

Вінницький національний технічний університет  
(повне найменування вищого навчального закладу)

Факультет інтелектуальних інформаційних технологій та автоматизації  
(повне найменування інституту, назва факультету (відділення))

Кафедра комп'ютерних наук  
(повна назва кафедри (предметної, циклової комісії))

## МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

на тему:

«Інформаційна технологія прогнозування курсу криптовалюти  
ETHEREUM»

Виконав: студент 2-го курсу,  
групи ЗКН-22м спеціальності 122

«Комп'ютерні науки»

(шифр і назва напрямку підготовки, спеціальності)

Сміленко В. О.  
(прізвище та ініціали)

Керівник: к.т.н., доц. кафедри КН

Паночишин Ю.М.

(прізвище та ініціали)

« 07 » 12 2023 р.

Опонент: к.т.н., проф. кафедри КСУ

Биков М.М.

(прізвище та ініціали)

« 07 » 12 2023 р.

**Допущено до захисту**

Завідувач кафедри КН

д.т.н., проф. Яровий А. А.

(прізвище та ініціали)

« 08 » 12 2023 р.

Вінниця ВНТУ - 2023 рік

Вінницький національний технічний університет  
Факультет інтелектуальних інформаційних технологій та автоматизації  
Кафедра комп'ютерних наук  
Рівень вищої освіти II-й (магістерський)  
Галузь знань – 12 Інформаційні технології  
Спеціальність – 122 Комп'ютерні науки  
Освітньо-професійна програма – Системи штучного інтелекту

**ЗАТВЕРДЖУЮ**  
Завідувач кафедри КН  
Д.т.н., проф. Яровий А.А.

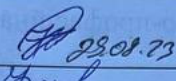
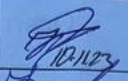
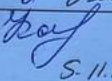
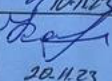
29.08 2023 року

### **ЗАВДАННЯ** **НА МАГІСТЕРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ**

Саміленку Вадиму Олександровичу  
(прізвище, ім'я, по батькові)

1. Тема роботи «Інформаційна технологія прогнозування курсу криптовалюти ETHEREUM»  
керівник роботи к.т.н., доц. кафедри КН, Паночишин Ю.М.  
затверджені наказом вищого навчального закладу від «28» 09 2023 року № 242
2. Строк подання студентом роботи 13.11. 2023 року
3. Вихідні дані до роботи:  
вихідні дані – Датасет Kaggle «Crypto - ETH : Advanced Analysis » з набором даних про курс Ethereum на світовому ринку за період з 01.01.2022 по 01.01.2023., мова програмування – об'єктно-орієнтована.
4. Зміст текстової частини:  
Вступ, аналіз предметної області прогнозування курсів валют, ідентифікація та вибір оптимальної моделі для прогнозування курсу криптовалюти Ethereum, програмна реалізація інформаційної технології прогнозування курсу Ethereum на основі машинного навчання, розроблення інформаційної технології аналізу та прогнозування курсу Ethereum у 2023 році, економічна частина, висновки, перелік використаних джерел, додатки
5. Перелік ілюстративного матеріалу (з точним зазначенням обов'язкових креслень)  
Алгоритм роботи програми прогнозування курсу валюти Ethereum, UML діаграма класів програми прогнозування курсу криптовалют, робочі вікна програми, результати тестування програми прогнозування курсу криптовалюти Ethereum.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціалита посада консультанта	Підпис, дата	
		завдання видав	виконан прийня
1-3 Спеціальна частина	к.т.н., доц. кафедри КН. Паночишин Ю.М.	 29.08.23	 29.11.23
4 Економічна частина	к.е.н., доц. каф. ЕПВМ Кавецький В.В.,	 5.11.23	 20.11.23

7. Дата видачі завдання 29.08 2023 року

КАЛЕНДАРНИЙ ПЛАН


№ з/п	Назва етапів магістерської кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Загальна характеристика поставленої задачі	01.09.23 - 08.09.23	Аналітичний огляд літературних джерел, досліджень, р 3 1
2	Вибір оптимальних інформаційних технологій	09.09.23 - 19.09.23	Розділ 2
3	Технології розробки	20.09.23 - 10.10.23	Розділ 2
4	Тренування моделей машинного навчання	11.10.23 - 28.10.23	Розділ 3
5	Оформлення матеріалів до захисту МКР	29.10.23 - 10.11.23	Розділ 4

Студент

  
(підпис)

Саміленко В. О.

Керівник роботи

  
(підпис)

Паночишин Ю.М.

## АНОТАЦІЯ

УДК 004.623

Саміленко В. О. Інформаційна технологія прогнозування курсу криптовалюти ETHEREUM. Магістерська кваліфікаційна робота зі спеціальності "Магістерська кваліфікаційна робота зі спеціальності 122 «Комп'ютерні науки», освітня програма «Системи штучного інтелекту» ". Вінниця: ВНТУ, 2023. 117 с.

Укр. мовою. Бібліогр.: 32 назв; рис.: 53; табл. 13

Дана магістерська кваліфікаційна робота присвячена вибору оптимальної моделі для прогнозування курсу криптовалюти Ethereum. На основі аналізу різних метрик, в результаті було визначено, яка модель є найбільш точною для задачі прогнозування курсу Ethereum.

Метою роботи є підвищення точності прогнозування курсу криптовалюти Ethereum за рахунок застосування методів машинного навчання.

У підсумку був сформований прогноз Ethereum.

Графічна частина складається з 5 плакатів.

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою «Інформаційна технологія прогнозування курсу криптовалюти ETHEREUM» становить 42,7 бала, що, свідчить про комерційну важливість проведення даних досліджень (рівень комерційного потенціалу розробки високий).

При оцінюванні за технічними параметрами, згідно узагальненого коефіцієнту якості розробки, науково-технічна розробка переважає існуючі аналоги приблизно в 2,80 рази.

Також термін окупності становить 2,03 р., що менше 3-х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

Ключові слова: криптовалюта, нейронна мережа, Ethereum.

## ABSTRACT

Samilenko V. O. Information technology for forecasting the ETHEREUM cryptocurrency exchange rate. "Master's thesis in the specialty 122 «Computer sciences», educational program «Artificial intelligence systems» ". Vinnytsia: VNTU, 2023. 117 p.

Ukraine language Bibliography: 32 titles; Fig.: 53; table 13

This master's thesis is devoted to the selection of the optimal model for forecasting the course of the Ethereum cryptocurrency. Based on the analysis of various metrics, as a result, it was determined which model is the most accurate for the task of forecasting the Ethereum exchange rate.

The purpose of the work is to increase the accuracy of forecasting the course of the Ethereum cryptocurrency through the use of machine learning methods.

As a result, the Ethereum forecast was formed.

The graphic part consists of 5 posters.

According to the conducted studies, the level of commercial development potential on the topic "Information technology for forecasting the ETHEREUM cryptocurrency rate" is 42.7 points, which indicates the commercial importance of conducting these studies (the level of commercial development potential is high).

When evaluated by technical parameters, according to the generalized coefficient of development quality, scientific and technical development outweighs existing analogues by approximately 2.80 times.

Also, the payback period is 2.03 years, which is less than 3 years, which indicates the commercial attractiveness of the scientific and technical development and may encourage a potential investor to finance the implementation of this development and its introduction to the market.

Keywords: cryptocurrency, neural network, Ethereum.

## ЗМІСТ

ВСТУП .....	5
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ПРОГНОЗУВАННЯ КУРСІВ ВАЛЮТ 7	7
1.1 Постановка задачі прогнозування цифрових валют .....	7
1.2 Огляд відомих методів прогнозування курсу валют .....	13
1.3 Обґрунтування вибору аналогу програми прогнозування курсу криптовалют.....	15
1.4 Висновок до розділу 1 .....	22
2 РОЗРОБКА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПРОГНОЗУВАННЯ КУРСУ КРИПТОВАЛЮТИ ETHEREUM .....	23
2.1 Обґрунтування вибору моделі машинного навчання.....	23
2.2 Синтез ознак на основі ознаки як часового ряду .....	25
2.3 Аналіз параметрів курсу криптовалюти як часового ряду .....	30
2.4 Аналіз аномалій ринку криптовалют .....	35
2.5 Висновок до розділу 2 .....	39
3 ПРОГРАМНА РЕАЛІЗАЦІЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПРОГНОЗУВАННЯ КУРСУ ETHEREUM НА ОСНОВІ МАШИННОГО НАВЧАННЯ.....	40
3.1 Обґрунтування вибору мови та середовища програмування для прогнозування курсу криптовалют. ....	40
3.2 Вибір спеціалізованої бібліотеки для програмної реалізації прогнозування курсів криптовалют.....	41
3.3 Прогнозування криптовалюти як часового ряду з використанням Facebook Prophet.....	45
3.4 Прогнозування цифрової валюти як часового ряду з використанням ARIMA.....	47
3.5 Прогнозування криптовалюти методами машинного навчання .....	51
3.6 Вибір оптимальної моделі .....	54
3.7 Висновок до розділу 3 .....	57

4	ТЕСТУВАННЯ ТА АНАЛІЗ РЕЗУЛЬТАТІВ РОБОТИ ПРОГРАМИ ПРОГНОЗУВАННЯ КУРСУ ETHEREUM У 2023 РОЦІ .....	58
4.1	Розроблення інформаційної технології для подальшого прогнозування..	58
4.2	Реалізація нейронної мережі .....	60
4.3	Тестування та аналіз результатів роботи програми прогнозування курсу криптовалюти ETHEREUM.....	63
4.5	Висновок до розділу 4 .....	72
5	ЕКОНОМІЧНА ЧАСТИНА .....	74
5.1	Проведення комерційного та технологічного аудиту науково-технічної розробки .....	74
5.2	Розрахунок узагальненого коефіцієнта якості розробки .....	78
5.3	Розрахунок витрат на проведення науково-дослідної роботи.....	80
5.3.1	Витрати на оплату праці.....	80
5.3.2	Відрахування на соціальні заходи .....	83
5.3.3	Сировина та матеріали.....	83
5.3.4	Розрахунок витрат на комплектуючі.....	85
5.3.5	Спецустаткування для наукових (експериментальних) робіт .....	85
5.3.6	Програмне забезпечення для наукових (експериментальних) робіт ..	86
5.3.7	Амортизація обладнання, програмних засобів та приміщень .....	87
5.3.8	Паливо та енергія для науково-виробничих цілей .....	88
5.3.9	Службові відрядження.....	89
5.3.10	Витрати на роботи, які виконують сторонні підприємства, установи і організації.....	90
5.3.11	Інші витрати.....	90
5.3.12	Накладні (загальновиробничі) витрати.....	91
5.4	Розрахунок економічної ефективності науково-технічної розробки при її можливій комерціалізації потенційним інвестором .....	92
5.5	Висновок до розділу 5 .....	96

	4
ВИСНОВКИ.....	98
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	100
Додаток А (обов'язковий) Протокол перевірки кваліфікаційної роботи на наявність текстових запозичень.....	105
Додаток Б (обов'язковий) Лістинг програми .....	106
Додаток В (обов'язковий) Ілюстративна частина.....	111



## ВСТУП

**Актуальність теми.** Розвиток ІТ-технологій надає значний вплив на різні сфери життя, у тому числі і на фінансові установи. Криптовалюти, як нові фінансові інструменти, швидко розвиваються та починають впроваджуватися в різних країнах як альтернатива традиційним грошам. Фінансові установи активно використовують можливості ІТ для удосконалення своєї діяльності та створення нових фінансових і економічних інструментів для взаємодії між людьми.

**Зв'язок роботи з науковими програмами, планами, темами.** Магістерська кваліфікаційна робота виконана відповідно до напрямку наукових досліджень кафедри комп'ютерних наук Вінницького національного технічного університету 22 К1 «Моделі, методи, технології та пристрої інтелектуальних інформаційних систем управління, економіки, навчання та комунікацій» та плану наукової та навчально-методичної роботи кафедри.

**Мета та завдання досліджень.** Метою роботи є підвищення точності прогнозування курсу криптовалюти Ethereum за рахунок використання методів машинного навчання, Facebook Prophet.

Для реалізації цієї мети потрібно розв'язати наступні задачі:

- виконати аналіз проблеми прогнозування курсу криптовалюти Ethereum;
- здійснити ідентифікацію та обрати оптимальну модель для прогнозування курсу криптовалюти Ethereum;
- розробити інформаційну технологію аналізу та прогнозування курсу криптовалюти Ethereum;

**Об'єктом дослідження** є процес прогнозування курсу криптовалюти Ethereum на криптовалютному ринку.

**Предметом дослідження** є інформаційна технологія та програмні засоби прогнозування курсу Ethereum у 2023 році та точність їх роботи.

