTM	13 337,28	9 395,86	8 900,56	11 066,31	13 525,54
RMSE	ARIMA	112,32	136,05	177,62	192,88	178,15
	RNN	226,48	116,04	180,55	192,43	162,13
	LSTM	115,49	96,93	94,34	105,20	116,30
MAPE	ARIMA	0,08	0,10	0,13	0,15	0,13
	RNN	0,16	0,08	0,13	0,12	0,10
	LSTM	0,08	0,06	0,06	0,07	0,08
SMAPE	ARIMA	8,28	10,33	12,55	14,75	11,69
	RNN	17,34	8,12	14,00	13,12	10,44
	LSTM	8,02	6,56	6,68	7,34	8,39

Як можна побачити з наведених даних, для короткострокового прогнозування (на 1 – 5 днів або тижнів) доцільно використовувати модель ARIMA. У більшій кількості випадків, результати отримані з її допомогою є точнішими (тобто мають кращі (менше) значення метрик, що були використані для оцінки одержаних прогнозів).

Обрані нейронні мережі (RNN та LSTM) починають демонструвати результативність, починаючи з прогнозування на періоди в кілька місяців, а вже з прогнозування на декілька кварталів, прогнозування за допомогою LSTM виходить на перший план.

У кількісному вираженні, ці результати розподілені наступним образом:

1) Загалом, з 300 отриманих оцінок, 244 кращіх (або 81%) припадає на результати отримані за допомогою моделі ARIMA. На другому місці знаходиться прогнозування за допомогою LSTM мережі – 46 (15%) кращіх метрик. Та останнє місце з 10 кращіми результатами займає прогнозування за допомогою RNN мережі.

2) При прогнозування на періоди 1 – 5 днів та тижнів безумовним лідером є модель ARIMA. Вона надає кращі (точніші) результати у 100% розглянутих випадків.

3) У розглянутих випадках, при прогнозуванні на 1 – 5 місяців, лідство залишається за моделлю ARIMA – за результатами 70 (93%) проведених оцінок. І лише 5 найкращих оцінок (7%) отримано за допомогою RNN нейронної мережі.

4) При використанні більших періодів (у цьому випадку кварталів) на перше місце за якістю виходять прогнози, що розроблені за допомогою мережі LSTM – вони є більш точними у 46 оцінках з 75 проведених (або 61%). Тоді як ARIMA переходить на друге місце відповідно до 24 (32%) результатів. І прогнози RNN залишаються на третьому місці за тими самими значеннями – 5 (7%) випадків.

Звідси бачимо, що обрані моделі можна застосовувати для прогнозування значень фондових індексів. При цьому модель ARIMA найкраще підходить для короткострокових прогнозів. Тоді як RNN та LSTM потребують додаткового вивчення та налаштування.

### 3. ВИСНОВКИ

У рамках дослідження було наведено теоретично підґрунтя застосування статистичної моделі ARIMA та нейронних мереж RNN і LSTM у питанні прогнозування значень фондових індексів на основі історичних даних. Наведено механізм оцінки одержаних результатів із використанням комплексних метрик: середньої абсолютної помилки (MAE), середньоквадратичної помилки (MSE), кореневої середньоквадратичної помилки (RMSE), середньої абсолютної відсоткової помилки (MAPE) та симетричної середньої абсолютної відсоткової помилки (SMAPE).

Аналіз показав, що кожна з моделей, маючи свої переваги та обмеження, може бути використана для прогнозування значень фондових індексів виходячи із специфіки певного показника та вимог до прогнозування. Модель ARIMA ефективна в умовах стаціонарних часових рядів і може виявляти значущі автокореляційні зв'язки на короткі періоди. Показує гарні результати при здійсненні прогнозування на коротко- та середньострокові інтервали часу. Моделі RNN і LSTM, з їхньою здатністю зберігати інформацію про попередній стан, є особливо корисними для виявлення довготривалих залежностей у даних. Однак, вони вимагають значних обчислювальних ресурсів і схильні до перенавчання, особливо при обмеженому наборі даних. Стають ефективнішими за ARIMA на довших інтервалах (починаючи з кварталів).

Це дослідження демонструє, що вибір моделі для прогнозування динаміки показників фондового ринку повинен базуватися на ретельному аналізі структури даних і специфічних цілей прогнозування. Використання комплексних метрик дозволяє об'єктивно оцінювати і порівнювати результати, надаючи важливу інформацію для прийняття обґрунтованих інвестиційних рішень на основі аналітичних моделей.