**Методи дослідження.** У роботі використані такі методи наукових досліджень: Random Forest Regressor, Linear Regression, KNeighbors Regressor,

Support vector Machines, Linear SVR, Bagging Regressor, XGB Regressor, MLP Regressor, ARIMA, Prophet, SHAP, Sklearn.

### **Наукова новизна одержаних результатів.**

1. Набула подальшого розвитку інформаційна технологія прогнозування курсу криптовалюти, яка відрізняється використанням моделі машинного навчання, що дозволило підвищити точність прогнозування курсу криптовалюти на криптовалютному ринку.

**Практичне значення** одержаних результатів полягає в тому, що на основі проведених досліджень розроблено програмне забезпечення прогнозування курсу криптовалюти.

**Достовірність теоретичних положень** магістерської кваліфікаційної роботи підтверджується коректністю постановки завдання, коректністю використання методу машинного навчання, експериментальними дослідженнями тестування програмної реалізації інформаційної технології прогнозування курсу криптовалюти.

**Особистий внесок здобувача.** Усі результати, що наведені у магістерській кваліфікаційній роботі, отримані самостійно. У працях, які написано у співавторстві, здобувачу належать: аналіз процесу прогнозування курсу криптовалюти Ethereum у 2023 році [1].

**Апробація результатів роботи.** Результати досліджень апробовані на конференції «Молодь в науці: дослідження, проблеми, перспективи-2024», Вінниця, жовтень 2023 – червень 2024 року [1].

**Публікації.** За результатами досліджень опубліковано одні тези доповіді на науково-технічній конференції [1].

# 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ПРОГНОЗУВАННЯ КУРСІВ ВАЛЮТ

## 1.1 Постановка задачі прогнозування цифрових валют

В сучасних умовах розвитку фінансового ринку, особливо у сфері цифрових валют, прогнозування курсів стає ключовим елементом для розуміння та управління ризиками інвестицій. Постановка задачі прогнозування цифрових валют включає в себе ряд важливих аспектів та завдань, які можна розглядати в рамках дослідження.

Криптовалюта - це цифрова валюта, яка використовує криптографію для забезпечення безпеки та перевірки транзакцій. Він не випускається жодною центральною владою, а його випуск і облік виконуються децентралізованою платіжною системою.

Однією з ключових переваг криптовалют є їх децентралізація. Традиційні валюти випускаються та регулюються центральними банками, які можуть використовувати свою владу для маніпулювання економікою. Криптовалюти, з іншого боку, не підконтрольні жодній центральній владі. Це може призвести до підвищення безпеки та прозорості, оскільки немає жодних централізованих систем, які можна було б зламати або піддатися корупції.

Криптовалюти використовують криптографію для захисту своїх транзакцій. Це означає, що вони можуть бути зашифровані таким чином, що їх може прочитати лише відправник і одержувач. Це може зробити їх менш вразливими до шахрайства та хакерських атак.

Криптовалюти можуть бути перераховані швидко та легко. Це може зробити їх привабливими для міжнародних платежів, які можуть зайняти кілька днів або навіть тижнів при використанні традиційних валют.

Однією з основних проблем криптовалют є їх висока волатильність. Ціни на криптовалюти можуть значно коливатися, що може зробити їх ризикованим

вкладенням. Наприклад, ціна Bitcoin за останні роки коливалася від кількох сотень до декількох тисяч доларів.

Криптовалюти є відносно новими, і їх правовий статус все ще не визначений у багатьох країнах. Це може призвести до юридичних проблем для людей, які використовують криптовалюти.

Видобуток криптовалюти може мати значний екологічний вплив. Видобуток Bitcoin, наприклад, вимагає використання потужних комп'ютерів, які споживають багато енергії. Це може призвести до підвищення викидів парникових газів.

Криптовалюти все ще перебувають на ранніх стадіях розвитку, і їх майбутнє ще не визначено. Однак вони мають потенціал змінити спосіб, яким ми здійснюємо платежі та зберігаємо цінність. Якщо криптовалюти зможуть подолати свої поточні проблеми, вони можуть стати широко прийнятими засобами обміну та зберігання цінності.

Криптовалюти - це складний і багатогранний об'єкт. Вони мають ряд потенційних переваг перед традиційними валютами, але також мають ряд потенційних недоліків. Майбутнє криптовалют ще не визначено, але вони мають потенціал змінити спосіб, яким ми здійснюємо платежі та зберігаємо цінність.

Найвідомішою криптовалютою є Bitcoin, яка була створена в 2009 році. З тих пір було створено тисячі інших криптовалют, кожна з яких має свої унікальні характеристики та особливості.

Криптовалюти - це відносно новий вид валюти, який використовує криптографію для забезпечення безпеки та перевірки транзакцій. Вони не випускаються жодною центральною владою, а їх випуск і облік виконуються децентралізованою платіжною системою.

Ціни на криптовалюти схильні до значної волатильності, що робить їх привабливими для інвесторів, які шукають високі прибутки. Однак висока волатильність також робить їх ризикованим вкладенням.

Прогнозування цін на криптовалюти є важливою задачею для інвесторів, які хочуть зменшити ризик своїх інвестицій. Це також важливо для бізнесів, які приймають криптовалюти як платіж.

Задача прогнозування цін на криптовалюти є складною, оскільки на ціни впливає безліч факторів, як внутрішніх, так і зовнішніх.

Найбільш ефективний метод прогнозування цін на криптовалюти залежить від конкретних обставин. Для інвесторів, які шукають точні прогнози, методи машинного навчання можуть бути найкращим вибором. Для бізнесів, які приймають криптовалюти як платіж, економетричні методи можуть бути більш придатними, оскільки вони краще розуміються та інтерпретуються.

Капіталізація ринку станом на 12.04.2023 р., де криптовалюта Ethereum займає друге місце (рис. 1.1) [2].

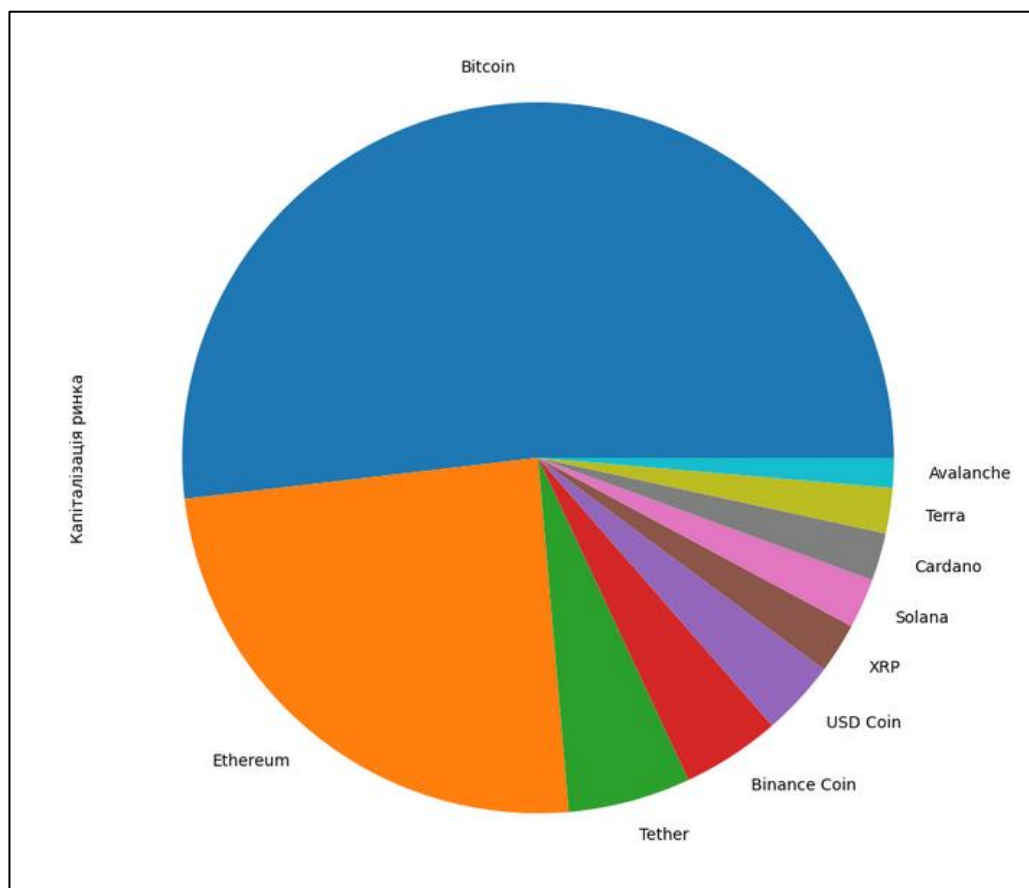


Рисунок 1.1 – Капіталізація ринку криптовалют за даними сайту Yahoo станом на 12.04.2023 р.

Ethereum, друга за величиною криптовалюта після біткойна, за останні кілька років привернула широку увагу і накопичила значні записи транзакцій. Однак основна структура мережі Ethereum все ще відносно не вивчена.

Розроблена інтелектуальна система прогнозування надає точні результати спрогнозованої вартості криптовалют. Подібна ефективність системи досягається завдяки виконанню поступових ітераційних етапів обробки даних. Основними етапами, які можливо виділити є наступні:

Етап збірки та обробки даних. Даний етап слугує початковою ланкою роботи системи. Існує багато ресурсів-агрегаторів, які надають можливість отримати бази даних з колекціями крипто валютних індексів. Саме на цих даних, система буде навчатись та прогнозувати у подальшому. Окрім цього, на даному етапі, з набору криптовалютних індексів, вилучаються усі некоректні значення.

Етап агрегації даних. Агрегування даних, потрібно для вилучення потрібних часових рядів з існуючого масиву даних. Завдяки розбиттю даних на необхідні часові ряди, система має можливість використання найбільш точної моделі прогнозування.

Етап навчання моделей. Виконується розбиття використаних у моделі даних на тренувальну та перевірочну вибірки. В подальшому, кожна з моделей використовує власний набір особливостей прогнозування, що дає різну точність прогнозування.

Етап прогнозування даних. Кінцевий етап роботи системи. Система робить вибір найбільш правдивої моделі прогнозування за допомогою виконання перехресної перевірки серед вже навчених моделей. Кожний з етапів хоч і є поступовим у виконанні, має можливість ітераційного застосування. На виконанні кожного з етапів, система має можливість повернутися до будь-якого з попередніх, для уточнення або зміни вже існуючого набору даних. Це дає можливість гнучкого навчання системи, що веде до високоточного прогнозування вартості криптовалют.

Також, обрано нечимало спроб покращити передбачуваність транзакцій в мережі Ethereum. Проведено докладний аналіз поведінки транзакцій, структури спільноти та передбачення посилянь (DANET) у мережі Ethereum. Дослідження охоплює аналіз змін у розподілі та накопиченні багатства в Ethereum Featured Transactional Network (EFTN), а також вивчення структури спільноти. Також, проводиться пошук моделі передбачуваності зв'язку на EFTN, використовуючи сучасні методи автоматичного кодування варіацій графіків. Результати експериментів ілюструють високу точність передбачення посилянь в мережах Ethereum.

Крім того, статистичні дані про використання мережі Ethereum представлені візуально та узагальнено шляхом експериментів, що дозволяє зробити припущення щодо поточного використання цієї технології та її майбутнього розвитку.

За останні десятиліття технологічні досягнення, такі як Інтернет речей (IoT), електронна комерція та цифрові платежі, значно вплинули на спосіб, яким ми взаємодіємо зі світом та обмінюємося грошима за різноманітні послуги. Зокрема, зростає популярність криптовалют, що є новим видом валюти, що використовується в цифровому середовищі. [5].

Курс криптовалюти Ethereum зазнає частих змін (рис. 1.2).

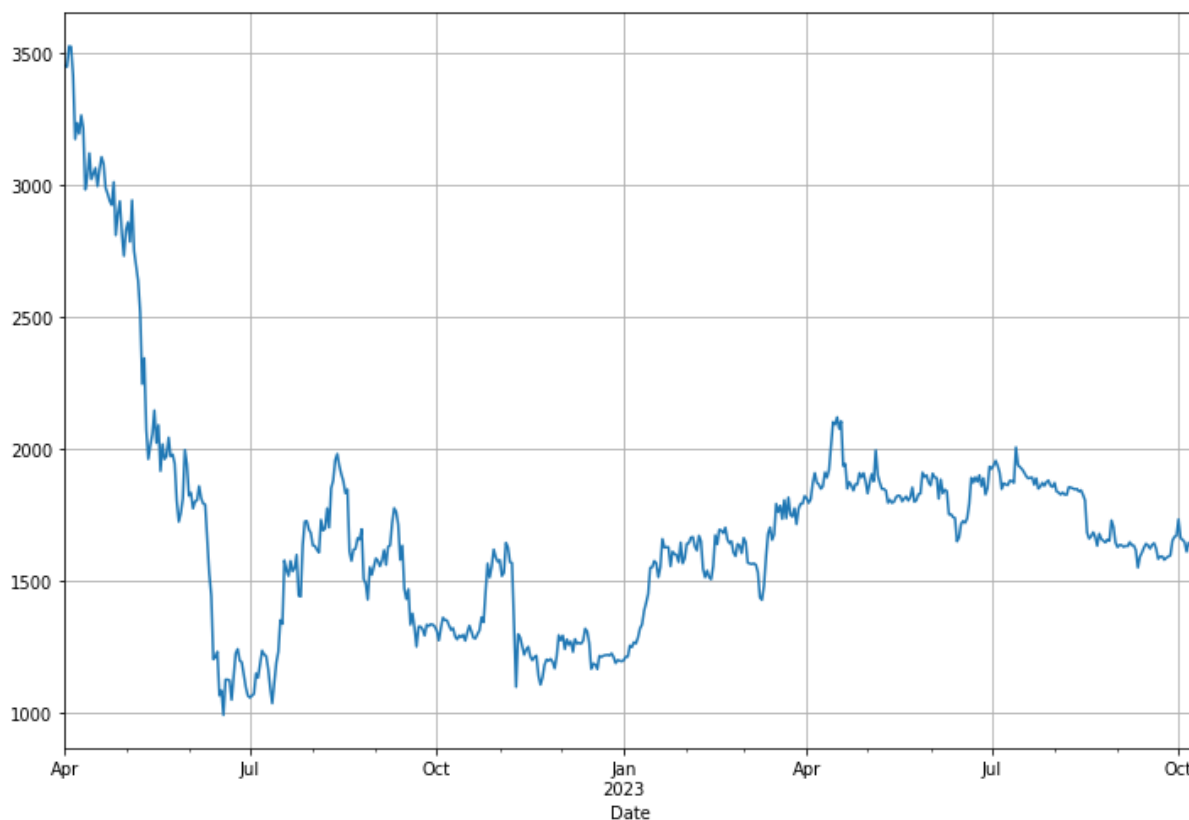


Рисунок 1.2 – Курс Ethereum за даними сайту Yahoo

Криптовалюти використовують технології блокчейнів і розподіленого реєстру, що унеможливорює будь-якому регулятору контролювати події в мережі та дозволяє їм відбуватися без обмежень на всій території користувачів.

Блокчейн можна описати як книгу транзакцій з цифровими активами, яка фіксує володіння, діяльність та час здійснення різних операцій. З технічного погляду, блокчейн представляє собою розподілену транзакційну базу даних, де глобально розподілені вузли зв'язані одноранговою мережею P2P для взаємозв'язку та виявлення однорангового зв'язку. P2P-мережі означають систему, де всі сторони є рівноправними учасниками, підключеними без централізованих посередників чи регулюючих органів. Це визначається як розподілена система, де окремі комп'ютери або вузли надають свої обчислювальні ресурси безпосередньо всім іншим учасникам мережі, утворюючи величезну глобальну електронну таблицю. [6].



## 1.2 Огляд відомих методів прогнозування курсу валют

Прогнозування курсу валют – це складне завдання, оскільки на нього впливають множина факторів, включаючи економічні, політичні та соціокультурні аспекти.

Для вирішення цієї задачі дослідники використовують різноманітні методи, починаючи від традиційних до більш сучасних. Нижче представлений огляд відомих методів прогнозування курсу валют:

### Основний аналіз:

Основний аналіз базується на вивченні фундаментальних факторів, які впливають на економіку країни, такі як стан господарства, інфляція, ставки, бюджетні та торговельні баланси.

Переваги: Дозволяє зрозуміти довгострокові тенденції та великі рухи на валютному ринку.

Недоліки: Може бути менше ефективним для короткострокового прогнозування, оскільки не враховує краткострокові волатильні фактори.

### Технічний аналіз:

Технічний аналіз використовує графіки, індикатори та патерни цін для прогнозування майбутніх рухів.

Переваги: Ефективний для короткострокового прогнозування, враховує психологічні аспекти ринку.

Недоліки: Не враховує фундаментальні фактори та може бути менше ефективним на ринках, що не підкоряються технічному аналізу.

### Моделі машинного навчання:

Використовують алгоритми машинного навчання для аналізу великих обсягів даних та виявлення закономірностей.

Переваги: Здатні адаптуватися до змінюючихся умов ринку, враховують велику кількість факторів.

Недоліки: Вимагають значних обчислювальних ресурсів, можуть виявитися недостатньо ефективними при недостатньому обсязі даних.

#### Економічні моделі:

Використовують статистичні моделі для аналізу та прогнозування курсів валют.

Переваги: Можуть бути ефективними для розуміння взаємозв'язків між економічними показниками та курсами валют.

Недоліки: Потребують чіткої специфікації економічних моделей, можуть недооцінювати нестандартні ситуації.

#### Сентимент-аналіз:

Вивчає громадську думку, новини та соціальні мережі для визначення ставлення ринку до валюти.

Переваги: Враховує "людський фактор" та може передбачити рухи цін відповідно до настроїв ринку.

Недоліки: Залежить від точності аналізу громадської думки та новин, може бути піддано впливу масових емоцій.

Обробка прогнозування курсів валют залишається складним завданням, що вимагає поєднання різних методів та постійного вдосконалення. Перевага полягає в поєднанні різноманітних підходів для отримання більш точних та надійних результатів в умовах постійної зміни на фінансових ринках.

Серед усіх методів ми визначили, що моделі машинного навчання можуть виявитися ефективними у прогнозуванні курсів валют з ряду причин:

Обробка великої кількості даних – машинне навчання здатне ефективно обробляти великі обсяги історичних та поточних даних, що важливо для прогнозування на фінансових ринках.

Врахування складних взаємодій – моделі машинного навчання здатні виявляти складні неявні взаємодії між різними факторами, які можуть впливати на курс валют.

Адаптація до змін – моделі машинного навчання можуть адаптуватися до змінюючихся умов ринку, що робить їх ефективними в умовах волатильності та непередбачуваності.

Розпізнавання складних патернів – алгоритми машинного навчання можуть розпізнавати складні патерни та тенденції в даних, які можуть залишитися непоміченими іншими методами.

Можливість використання різних типів даних – машинне навчання може працювати з різними типами даних, включаючи числові, текстові, категоріальні та інші, що робить його гнучким у використанні.

Оптимізація параметрів – можливість оптимізації параметрів дозволяє покращити точність та ефективність моделей машинного навчання під час їхнього тренування.

Здатність до автоматизації – машинне навчання може бути використано для автоматизації процесу прогнозування, зменшуючи витрати часу та зусиль.

Однак важливо пам'ятати, що моделі машинного навчання також мають свої обмеження і вимагають уважної підготовки даних, вибору відповідних алгоритмів та налагодження параметрів для досягнення оптимальних результатів. При використанні машинного навчання важливо враховувати контекст ринку та особливості фінансових інструментів, які аналізуються.

### **1.3 Обґрунтування вибору аналогу програми прогнозування курсу криптовалют**

В сучасному світі фінансів та інвестицій, точне прогнозування курсів криптовалют стає визначальним елементом прийняття обґрунтованих фінансових рішень. Однак у світі зростаючої складності ринку і динамічних змін, вибір ефективного інструменту для прогнозування стає ключовим завданням. В даному випадку обговоримо створену програму для прогнозування курсу криптовалют створену із допомогою Facebook Prophet та ARIMA, аргументуючи її вибір з різних позицій.

Гнучкість та Простота Використання: прогнозування курсів криптовалют часто стає завданням для різних категорій користувачів - від досвідчених аналітиків до новачків в області фінансів.

Автоматична Обробка Сезонності: у світі фінансів, особливо на валютних ринках, важливо враховувати сезонні та циклічні тенденції, що впливають на курси криптовалют. Автоматично визначає сезонність та свята, вбудовуючи ці впливи в модель прогнозування. Це особливо корисно для точного передбачення руху курсів криптовалют, які можуть підкорятися регулярним змінам у певні періоди.

Моделювання Погодженого Тренду: погоджений тренд важливий для адаптації до змін у довгострокових тенденціях. Дозволяє включати цей тренд у модель, що поліпшує точність прогнозів для криптовалют, які демонструють чіткі тренди з часом. Ця можливість аналізувати довгострокові тенденції сприяє більш точному прогнозуванню та забезпечує користувачів інформацією для ухвалення обдуманих фінансових рішень.

Обробка Відсутності Даних та Викидів: прогнозування курсів криптовалют часто стикається з пропущеними даними та аномаліями, що може впливати на точність результатів. Програма розроблена для ефективної обробки відсутності даних та викидів, що робить його стійким до аномалій у вхідних даних. Це важливо для підтримки точності прогнозів в реальних умовах фінансового ринку.

Широкі Можливості Налаштувань: програма надає користувачам можливість налаштування різних параметрів моделі, що дозволяє персоналізувати аналіз для конкретного ринкового середовища. Можливість налаштування робить створену програму гнучкою і придатною для різних умов ринку та стилів аналізу.

Розглянемо LSTM (Long Short-Term Memory) модель у Deep Learning як ефективний аналог для прогнозування курсу криптовалют.

Здатність до Аналізу Складних Залежностей: LSTM, що є типом рекурентних нейронних мереж, демонструє високу здатність до аналізу та моделювання складних нелінійних залежностей в часових рядах. Валютні курси часто характеризуються складними, неочікуваними змінами, і LSTM може ефективно адаптуватися до цих динамічних умов.

Обробка Довгострокових Залежностей: LSTM спеціально розроблені для ефективної обробки довгострокових залежностей в часових рядах. В умовах фінансового ринку, де існують тенденції та циклічні коливання, ця здатність є ключовою для точного прогнозування та адаптації до довгострокових тенденцій.

Адаптація до Змін в Інформації: LSTM може адаптуватися до змін в інформації, запам'ятовуючи та забуваючи інформацію за потреби. Це важливо в умовах, коли фінансові ринки можуть раптово змінюватися, і модель повинна швидко реагувати на нові умови.

Можливість Прогнозування з Різних Джерел Даних: LSTM може працювати з різними типами даних, включаючи числові, текстові та категоріальні, що робить його універсальним для прогнозування курсів криптовалют, які можуть бути впливовані різними чинниками.

Робастність до Шуму та Викидів: моделі LSTM можуть бути навчені розпізнавати та адаптуватися до шуму та викидів у даних, зменшуючи їхній вплив на процес прогнозування. Це важливо для покращення стійкості моделі та уникнення неточностей у прогнозах.

Створена нами програма в поєднанні з Facebook Prophet та ARIMA визначається трьома параметрами  $p$ ,  $d$  і  $q$ , які описують три основні компоненти моделі.

Інтегрований (I в ARIMA): кількість відмінностей, необхідних для досягнення стаціонарності, визначається параметром  $d$ . Нехай вихідні ознаки будуть  $Y_t$ , де  $t$  є індексом у послідовності. Ми створюємо стаціонарний часовий ряд, використовуючи наступні перетворення для різних значень  $d$ .

Для  $d=0$

$$y_t = Y_t , \quad (1.1)$$

У цьому випадку ряд вже нерухомий і нам нічого робити.

Для  $d=1$

$$y_t = Y_t - Y_{t-1}, \quad (1.2)$$

Це найбільш типове перетворення.

Для  $d=2$

$$y_t = Y_t - Y_{t-1} - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}, \quad (1.3)$$

Диференціювання можна розглядати як дискретну версію диференціювання. Для  $d=1$  нові функції представляють, як змінюються значення. У той час як для  $d=2$  нові функції представляють швидкість зміни, як і друга похідна в численні. Вищесказане також можна узагальнити до  $d>2$ , але це рідко використовується на практиці.

Минулі помилки передбачення:

$$\varepsilon_i = y_i - \bar{y}_i, \quad (1.4)$$

Поєднання трьох компонентів дає модель ARIMA( $p, d, q$ ). Точніше, ми спочатку інтегруємо часові ряди, а потім додаємо моделі ARiMA і вивчаємо відповідні коефіцієнти.

Щоб порівняти нашу програму та аналог, який застосовує модель LSTM ми збираємося використовувати дані фондової біржі для Bajaj Finserv Ltd, індійської фінансової компанії, щоб порівняти моделі. Набір даних охоплює період з 2008 року до кінця 2021 року. Він містить щоденну ціну акцій (середнє, найнижче та високе значення), а також загальний обсяг і оборот акцій, що торгуються (рис. 1.3).

```
1 df.head()
```

	Date	Symbol	Series	Prev Close	Open	High	Low	Last	Close	VWAP	Volume	Turnover	Trades	Deliverable Volume	%Deliverble
	2008-05-26	BAJAJFINSV	EQ	2101.05	600.00	619.00	501.0	505.1	509.10	548.85	3145446	1.726368e+14	NaN	908264	0.2888
	2008-05-27	BAJAJFINSV	EQ	509.10	505.00	610.95	491.1	564.0	554.65	572.15	4349144	2.488370e+14	NaN	677627	0.1558
	2008-05-28	BAJAJFINSV	EQ	554.65	564.00	665.60	564.0	643.0	640.95	618.37	4588759	2.837530e+14	NaN	774895	0.1689
	2008-05-29	BAJAJFINSV	EQ	640.95	656.65	703.00	608.0	634.5	632.40	659.60	4522302	2.982921e+14	NaN	1006161	0.2225
	2008-05-30	BAJAJFINSV	EQ	632.40	642.40	668.00	588.3	647.0	644.00	636.41	3057669	1.945929e+14	NaN	462832	0.1514

Рисунок 1.3 – Підвибірка набору даних

На базі даних фондової біржи ми можемо спрогнозувати зміни середньозваженої ціни (VWAP) наприкінці кожного дня (рис. 1.4).