Проведена робота може бути використана як база для подальшого дослідження механізмів прогнозування динаміки індексів фондового ринку та перевірки та налаштування зазначених моделей прогнозування.



#### 4. СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Wu, W.; Chen, W.; Liu, B. Using BP neural network to predict the rise and fall of the stock market. *Dalian Univ. Technol. J.* 2001, 9–15.
2. Ticknor, J.L. A Bayesian regularized artificial neural network for stock market forecasting. *Expert Syst. Appl.* 2013, 40, 5501–5506.
3. Zhang, D.H.; Lou, S. The application research of neural network and BP algorithm in stock price pattern classification and prediction. *Futur. Gener. Comp. Syst.* 2021, 115, 872–879.
4. Tay, F.E.H.; Cao, L.J. Application of support vector machines in financial time series forecasting. *Omega-Int. J. Manage. Sci.* 2001, 29, 309–317.
5. Ran, Y.; Jiang, H. Stock Prices Prediction based on Back Propagation Neural Network and Support Vector Regression. *J. Shanxi Univ.* 2018, 41, 1–14.
6. Kim, K.-j. Financial time series forecasting using support vector machines. *Neurocomputing* 2003, 55, 307–319.
7. Hinton, G.E.; Salakhutdinov, R.R. Reducing the Dimensionality of Data with Neural Networks. *Science* 2006, 313, 504–507.
8. Singh, R.; Srivastava, S. Stock prediction using deep learning. *Multimed. Tools Appl.* 2017, 76, 18569–18584.
9. Kraus, M.; Feuerriegel, S. Decision support from financial disclosures with deep neural networks and transfer learning. *Decis. Support Syst.* 2017, 104, 38–48.
10. Chong, E.; Han, C.; Park, F.C. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Syst. Appl.* 2017, 83, 187–205.
11. Cui, D. A Study on the Prediction of Stock Price based Deep Belief Networks. Master's Thesis, Huazhong University of Science & Technology, Wuhan, China, 2016.
12. Liu, Q. Short Term Stock Price Forecasting based on Fuzzy Deep Learning Network Algorithm. Master's Thesis, Harbin Institute of Technology, Harbin, China, 2016.

13. Li, X.M.; Yang, L.; Xue, F.Z.; Zhou, H.J. Time Series Prediction of Stock Price Using Deep Belief Networks with Intrinsic Plasticity. In Proceedings of the 29th Chinese Control and Decision Conference (CCDC), IEEE, Chongqing, China, 28–30 May 2017; pp. 1237–1242.
14. Tsantekidis, A.; Passalis, N.; Tefas, A.; Kannianen, J.; Gabbouj, M.; Iosifidis, A. Forecasting Stock Prices from the Limit Order Book using Convolutional Neural Networks. In Proceedings of the IEEE 19th Conference on Business Informatics CBI, Thessalonki, Greece, 24–26 July 2017; Volume 1, pp. 7–12.
15. Sim, H.S.; Kim, H.I.; Ahn, J.J. Is Deep Learning for Image Recognition Applicable to Stock Market Prediction? *Complexity* 2019, 10, 4324878.
16. Chen, W.; Jiang, M.R.; Zhang, W.G.; Chen, Z.S. A novel graph convolutional feature based convolutional neural network for stock trend prediction. *Inf. Sci.* 2021, 556, 67–94.
17. Luca, D.P.; Oleksandr, H. Artificial neural networks architectures for stock price prediction: Comparisons and applications. *Int. J. Circuits Syst. Signal Process* 2016, 10, 403–413.
18. Hsieh, T.J.; Hsiao, H.F.; Yeh, W.C. Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm. *Appl. Soft. Comput.* 2011, 11, 2510–2525.
19. Rather, A.M.; Agarvval, A.; Sastry, V.N. Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Syst. Appl.* 2015, 42, 3234–3241.
20. Qin, Y.S.; Song, D.J.; Cheng, H.F.; Cheng, W.; Jiang, G.F.; Cottrell, G.W. A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction. In Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI), Melbourne, Australia, 19–25 August 2017; pp. 2627–2633.
21. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* 1997, 9, 1735–1780.
22. Yang, Q.; Wang, C. A Study on Forecast of Global Stock Indices Based on Deep LSTM Neural Network. *Stat. Res.* 2019, 36, 65–77.