Рисунок 1.4 – Графік часових рядів значень VWAP

Для оцінки ми розділили часові ряди на навчальний і тестовий часові ряди, де навчальний ряд складається з даних до кінця 2018 року (рис. 1.5).

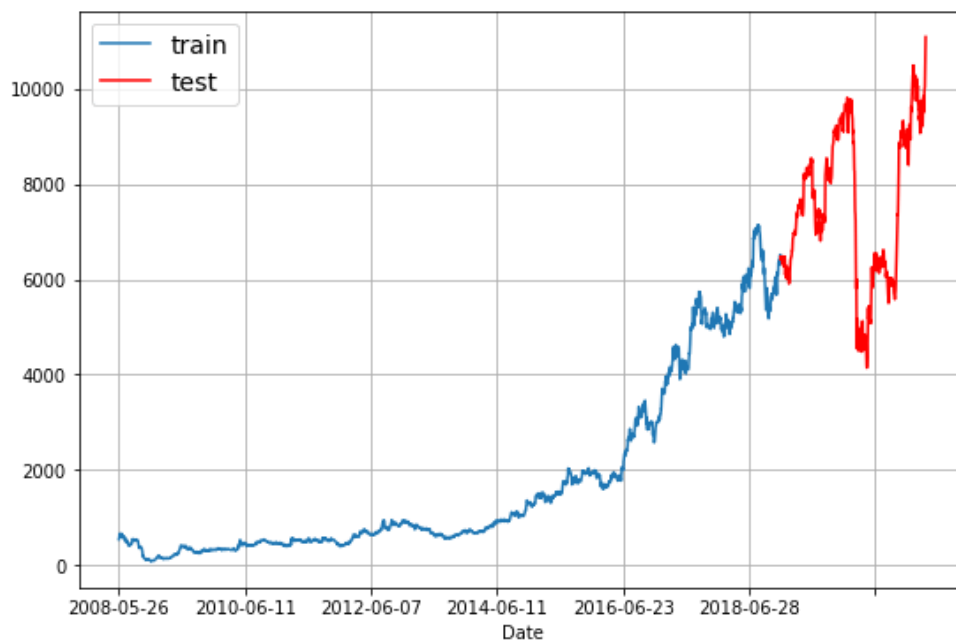


Рисунок 1.5 – Навчальні та тестові підмножини часового ряду VWAP

Точні прогнози, порівняні з істинними значеннями, можна побачити на наступних зображеннях (1.6 – 1.8).

Ми спостерігаємо, що моделі охоплюють загальну тенденцію часового ряду, але LSTM, здається, працює за кривою, тобто їй потрібно більше, щоб пристосуватися до зміни тенденції, що виявляє її неточності у випадку прогнозування криптовалют на короткий термін.

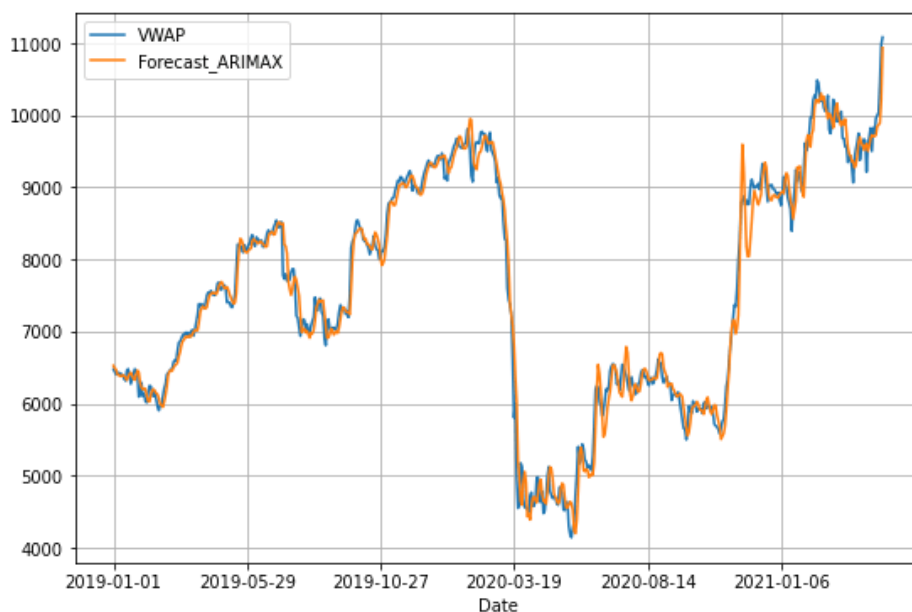


Рисунок 1.6 – Прогнози ARIMA



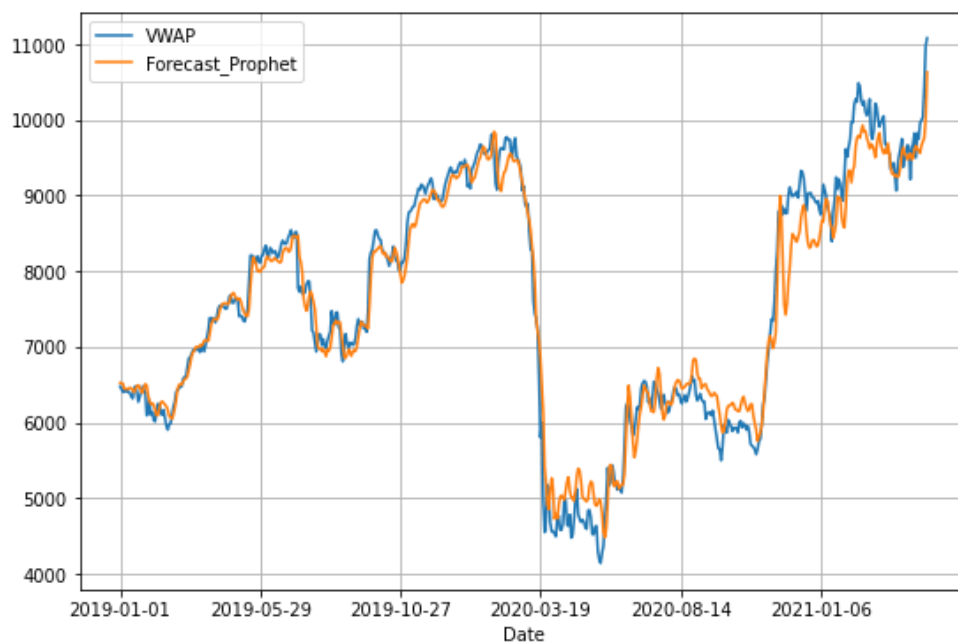


Рисунок 1.7 – Передбачення пророка

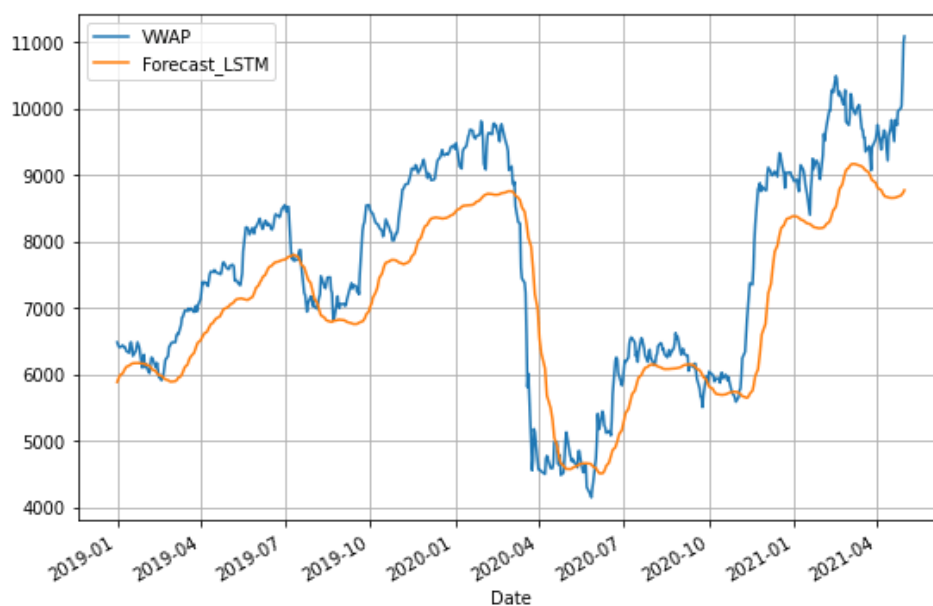


Рисунок 1.8 – Прогноз LSTM

Рекурентні нейронні мережі на основі LSTM є найгіршим підходом до навчання на основі часових рядів, які ми використовуємо в нашій програмі для прогнозування курсу криптовалют на короткий термін. У таких випадках недоліком є те, що RNN на основі LSTM важко інтерпретувати, і важко

інтуїтивно зрозуміти їхню поведінку. Крім того, для досягнення хороших результатів потрібна ретельна настройка гіперпараметрів.

#### **1.4 Висновок до розділу 1**

У першому розділі розглянуто постановку задачі прогнозування цифрових криптовалют та огляд відомих методів прогнозування курсу криптовалют. Було показано, що ціни на криптовалюти схильні до значної волатильності, що робить їх привабливими для інвесторів, але також і ризикованим вкладенням. Проведено огляд відомих методів вирішення задачі. Найперспективнішими були вибрані моделі машинного навчання Prophet, ARIMA для прогнозування курсу криптовалют на основі часових рядів. Недоліком моделі LSTM необхідно враховувати її складність навчання, вимогливість до обчислювальних ресурсів, чутливість до шуму в даних та складність розуміння та інтерпретації прогнозів.

## 2 РОЗРОБКА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПРОГНОЗУВАННЯ КУРСУ КРИПТОВАЛЮТИ ETHEREUM

### 2.1 Обґрунтування вибору моделі машинного навчання

У першому розділі було показано, що для задачі прогнозування курсу криптовалют варто обрати метод машинного навчання оскільки він має переваги перед іншими методами прогнозування курсу криптовалют. Але є багато типів машинного навчання. Тому потрібно визначити який саме тип машинного навчання слід обрати з великої кількості їх різновидів. Розглянемо типи машинного навчання та оберемо, що найбільш підходить для розв'язання поставленої задачі.

Машинне навчання (ML) – це галузь штучного інтелекту, яка дозволяє системам навчатися на даних і поліпшувати свою продуктивність з часом без явного програмування. ML використовується в широкому спектрі застосувань, включаючи розпізнавання образів, розпізнавання мови, природний мовний оброблення, рекомендаційні системи та прогнозування.

Існує два основних типи машинного навчання:

– Навчання з учителем – це тип машинного навчання, при якому система навчається на наборі даних, що містить правильні відповіді. Система використовує ці правильні відповіді, щоб навчитися прогнозувати правильні результати для нових даних.

– Навчання без учителя – це тип машинного навчання, при якому система навчається на наборі даних, що не містить правильних відповідей. Система використовує цей набір даних, щоб навчитися виявляти закономірності та тенденції в даних.

Навчання з учителем має ряд переваг, включаючи:

– Точність – системи, навчені з учителем, як правило, більш точні, ніж системи, навчені без учителя. Це пов'язано з тим, що системи, навчені з

учителем, мають доступ до правильних відповідей, які вони можуть використовувати для навчання.

– Зрозумілість – системи, навчені з учителем, як правило, більш зрозумілі, ніж системи, навчені без учителя. Це пов'язано з тим, що системи, навчені з учителем, використовують алгоритми, які можна легко зрозуміти та інтерпретувати.

Навчання без учителя має ряд переваг, включаючи:

– Застосовність – системи, навчені без учителя, як правило, більш універсальні, ніж системи, навчені з учителем. Це пов'язано з тим, що системи, навчені без учителя, не потребують набору даних, що містить правильні відповіді.

– Необхідність даних – системи, навчені без учителя, як правило, потребують менших наборів даних, ніж системи, навчені з учителем. Це пов'язано з тим, що системи, навчені без учителя, можуть вчитися на даних, що не містять правильних відповідей.

Вибір типу машинного навчання залежить від конкретних потреб завдання. Якщо точність та зрозумілість є важливими факторами, то навчання з учителем може бути найкращим вибором. Якщо універсальність та потреба в даних є важливими факторами, то навчання без учителя може бути найкращим вибором.

Машинне навчання використовується в широкому спектрі застосувань:

Розпізнавання образів – машинні моделі можуть використовуватися для розпізнавання об'єктів на зображеннях. Наприклад, машинні моделі використовуються для розпізнавання осіб, розпізнавання дорожніх знаків та розпізнавання товарів на полицях магазинів.

Розпізнавання мови – машинні моделі можуть використовуватися для розпізнавання мови. Наприклад, машинні моделі використовуються для розпізнавання мови голосових помічників, розпізнавання мови в чат-ботах та розпізнавання мови в системах перекладу.

Природний мовний оброблення – машинні моделі можуть використовуватися для обробки природної мови. Наприклад, машинні моделі

використовуються для перекладу мов, створення граматично правильних речень та розуміння запитів користувачів.

Рекомендаційні системи. – машинні моделі можуть використовуватися для створення рекомендацій для користувачів. Наприклад, машинні моделі використовуються для рекомендацій товарів на веб-сайтах, рекомендацій фільмів на потокових сервісах та рекомендацій музики на музичних сервісах.

Прогнозування часових рядів – машинні моделі можуть використовуватися для прогнозування майбутніх подій. Наприклад, машинні моделі використовуються для прогнозування погоди, прогнозування попиту на товари та прогнозування фінансових ринків.

Машинне навчання – це потужний інструмент, який може використовуватися для вирішення широкого спектру задач. Завдяки своїм перевагам, машинне навчання стає все більш важливим у сучасному світі.

З переліку типів машинного навчання нам найкраще підходить для виконання нашого завдання метод прогнозування часових рядів.

## **2.2 Синтез ознак на основі ознаки як часового ряду**

Для прогнозування цін на криптовалюту необхідно провести аналіз параметрів курсу криптовалюту як часового ряду.

Середнє значення курсу криптовалюту може бути використано для оцінки загальної тенденції курсу. Якщо середнє значення зростає, то це означає, що курс, як правило, зростає. Якщо середнє значення падає, то це означає, що курс, як правило, падає

Стандартне відхилення курсу криптовалюту може бути використано для оцінки мінливості курсу. Якщо стандартне відхилення високе, то це означає, що курс часто коливається. Якщо стандартне відхилення низьке, то це означає, що курс загалом стабільний.

Максимальне та мінімальне значення курсу криптовалюту можуть бути використані для оцінки діапазону курсу. Якщо максимальне значення значно

вище мінімального значення, то це означає, що курс схильний до значної волатильності.

Щоб зрозуміти, що таке розробка функцій на інтуїтивному рівні і чому вона необхідна, може бути корисно розшифрувати, як люди розуміють дані. Люди мають здатність, випереджаючи машину, знаходити складні закономірності чи відносини настільки, що можемо їх бачити, навіть коли їх насправді не існує.

Хоча розуміння даних і цільової проблеми є невід'ємною частиною інженерії функцій у машинному навчанні, і насправді немає жорстких правил щодо того, як цього досягти, слід знати наступні методи розробки функцій:

- Імпутація використовується для обробки відсутніх значень у наборі даних. Хоча одним із варіантів вирішення цієї проблеми є видалення записів, де відсутні значення, це може привести до втрати цінної інформації. У цьому випадку імпутація стає корисним методом.

- Дискретизація, зазвичай, включає в себе групування значень даних логічним чином у контейнери чи сегменти. Цей процес, також відомий як бінінг, може застосовуватися як до числових, так і до категоріальних значень. Він спрямований на уникнення переобладнання даних, хоча це може призвести до втрати деталізації вихідного набору даних.

- Категоричне кодування — це техніка, що використовується для перетворення категоріальних ознак у числові значення, що зазвичай легше сприймається алгоритмами. Одне гаряче кодування (ONE) є популярним методом категоричного кодування. Здесь категоріальні значення перетворюються на прості числові 1 і 0, зберігаючи при цьому інформацію. Однак ONE може призвести до значного збільшення кількості ознак і створення високо корельованих функцій, що може вплинути на ефективність моделі.

- Розбиття функцій на складові частини може підвищити цінність для вивчення конкретної цілі. Наприклад, в окремих випадках дата може бути більш важливою для прогнозування, ніж комбінація дати і часу. Розуміння взаємодії

різних аспектів функцій може допомогти оптимізувати модель та зробити її більш точною.

– Викиди - це надзвичайно великі або малі значення в наборі даних, які рідко зустрічаються в звичайних сценаріях. Оскільки ці відхилення можуть негативно вплинути на точність прогнозування, їх важливо належним чином обробляти, наприклад, використовуючи методи виявлення аномалій або нормалізації даних.

– Методи перетворення змінних грають важливу роль у нормалізації викривлених даних. Одним із широко використовуваних методів є логарифмічне перетворення, яке допомагає стиснути більші числа і розширити менші, зменшуючи перекося значень. Це особливо ефективно для важких розподілів даних. Інші використовувані перетворення включають перетворення квадратного кореня та перетворення Бокса-Кокса, яке узагальнює перші два.

– Масштабування функцій важливо через чутливість деяких алгоритмів машинного навчання до масштабу вхідних значень. Цей прийом, іноді відомий як нормалізація ознак, забезпечує однаковий масштаб для різних ознак, уникаючи впливу їхнього величини на результати алгоритму.

– Створення нових функцій з вже існуючих може включати в себе прості математичні операції, такі як агрегування для отримання середнього значення, медіани, моди, суми, різниці чи навіть добутку двох значень. Вибір цих характеристик для відповідності цілі може суттєво впливати на продуктивність моделі при ретельному підборі.

Як правило, усі датасети з даними на криптовалютних ринках містять базові ознаки, подані на рисунку 2.1

- «Open» – ціна відкриття інтервалу;
- «High» – максимум ціни інтервалу;
- «Low» – мінімум ціни інтервалу;
- «Close» – ціна закриття інтервалу;
- «Volume» – загальний об`єм;

Date	Open	High	Low	Close	Volume	Close_diff
2022-04-02	3449.788574	3521.284668	3442.000244	3445.059326	23571556215	-4.492920
2022-04-03	3444.810547	3573.960205	3421.259766	3522.833496	15333808649	77.774170
2022-04-04	3522.364990	3535.148193	3422.000977	3521.241211	18209969743	-1.592285
2022-04-05	3521.239746	3546.706787	3410.547607	3411.792480	16681503199	-109.448730
2022-04-06	3411.672119	3411.672119	3171.205078	3171.691895	25632563639	-240.100586
...	...	...	...	...	...	...
2023-10-06	1611.365723	1659.609863	1611.365723	1645.831543	4941208729	34.355103
2023-10-07	1645.810913	1648.183716	1631.158203	1634.511230	2578994988	-11.320312
2023-10-08	1634.657349	1641.182373	1618.189209	1633.548584	3315554175	-0.962646
2023-10-09	1633.454834	1635.447876	1553.006958	1579.806641	7020801716	-53.741943
2023-10-10	1580.112061	1593.741333	1553.031738	1567.713013	5254966125	-12.093628

Рисунок 2.1 – Базові ознаки датасету по криптовалюти Ethereum за даними сайту Yahoo

«OHLC-ознаки» – ними будують так звану OHLC-діаграму у вигляді свічок червоного (падіння курсу) та зеленого (зростання) кольорів (рис. 2.2).

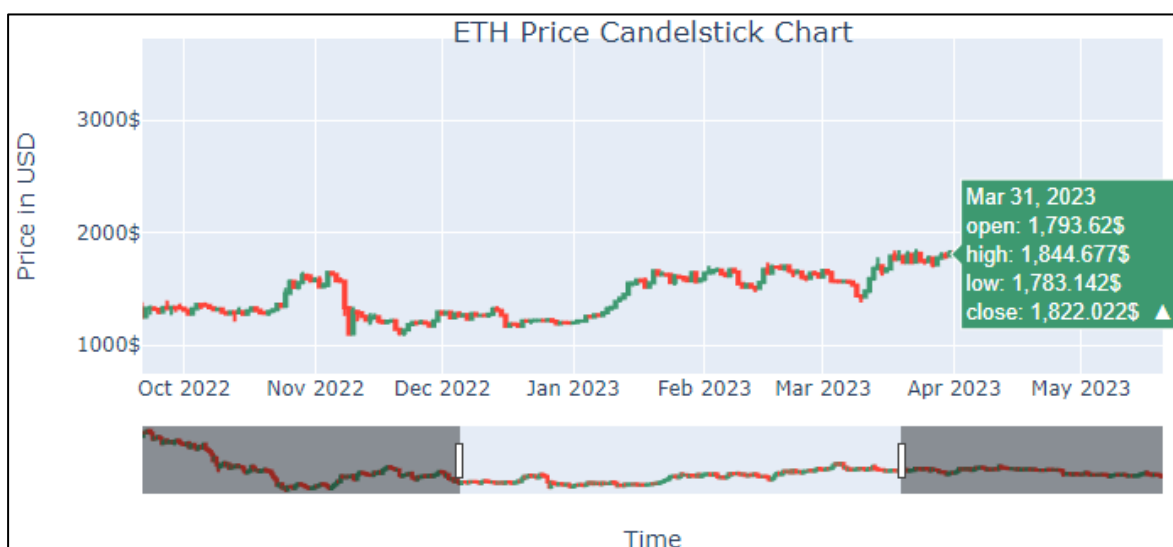


Рисунок 2.2 – OHLC-діаграма криптовалюти Ethereum

API на Yahoo містить ще «Adj Close», яке, як правило, він дорівнює «Close» і тому його видаляють.

Бібліотека TSFRESH є потужним інструментом для автоматичного створення різноманітних ознак з часових рядів. Застосовуючи цю бібліотеку до



ознаки "Close" і використовуючи фільтр для вилучення ознак без значень (NaN) та тих, що мають сталий рівень стандартного відхилення (які залишаються незмінними для всього ряду спостережень), можна отримати низку нових ознак.

На рисунку 2.3 відображений результат застосування TSFRESH до ознаки "Close" після використання зазначеного фільтра. Це може включати такі статистичні характеристики, як середнє значення, дисперсія, ексцес, медіана, а також агреговані значення для різних таймфреймів, таких як тиждень, місяць або день.

```
[ 'index__sum_values',
  'index__abs_energy',
  'index__median',
  'index__mean',
  'index__root_mean_square',
  'index__maximum',
  'index__absolute_maximum',
  'index__minimum',
  'index__benford_correlation',
  'index__quantile__q_0.1',
  'index__quantile__q_0.2',
  'index__quantile__q_0.3',
  'index__quantile__q_0.4',
  'index__quantile__q_0.6',
  'index__quantile__q_0.7',
  'index__quantile__q_0.8',
  'index__quantile__q_0.9',
  'index__cwt_coefficients__coeff_0_w_2_widths_(2, 5, 10, 20)',
  'index__cwt_coefficients__coeff_0_w_5_widths_(2, 5, 10, 20)',
  'index__cwt_coefficients__coeff_0_w_10_widths_(2, 5, 10, 20)',
  'index__cwt_coefficients__coeff_0_w_20_widths_(2, 5, 10, 20)',
  'index__fft_coefficient__attr_"real"__coeff_0',
  'index__fft_coefficient__attr_"abs"__coeff_0',
  'index__value_count__value_0',
  'index__value_count__value_1',
  'index__range_count__max_1__min_-1',
  'index__count_below__t_0',
  'Close__sum_values',
  'Close__abs_energy',
  'Close__median',
  'Close__mean',
  'Close__root_mean_square',
  'Close__maximum',
  'Close__absolute_maximum',
  'Close__minimum',
  'Close__benford_correlation',
  'Close__quantile__q_0.1',
  'Close__quantile__q_0.2',
  'Close__quantile__q_0.3',
  'Close__quantile__q_0.4',
  'Close__quantile__q_0.6',
  'Close__quantile__q_0.7',
  'Close__quantile__q_0.8',
  'Close__quantile__q_0.9',
  'Close__cwt_coefficients__coeff_0_w_2_widths_(2, 5, 10, 20)',
  'Close__cwt_coefficients__coeff_0_w_5_widths_(2, 5, 10, 20)',
  'Close__cwt_coefficients__coeff_0_w_10_widths_(2, 5, 10, 20)',
  'Close__cwt_coefficients__coeff_0_w_20_widths_(2, 5, 10, 20)',
  'Close__fft_coefficient__attr_"real"__coeff_0',
  'Close__fft_coefficient__attr_"abs"__coeff_0',
  'Date']
```

Рисунок 2.3 – Нові ознаки для ознаки «Close», згенеровані бібліотекою TSFRESH курсу криптовалюти Ethereum

Відібрано 57 ознак, у тому числі. 5 базових та 50 ознак з TSFRESH, на основі базової інформації на крипторинку. У наступному підрозділі розглянемо зовнішні фактори впливу.