23. Lee, M.C. Using support vector machine with a hybrid feature selection method to the stock trend prediction. *Expert Syst. Appl.* 2009, 36, 10896–10904.
24. Nti, I.K.; Adekoya, A.F.; Weyori, B.A. A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction. *J. Big Data* 2021, 8, 17.
25. Sakhare, N.N.; Imambi, S.S. Performance analysis of regression-based machine learning techniques for prediction of stock market movement. *Int. J. Recent Technol. Eng.* 2019, 7, 655–662.
26. Hu, Z.; Zhao, Y.; Khushi, M. A Survey of Forex and Stock Price Prediction Using Deep Learning. *Appl. Syst. Innov.* 2021, 4, 9.
27. Hiransha, M.; Gopalakrishnan, E.A.; Menon, V.K.; Soman, K.P. NSE Stock Market Prediction Using Deep-Learning Models. *Procedia Comput. Sci.* 2018, 132, 1351–1362.
28. Patel, R.; Choudhary, V.; Saxena, D.; Singh, A.K. Review of Stock Prediction using machine learning techniques. In *Proceedings of the 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, 3–5 June 2021; pp. 840–847.
29. Kamath, U.; Liu, J.; Whitaker, J. *Deep Learning for NLP and Speech Recognition*; Springer: Cham, Switzerland, 2019; pp. 575–613.
30. Manolakis, D.; Bosowski, N.; Ingle, V.K. Count Time-Series Analysis: A Signal Processing Perspective. *IEEE Signal Process. Mag.* 2019, 36, 64–81.
31. Kabbani, T.; Duman, E. Deep Reinforcement Learning Approach for Trading Automation in the Stock Market. *IEEE Access* 2022, 10, 93564–93574.
32. Moghar, A.; Hamiche, M. Stock Market Prediction Using LSTM Recurrent Neural Network. *Procedia Comput. Sci.* 2020, 170, 1168–1173.
33. Ren, Y.; Liao, F.; Gong, Y. Impact of News on the Trend of Stock Price Change: An Analysis based on the Deep Bidirectional LSTM Model. *Procedia Comput. Sci.* 2020, 174, 128–140.
34. Jin, Z.; Yang, Y.; Liu, Y. Stock closing price prediction based on sentiment analysis and LSTM. *Neural Comput. Appl.* 2019, 32, 9713–9729.

35. Parray, I.R.; Khurana, S.S.; Kumar, M.; Altalbe, A.A. Time series data analysis of stock price movement using machine learning techniques. *Soft Comput.* 2020, 24, 16509–16517.

## 5. ДОДАТКИ

### Додаток А

Версії програмного забезпечення, що використані при розробці моделей

- Python версія: 3.11.3
- pip версія: 24.0
- keras версія: 3.2.1
- matplotlib версія: 3.7.1
- numpy версія: 1.24.3
- pandas версія: 2.0.1
- pmdarima версія: 2.0.4
- sklearn версія: 0.0.post4
- scikit-learn версія: 1.4.2

## Додаток Б

### Приклад коду для побудови моделі ARIMA

```
import numpy as np
import pandas as pd
import warnings
import copy

from pmdarima import auto_arima

from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import root_mean_squared_error
from sklearn.metrics import mean_absolute_percentage_error

from pmdarima.metrics import smape

def arma_model(train, test, pred_num):

    AUTO_MOD = auto_arima(train["Price"], seasonal = False, trace = False)

    def forecast_in_steps(MODEL, num_periods, print_data = False):
        fc = []
        conf_int = []
        fc, conf_int = MODEL.predict(n_periods = num_periods,
                                     return_conf_int = True)
        return (fc.tolist(), np.asarray(conf_int).tolist())

    ERRORS_ALL = []

    forecasts = []
    confidence_intervals = []
    TEST_LIST = test["Price"]

    MODEL_4LOOP = copy.deepcopy(AUTO_MOD)
    while len(TEST_LIST) > 0:
        new_ob = TEST_LIST[0: pred_num]
        TEST_LIST = TEST_LIST[pred_num:]
        fc = []
        conf = []
        fc, conf = forecast_in_steps(MODEL_4LOOP, pred_num)
        forecasts = forecasts + fc
        confidence_intervals = confidence_intervals + conf
        MODEL_4LOOP.update(new_ob)

    FCOBS_DF = pd.DataFrame()
    FCOBS_DF["Forecast"] = forecasts[0: len(test["Date"])]
    FCOBS_DF["Date"] = test["Date"]
    FCOBS_DF = FCOBS_DF.set_index(FCOBS_DF["Date"])

    MAE = mean_absolute_error(test["Price"], FCOBS_DF["Forecast"])
    MSE = mean_squared_error(test["Price"], FCOBS_DF["Forecast"])
    RMSE = root_mean_squared_error(test["Price"], FCOBS_DF["Forecast"])
    MAPE = mean_absolute_percentage_error(test["Price"], FCOBS_DF["Forecast"])
    SMAPE = smape(test['Price'], FCOBS_DF['Forecast'])

    MAE_LIST = [pred_num, "MAE", MAE]
    MSE_LIST = [pred_num, "MSE", MSE]
```

```
RMSE_LIST = [pred_num, "RMSE", RMSE]
MAPE_LIST = [pred_num, "MAPE", MAPE]
SMAPE_LIST = [pred_num, "SMAPE", SMAPE]

ERRORS_ALL.append(MAE_LIST)
ERRORS_ALL.append(MSE_LIST)
ERRORS_ALL.append(RMSE_LIST)
ERRORS_ALL.append(MAPE_LIST)
ERRORS_ALL.append(SMAPE_LIST)

return FCOBS_DF["Forecast"], ERRORS_ALL
```

## Додаток В

### Приклад коду для побудови моделі RNN

```
import numpy as np
import pandas as pd
import warnings

from keras.models import Sequential
from keras.layers import Dense
from keras.layers import SimpleRNN
from keras.layers import Dropout
from keras.utils import set_random_seed

from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error
from sklearn.metrics import root_mean_squared_error
from sklearn.metrics import mean_absolute_percentage_error

from pmdarima.metrics import smape

def rnn_model(data, train, pred_num, times = 1, rd_seed = np.nan):

    warnings.filterwarnings("ignore")

    if not np.isnan(rd_seed):
        set_random_seed(rd_seed)

    ERRORS_ALL = []

    VALID_DATA = data[len(train) - (times + (pred_num - 1)):]
    TRAIN_VALUES = train["Price"].values
    TRAIN_VALUES = np.reshape(TRAIN_VALUES, (-1, 1))

    scaler = MinMaxScaler(feature_range = (0, 1))
    TRAIN_VALUES_SCALED = scaler.fit_transform(TRAIN_VALUES)

    X_TRAIN = []
    Y_TRAIN = []
    for ITER in range(times, len(train) - (pred_num - 1)):
        PRED = ITER + pred_num - 1
        X_TRAIN.append(TRAIN_VALUES_SCALED[ITER - times: ITER, 0])
        Y_TRAIN.append(TRAIN_VALUES_SCALED[PRED, 0])

    X_TRAIN, Y_TRAIN = np.array(X_TRAIN), np.array(Y_TRAIN)
    X_TRAIN = np.reshape(X_TRAIN, (X_TRAIN.shape[0], X_TRAIN.shape[1], 1))
    Y_TRAIN = np.reshape(Y_TRAIN, (Y_TRAIN.shape[0], 1))

    REGRESSOR = Sequential()

    REGRESSOR.add(SimpleRNN(units = 50,
                            activation = "tanh",
                            return_sequences = True,
                            input_shape = (X_TRAIN.shape[1], 1)))
    REGRESSOR.add(Dropout(0.2))

    REGRESSOR.add(SimpleRNN(units = 50,
                            activation = "tanh",
                            return_sequences = True))
```



```

REGRESSOR.add(Dropout(0.2))

REGRESSOR.add(SimpleRNN(units = 50,
                        activation = "tanh",
                        return_sequences = True))
REGRESSOR.add(Dropout(0.2))

REGRESSOR.add(SimpleRNN(units = 50))
REGRESSOR.add(Dropout(0.2))

REGRESSOR.add(Dense(units = 1))

REGRESSOR.compile(optimizer = "adam",
                  loss = "mean_squared_error",
                  metrics = ["accuracy"])

RNN_MODEL = REGRESSOR.fit(X_TRAIN, Y_TRAIN, epochs = 50, batch_size = 32)

Y_PRED = REGRESSOR.predict(X_TRAIN)
Y_PRED = scaler.inverse_transform(Y_PRED)
Y_TRAIN = scaler.inverse_transform(Y_TRAIN)

VALID_VALUES = VALID_DATA["Price"].values
VALID_VALUES = np.reshape(VALID_VALUES, (-1, 1))
VALID_VALUES_SCALED = scaler.fit_transform(VALID_VALUES)

X_TEST = []
Y_TEST = []

for ITER in range(times, len(VALID_DATA) - (pred_num - 1)):
    PRED = ITER + STEP - 1
    X_TEST.append(VALID_VALUES_SCALED[ITER - times: ITER, 0])
    Y_TEST.append(VALID_VALUES_SCALED[PRED, 0])

X_TEST, Y_TEST = np.array(X_TEST), np.array(Y_TEST)
X_TEST = np.reshape(X_TEST, (X_TEST.shape[0], X_TEST.shape[1], 1))
Y_TEST = np.reshape(Y_TEST, (-1, 1))
Y_PRED_OF_TEST = REGRESSOR.predict(X_TEST)
Y_PRED_OF_TEST = scaler.inverse_transform(Y_PRED_OF_TEST)

VALID_2CHECK = VALID_DATA["Price"].iloc[times + (pred_num - 1):]

MAE = mean_absolute_error(VALID_2CHECK, Y_PRED_OF_TEST)
MSE = mean_squared_error(VALID_2CHECK, Y_PRED_OF_TEST)
RMSE = root_mean_squared_error(VALID_2CHECK, Y_PRED_OF_TEST)
MAPE = mean_absolute_percentage_error(VALID_2CHECK, Y_PRED_OF_TEST)
SMAPE = smape(VALID_2CHECK, Y_PRED_OF_TEST)

MAE_LIST = [STEP, "MAE", MAE]
MSE_LIST = [STEP, "MSE", MSE]
RMSE_LIST = [STEP, "RMSE", RMSE]
MAPE_LIST = [STEP, "MAPE", MAPE]
SMAPE_LIST = [STEP, "SMAPE", SMAPE]

ERRORS_ALL.append(MAE_LIST)
ERRORS_ALL.append(MSE_LIST)
ERRORS_ALL.append(RMSE_LIST)
ERRORS_ALL.append(MAPE_LIST)
ERRORS_ALL.append(SMAPE_LIST)

return Y_PRED_OF_TEST, ERRORS_ALL