### 2.3 Аналіз параметрів курсу криптовалюти як часового ряду

Для прогнозування курсу криптовалюти Ethereum, яка характеризується ознакою "Close", важливо визначити стаціонарність часового ряду та виявити можливість його стаціонаризації. Один із підходів до цього — використання критерію Діка-Фуллера (Dickey-Fuller).

Для перевірки чи є ряд стаціонарним використовується критерій Діка-Фуллера (Dickey-Fuller) (рис. 2.4) [19].

Тест ADF (Augmented Dickey-Fuller) є статистичним тестом значущості, який використовується для перевірки гіпотез про стаціонарність часового ряду. Результатом тесту є значення  $p$ , яке служить показником статистичної значущості. Зазвичай висновок формулюється на основі порівняння  $p$ -значення з певним пороговим значенням (наприклад, 0.05).

Якщо  $p$ -значення менше порогового, то є достатній емпіричний доказ того, що часовий ряд стаціонарний, і відхилення від його стаціонарності є статистично незначущими. Якщо  $p$ -значення велике, то ми не можемо відкинути нульову гіпотезу про нестаціонарність, і можливо, ряд потребує стаціонаризації для подальшого аналізу та моделювання.

```
result = adfuller(series.values)

print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
print('Critical Values:')
for key, value in result[4].items():
    print('\t%s: %.3f' % (key, value))

if (result[1] <= 0.05) & (result[4]['5%'] > result[0]):
    print("\u001b[32mStationary\u001b[0m")
else:
    print("\x1b[31mNon-stationary\x1b[0m")
```

Рисунок 2.4 – Програмний код для перевірки на стаціонарність курсу криптовалюти Ethereum, ряд даних series

Результат застосування цього критерію до ознаки «Close» (рис. 2.5).

```
ADF Statistic: -3.750036
p-value: 0.003462
Critical Values:
    1%: -3.442
    5%: -2.867
    10%: -2.570
Stationary
```

Рисунок 2.5 – Результат перевірки курсу криптовалюти Ethereum на стаціонарність

Як видно, з рисунку 2.5, ряд є стаціонарним. Зокрема, видно, що нульова гіпотеза про те, що він є стаціонарним, має значущість p-value 0,03, тобто менше, ніж 0,05 (рис. 2.6 - 2.7).

```
ADF Statistic: -5.500507
p-value: 0.000002
Critical Values:
    1%: -3.442
    5%: -2.867
    10%: -2.570
Stationary
```

Рисунок 2.6 – Перша різниця курсу криптовалюти Ethereum за стаціонарністю

```
ADF Statistic: -9.630940
p-value: 0.000000
Critical Values:
    1%: -3.443
    5%: -2.867
    10%: -2.570
Stationary
```

Рисунок 2.7 – Друга різниця курсу криптовалюти Ethereum за стаціонарністю

Можемо спостерігати, що перша і друга різниці ряду Ethereum за 2023 р. є стаціонарними часовими рядами. Оскільки важливим є якомога менший порядок, беремо і далі будемо моделювати та прогнозувати саме першу різницю «Close\_diff», яка на відміну від оригінального «Close» вже є стаціонарною (рис. 2.8).

Date	Open	High	Low	Close	Volume	Close_diff
2022-04-02	3449.788574	3521.284668	3442.000244	3445.059326	23571556215	-4.492920
2022-04-03	3444.810547	3573.960205	3421.259766	3522.833496	15333808649	77.774170
2022-04-04	3522.364990	3535.148193	3422.000977	3521.241211	18209969743	-1.592285
2022-04-05	3521.239746	3546.706787	3410.547607	3411.792480	16681503199	-109.448730
2022-04-06	3411.672119	3411.672119	3171.205078	3171.691895	25632563639	-240.100586
...	...	...	...	...	...	...
2023-10-06	1611.365723	1659.609863	1611.365723	1645.831543	4941208729	34.355103
2023-10-07	1645.810913	1648.183716	1631.158203	1634.511230	2578994988	-11.320312
2023-10-08	1634.657349	1641.182373	1618.189209	1633.548584	3315554175	-0.962646
2023-10-09	1633.454834	1635.447876	1553.006958	1579.806641	7020801716	-53.741943
2023-10-10	1580.112061	1593.741333	1553.031738	1567.713013	5254966125	-12.093628

Рисунок 2.8 – Курс криптовалюти Ethereum для першої різниці «Close\_diff»

Варто зазначити, що як цільову ознаку для часових рядів варто використовувати саме «Close», а для класичних моделей машинного навчання наприклад дерев рішень, нейронних мереж тощо, які обробляють багато ознак – «Close\_diff».

Для перевірки чи є ряд сезонним використовується команда `seasonal.seasonal_decompose` бібліотеки `Statsmodels.tsa` для декомпозиції ряду на складові, у тому числі – на сезонність, з використанням ковзного середнього [20] (рис. 2.9). Перевірка на сезонність полягає у декомпозиції ряду на тренд «Trend», сезонну складову «Seasonal» та залишок «Resid» (рис. 2.9).

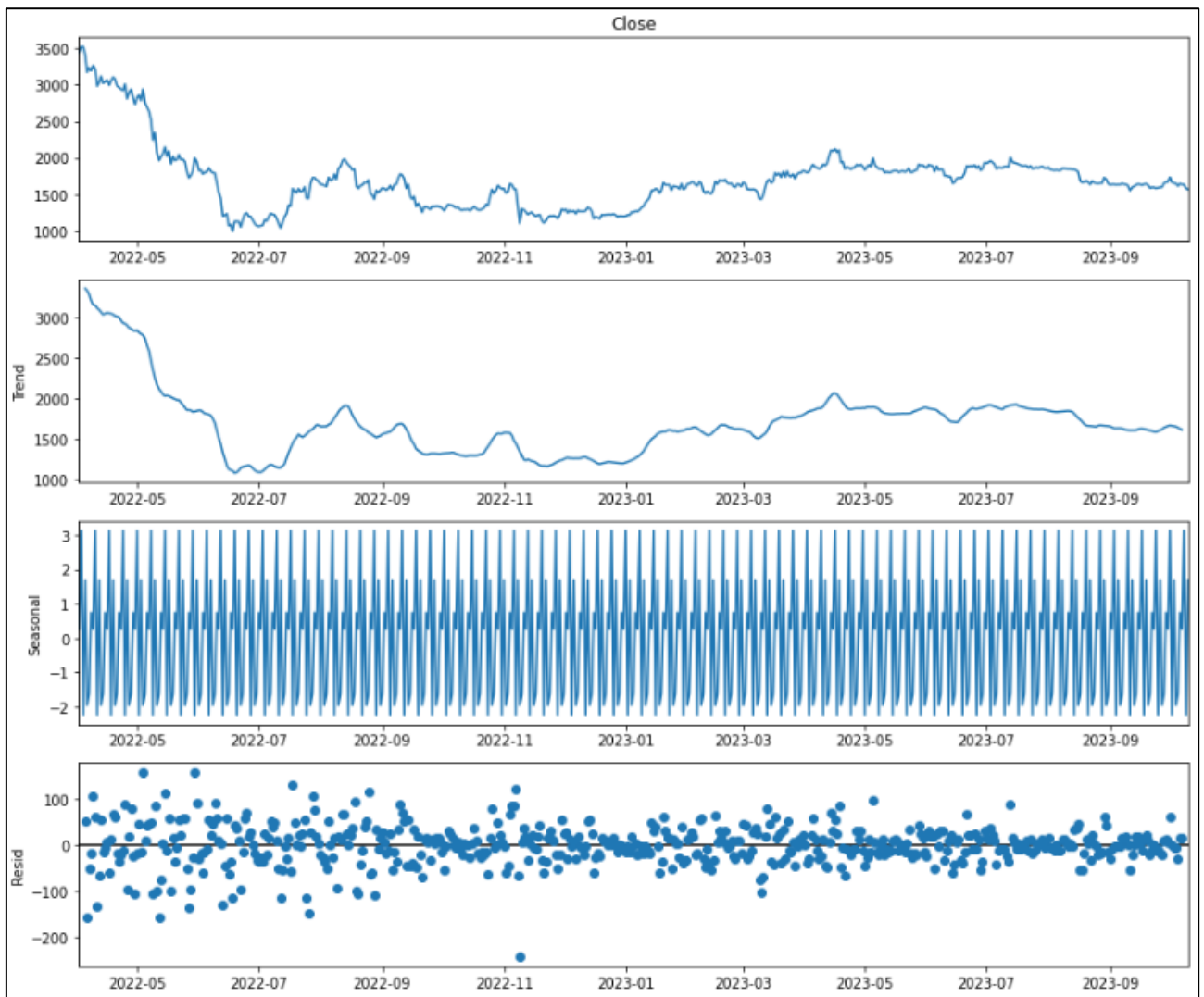


Рисунок 2.9 – Декомпозиція часового ряду на «Trend», «Seasonal» та «Resid», курсу криптовалюти Ethereum

З рисунку 2.9 видно, що через велику кількість даних, важко точно визначити період сезонності. Для цього здійснимо декомпозицію та візуалізуємо тільки дані за останній місяць (рис. 2.10).

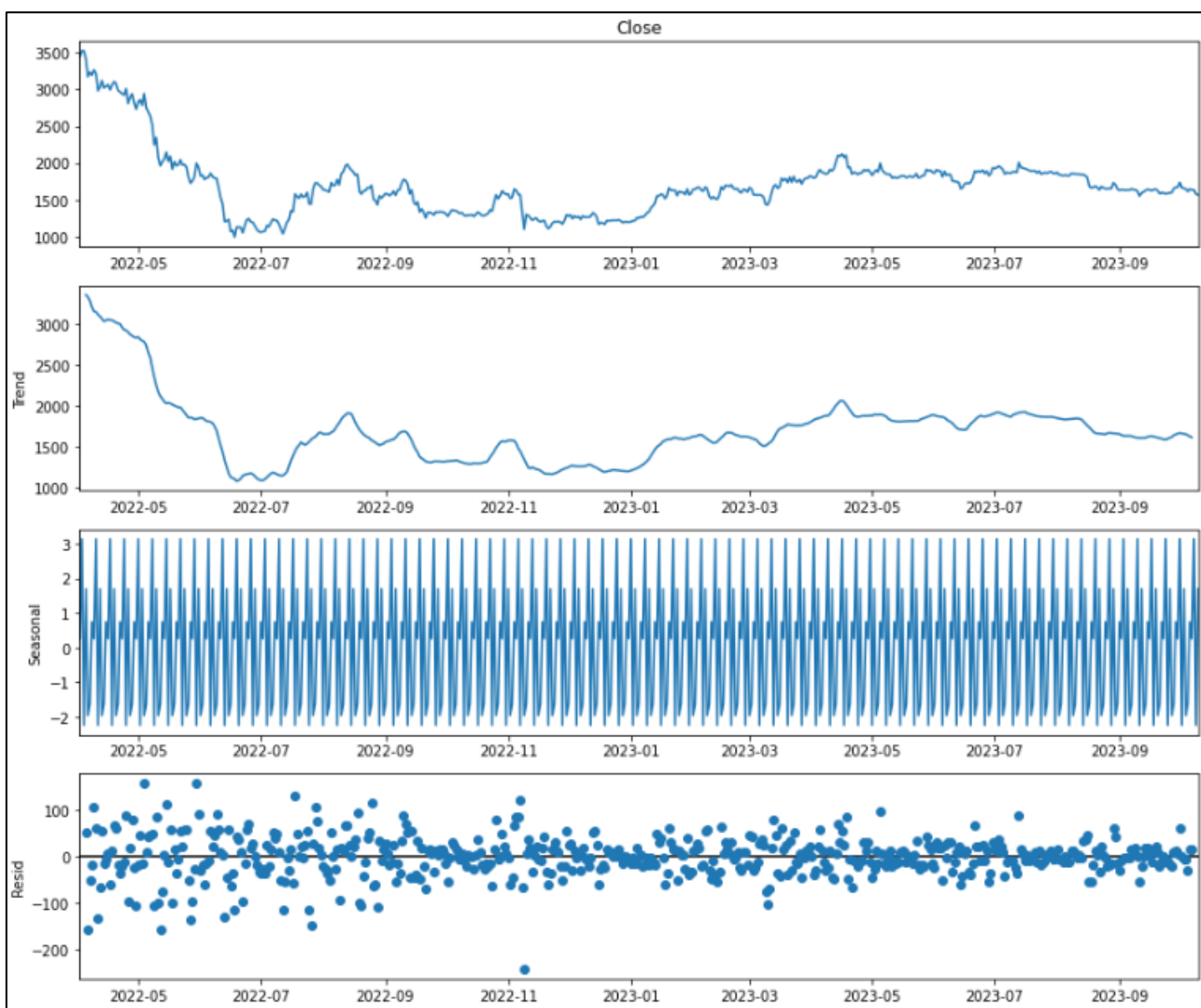


Рисунок 2.10 – Декомпозиція часового ряду курсу криптовалюти Ethereum за грудень 2023 року на складові.