```

## АНОТАЦІЯ

Дипломна робота складається зі вступу, основної частини, висновка, списку використаних джерел та додатків. Обсяг роботи становить 57 сторінок. Основна частина займає 40 сторінок та містить 12 рисунків і 12 таблиць. Документ містить 35 посилань на використані джерела.

Дане дослідження присвячено вивченню методів оцінювання та прогнозування в дослідженні процесів на фондовому ринку. Метою дослідження є розробка, валідація і перевірка ефективності моделей ARIMA, RNN і LSTM для прогнозування рухів трьох міжнародних фондових індексів: BUMIX (Budapest Stock Index), NYSE (New York Stock Exchange Index) і Blue-Chip SBITOP (Slovenia Stock Index).

Результатом дослідження є оцінка та порівняння створених моделей прогнозування значень фондових індексів на періоди від 1 до 5 днів, тижнів, місяців та кварталів.

Приводиться достатня кількість теоретичної інформації стосовно обраних методів проведення дослідження, а також описуються підходи до оцінки результатів таких методів.

Ключові слова: фондовий ринок, передбачення, ARIMA, RNN, LSTM.

## ABSTRACT

The thesis consists of an introduction, the main part, a conclusion, a list of used sources and appendices. The volume of work is 57 pages. The main part occupies 40 pages and contains 12 figures and 12 tables. The document contains 35 references to used sources.

This study is devoted to the study of assessment and forecasting methods in the study of processes on the stock market. The purpose of the study is to develop, validate and verify the effectiveness of ARIMA, RNN and LSTM models for forecasting the movements of three international stock indices: BUMIX (Budapest Stock Index), NYSE (New York Stock Exchange Index) and Blue-Chip SBITOP (Slovenia Stock Index).

The result of the study is an assessment and comparison of the created models for forecasting the values of stock indices for periods from 1 to 5 days, weeks, months and quarters.

A sufficient amount of theoretical information regarding the chosen methods of conducting the research is provided, as well as the approaches to evaluating the results of such methods are described.

Keywords: stock market, prediction, ARIMA, RNN, LSTM.