Проаналізуємо який відсоток значень описує складова «Seasonal». Розглянемо весь ряд спостережень на рис. 2.9. З нього видно, що максимальна амплітуда значень цієї складової від  $-7.5$  до  $7.5$ , тобто – десь 15, у тому числі, час як основного ряду – приблизно від  $-200$  до  $400$ .

Отже, сезонна складова описує приблизно 3.5% даних, отже нею можна знехтувати, але можна спробувати ще пошукати період сезонності за більш складними моделями.

## 2.4 Аналіз аномалій ринку криптовалют

Проаналізуємо аномалії ринку. Аналіз періоду досліджень показав, що мав місце ряд аномалій (рис. 2.11-2.14).



Рисунок 2.11 – Аномалія 28.01.2022 р. різкої зміни курсу криптовалюти



Рисунок 2.12 – Аномалія 14.03.2022 р. різкої зміни курсу криптовалюти



Рисунок 2.13 – Аномалія 29.03.2022 р. різкої зміни курсу криптовалюти

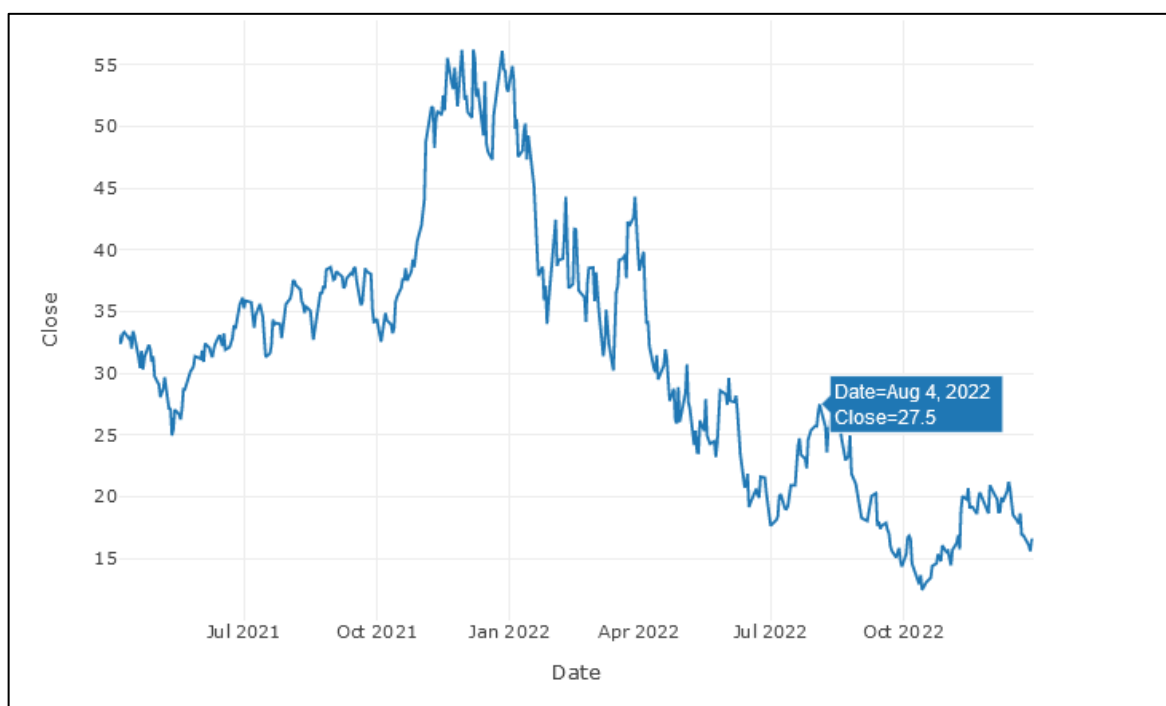


Рисунок 2.14 – Аномалія 04.08.2022 р. різкої зміни курсу криптовалюти

Можливі причини таких змін курсу:

Вплив на курс криптовалюти Ethereum (ETH) може бути значним у зв'язку з проведенням важливих технічних оновлень або протоколів в мережі Ethereum.



Дата 28 січня 2022 року вказує на можливу подію, яка може бути ключовою причиною зміни курсу ETH.

Технічні оновлення в мережі Ethereum можуть включати важливі зміни в протоколі, функціональності, безпеці чи масштабованості. Якщо такі зміни сприймаються ринком як поліпшення або розвиток мережі, це може призвести до збільшення попиту на Ethereum.

Зокрема, анонси або введення в дію технічних оновлень можуть залучити інтерес учасників ринку, інвесторів та розробників, що призведе до зміни психології ринку та динаміки попиту та пропозиції. Також важливо враховувати, що реакція ринку на такі події може бути вплинута різними факторами, включаючи загальний стан криптовалютного ринку та світові події.

– 14.03.2022 р. Ethereum аносувало важливе оновлення свого протоколу на 14 березня 2022 року. Це оновлення може включати поліпшення в області масштабованості, швидкості транзакцій, ефективності або безпеки мережі.

Вказана дата вказує на важливе оновлення протоколу Ethereum, яке може вплинути на різні аспекти мережі. Оновлення може призвести до поліпшень у масштабованості, прискоренні обробки транзакцій, підвищенні ефективності або підвищенні рівня безпеки. Реакція ринку на такі оновлення може визначитися тим, наскільки добре сприймаються ці зміни спільнотою та ринковими учасниками.

– 29.03.2022 р Зміни на ринку криптовалют: В цілому, загальний стан ринку криптовалют має великий вплив на курс Ethereum. Негативний настрій або падіння цін на інші популярні криптовалюти вплинуло на сприйняття ринковими учасниками Ethereum, що призвело до падіння його курсу.

Стан ринку криптовалют є важливим фактором для курсу Ethereum та інших криптовалют. Негативний настрій або зниження цін на інші криптовалюти може викликати об'ємне продажі та призвести до падіння курсу Ethereum через взаємозв'язок між різними цифровими активами на ринку.

– 04.08.2022 р. Оголошення з боку уряду або регуляторних органів щодо введення обмежень або заборон на використання криптовалют, включаючи Ethereum.

Ця подія вказує на важливий фактор впливу на курс Ethereum - оголошення з боку уряду або регуляторних органів. Введення обмежень або повна заборона на використання криптовалют може спричинити значний вплив на їхню легальність, прийняття та використання на ринку. Реакція ринку Ethereum на це оголошення може бути негативною, викликаючи великі коливання вартості через обурення та нестабільність серед інвесторів та користувачів криптовалют.

На рисунку 2.15 наведено сформований датасет із таких дат аномальних змін, який варто використати під час прогнозування курсу валюти.

	ds	lower_window	upper_window	prior_scale	holiday
0	2022-02-18	0	0	10	anomalous_dates
1	2022-02-28	0	0	10	anomalous_dates
2	2022-04-15	0	0	10	anomalous_dates
3	2022-05-11	0	0	10	anomalous_dates
4	2022-05-23	0	0	10	anomalous_dates
5	2022-09-28	0	0	10	anomalous_dates
6	2022-11-08	0	0	10	anomalous_dates
7	2022-12-14	0	0	10	anomalous_dates

Рисунок 2.15 – Дати аномальних змін курсу криптовалюти та їх параметри

На рисунку 2.16 наведено комбінований графік курсу криптовалюти та аномальних дат з рисунку 2.15.

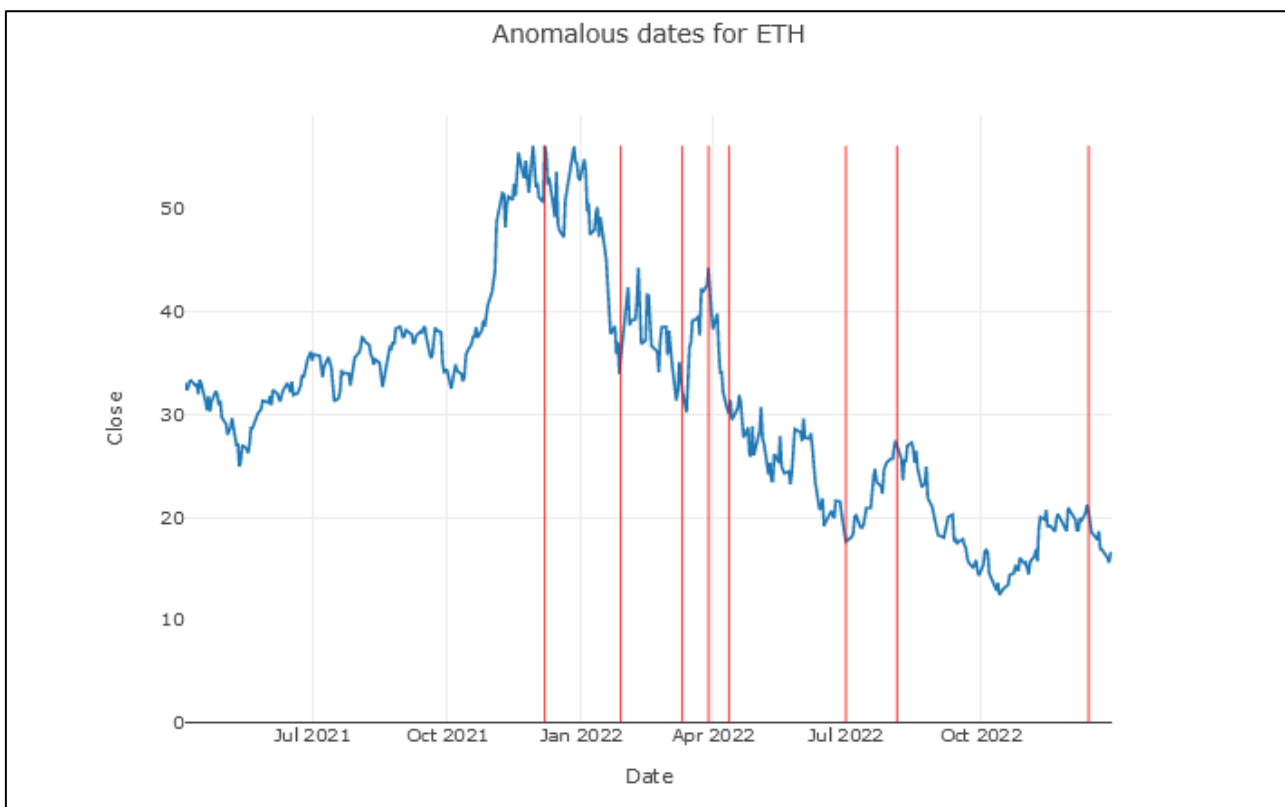


Рисунок 2.16 – Дати аномальних даних та курс криптовалюти

Комбінований графік курсу криптовалюти та аномальних дат - це потужний інструмент, який може допомогти зрозуміти, як аномальні дати впливають на курс криптовалюти. Він також може допомогти виявити закономірності в поведінці курсу криптовалюти.

## 2.5 Висновок до розділу 2

У другому розділі здійснено аналіз параметрів курсу криптовалюти як часового ряду, також спрогнозовано курс за допомогою ряду методів машинного навчання. Проведено аналіз аномалій курсу цифрової валюти Ethereum, аналіз впливу зовнішніх факторів на курс цифрової валюти Ethereum та огляд існуючих технологій прогнозування криптовалюти Ethereum. За допомогою синтезу ознак відібрано 57 ознак на основі базової інформації про криптовалюту Ethereum.

## **3 ПРОГРАМНА РЕАЛІЗАЦІЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПРОГНОЗУВАННЯ КУРСУ ETHEREUM НА ОСНОВІ МАШИННОГО НАВЧАННЯ**

### **3.1 Обґрунтування вибору мови та середовища програмування для прогнозування курсу криптовалют.**

Спочатку розглянемо мову програмування Python.

Python - це інтерпретована, інтерактивна, об'єктно-орієнтована мова програмування високого рівня із суворою динамічною типізацією. Розроблена в 1990 році Гвідо ван Россумом, Python стала однією з найпопулярніших мов програмування у світі.

Python використовується для широкого спектру задач, включаючи веб-розробку, машинне навчання, науку про дані та автоматизацію завдань.

Інтерпретована мова означає, що код Python виконується безпосередньо комп'ютером, без необхідності компіляції в машинний код. Це робить Python більш швидким та зручним для розробки, ніж компільовані мови.

Інтерактивна мова означає, що код Python можна виконувати безпосередньо в інтерпретаторі Python. Це дозволяє швидко тестувати та експериментувати з кодом.

Об'єктно-орієнтована мова означає, що код Python організований у вигляді об'єктів. Це робить код більш структурованим та зрозумілим.

Суворі динамічна типізація означає, що типи змінних визначаються під час виконання програми. Це робить код більш гнучким і простим для читання.

Python має ряд переваг, які роблять його популярним вибором для розробників, він легкий у вивченні. Синтаксис Python простий і зрозумілий, що робить його легким у вивченні для новачків.

Підтримка великої спільноти – Python має велику та активну спільноту розробників, яка надає підтримку та ресурси для вивчення та використання мови.

Модульність – Python підтримує модульну структуру, що дозволяє легко використовувати і повторно використовувати код.

Підтримка інтегрованого середовища розробки (IDE). Існує багато IDE для Python, які роблять розробку програм простішою та ефективнішою.

Python використовується для широкого спектру задач, включаючи:

Веб-розробка – Python є популярним вибором для розробки веб-додатків та веб-сервісів.

Машинне навчання – Python є однією з найпопулярніших мов для машинного навчання.

Наука про дані – Python є потужним інструментом для обробки та аналізу даних.

Автоматизація завдань – Python можна використовувати для автоматизації завдань, таких як тестування, розсилка електронної пошти та управління файлами.

Python є мовою програмування, яка має безліч переваг. Він простий у вивченні, має велику і активну спільноту, підтримує модульну структуру і має багато IDE. Ці переваги роблять Python популярним вибором для розробників, які працюють над проектами різного масштабу і складності.

Python – це потужний і універсальний інструмент, який можна використовувати для вирішення широкого спектру задач. Він є хорошим вибором для даної роботи.

### **3.2 Вибір спеціалізованої бібліотеки для програмної реалізації прогнозування курсу криптовалют.**

При виборі відповідної бібліотеки для реалізації програмного прогнозування курсу криптовалют слід враховувати такі чинники:

Можливості бібліотеки – бібліотека повинна надавати широкий спектр функцій для прогнозування цін на криптовалюти. Вона повинна включати в себе

методи для збору та обробки даних, для розробки моделей машинного навчання, а також для оцінки точності прогнозів.

Легкість використання – бібліотека повинна бути легкою у використанні, навіть для початківців. Вона повинна мати зрозумілу документацію та приклади коду.

Підтримка – бібліотека повинна мати активну спільноту розробників, яка надає підтримку та ресурси.

Для виконання нашої роботи було обрано та протестовано бібліотеки з великими можливостями такі як Facebook Prophet та ARIMA

Facebook Prophet - це відкрита бібліотека для прогнозування часових рядів, розроблена компанією Facebook. Вона надає інструменти для збору та обробки даних, для розробки моделей екстраполяції, а також для оцінки точності прогнозів.

Prophet заснований на моделі екстраполяції, яка включає в себе такі компоненти:

Тенденція – це довгостроковий тренд, який може бути лінійним або нелінійним.

Сезонність – це регулярні коливання, які повторюються через певні інтервали часу, наприклад, щодня, щотижня або щомісяця.

Екстраординарні події – це одноразові події, які можуть впливати на дані, наприклад, свята або природні катаклізми.

Prophet використовує ці компоненти для створення моделі, яка може прогнозувати майбутні значення часового ряду.

Легкість використання – Prophet є легкою у використанні, навіть для початківців. Вона має зрозумілу документацію та приклади коду.

Широкий спектр функцій – Prophet надає широкий спектр функцій для прогнозування часових рядів. Вона може бути використана для прогнозування різних типів часових рядів, включаючи часові ряди з сезонністю, часові ряди з аномальними даними та часові ряди з трендом.

Точність – Prophet показала високу точність прогнозів для різних типів часових рядів.

Prophet може бути використана для прогнозування різних типів часових рядів, включаючи:

Продажі. Prophet може бути використана для прогнозування продажів товарів або послуг.

Ціни – Prophet може бути використана для прогнозування цін на товари або послуги, включаючи криптовалюти.

Урожайність – Prophet може бути використана для прогнозування врожайності сільськогосподарських культур.

Погода – Prophet може бути використана для прогнозування погодних умов.

Facebook Prophet – це потужний інструмент для прогнозування часових рядів. Вона є легкою у використанні, має широкий спектр функцій і показує високу точність прогнозів. Prophet може бути використана для прогнозування різних типів часових рядів, включаючи продажі, ціни, урожайність і погоду.

ARIMA – це метод прогнозування часових рядів, який використовується для прогнозування майбутніх значень часового ряду на основі його минулої поведінки. Він заснований на припущенні, що часовий ряд можна описати як суму чотирьох компонентів:

Тенденція – це довгостроковий тренд, який може бути лінійним або нелінійним.

Сезонність – це регулярні коливання, які повторюються через певні інтервали часу, наприклад, щодня, щотижня або щомісяця.

Нестабільність – це коливання, які не є регулярними і неможливо передбачити.

Окремі події – це одноразові події, які можуть впливати на дані, наприклад, свята або природні катаклізми.

ARIMA використовує ці компоненти для створення моделі, яка може прогнозувати майбутні значення часового ряду.

ARIMA розшифровується як Autoregressive Integrated Moving Average, що означає Авторегресивний інтегрований ковзний середній. ARIMA-модель складається з трьох параметрів:

$p$  – це порядок авторегресії, який визначає кількість попередніх значень часового ряду, які використовуються для прогнозування поточних значень.

$d$  – це порядок інтеграції, який визначає, чи потрібно часовий ряд інтегрувати один або кілька разів, щоб усунути нестабільність.

$q$  – це порядок ковзного середнього, який визначає кількість попередніх відхилень від тренду, які використовуються для прогнозування поточних значень.

ARIMA має ряд переваг, які роблять її популярним вибором для прогнозування часових рядів:

Ефективність – ARIMA може бути ефективною для прогнозування різних типів часових рядів, включаючи часові ряди з трендом, часові ряди з сезонністю та часові ряди з нестабільністю.

Точність – ARIMA може показувати високу точність прогнозів для добре підібраних моделей.

Зручність – ARIMA є відносно легкою у використанні, навіть для початківців.

ARIMA може бути використана для прогнозування різних типів часових рядів, включаючи:

Продажі – ARIMA може бути використана для прогнозування продажів товарів або послуг.

Ціни – ARIMA може бути використана для прогнозування цін на товари або послуги, включаючи криптовалюти.

Урожайність – ARIMA може бути використана для прогнозування врожайності сільськогосподарських культур.



Погода – ARIMA може бути використана для прогнозування погодних умов.

ARIMA – це потужний метод прогнозування часових рядів, який може бути ефективним для прогнозування різних типів часових рядів. Він є відносно простим у використанні і може показувати високу точність прогнозів.

### 3.3 Прогнозування криптовалюти як часового ряду з використанням Facebook Prophet

Facebook Prophet — це алгоритм прогнозування часових рядів з відкритим вихідним кодом. Він використовує комбінацію класичних підходів з нововведеннями і є ефективним для моделювання часових рядів з кількома сезонними впливами. Проте він уникає деяких недоліків, характерних для інших алгоритмів. Основа Prophet — це сума трьох часових функцій (зростання  $g(t)$ , сезонність  $s(t)$ , свята  $h(t)$ ) та термін помилки  $\epsilon_t$ .

$$y(t) = g(t) + h(t) + \epsilon_t, \quad (3.1)$$

Модель Facebook Prophet є ефективною тільки для самих значень ряду, а не для його різниці.

```
Origin dataset has 365 rows and 2 features
Get training dataset with 345 rows
Get validation dataset with 10 rows
Get test dataset with 10 rows
```

Рисунок 3.1 – Розмір побудованих навчального, верифікованого та тестового датасетів з усіма ознаками курсу криптовалюти Ethereum

Враховуючи висновки з розділу 2.3, визначаємо параметри моделі наступним чином:

- Для річної та тижневої сезонності встановлюємо відсутність стандартних параметрів: `yearly_seasonality = False`, `weekly_seasonality = False`;
- Відсутність щоденної сезонності: `daily_seasonality = False`;
- Оскільки ряд має високу дисперсію і є дуже волатильним, потрібно встановити низьку регуляризацію: `changepoint_prior_scale = 0,5`;
- Додаємо параметри сезонності, які будемо підбирати в циклі.

Результати перебору варіантів параметрів моделі та визначення оптимальних параметрів та структури моделі представлені на рисунку 3.2.

```
)
model.add_seasonality(name='seasonality', period=period_days,
                      fourier_order=fourier_order_seasonality,
                      mode = 'multiplicative', prior_scale = 0.5)
```

Рисунок 3.2 – Структура моделі та ідентифіковані оптимальні параметри за технологією Facebook Prophet для прогнозування курсу криптовалюти Ethereum

На рисунку 3.3 наведено приклад результату прогнозування валідаційних даних за цією моделлю.

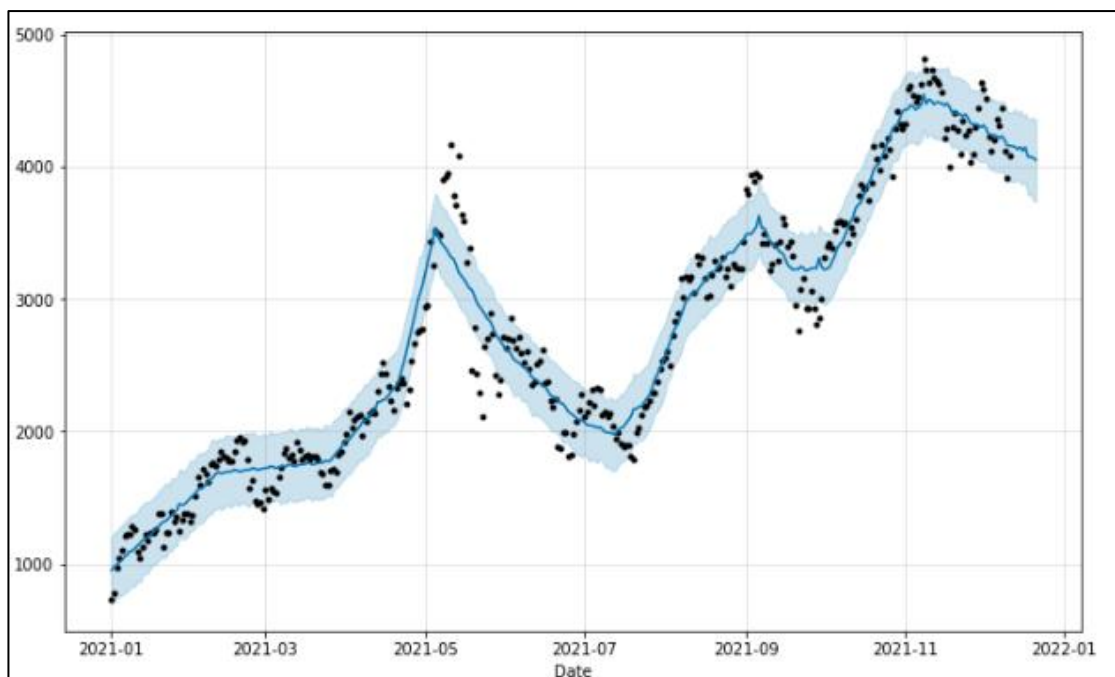


Рисунок 3.3 – Виконано прогнозування курсу валюти за найкращою моделлю за верифікованими даними на основі Facebook Prophet

У підсумку, у таблицю результатів вносяться прогнози усіх моделей які були побудовані і будуть проаналізовані разом з іншими моделями – це все буде наведено нижче.

### **3.4 Прогнозування цифрової валюти як часового ряду з використанням ARIMA**

Побудуємо модель для прогнозування ознаки "Close" курсу криптовалюти як часового ряду, використовуючи ARIMA-модель.

Моделі ARIMA надають інший підхід до прогнозування часових рядів. Експоненційне згладжування та ARIMA-моделі є двома найбільш широко використовуваними підходами до прогнозування часових рядів і надають додаткові інструменти для розв'язання проблеми. Моделі експоненційного згладжування засновані на описі тенденції та сезонності в даних, тоді як ARIMA спрямована на аналіз автокореляції в даних [22].

Спробуємо самостійно оцінити значення параметрів  $p$ ,  $d$ ,  $q$ .

Значення параметра  $d$  можна оцінити двома способами.

Використання тесту Діка–Фуллера:

- Якщо даний ряд є стаціонарним:  $d = 0$ ;
- Якщо стаціонарною буде перша різниця ряду:  $d = 1$ ;
- Якщо стаціонарною буде друга різниця ряду:  $d = 2$ , і так далі.

Візуальна оцінка за значеннями автокореляційної, часткової автокореляційної та повної автокореляційної функцій:

- Обирається те значення  $d$ , при якому значення  $p$ -value найшвидше потрапляють у "коридор" значень  $p$ -value = 0,05 (рис. 3.4).

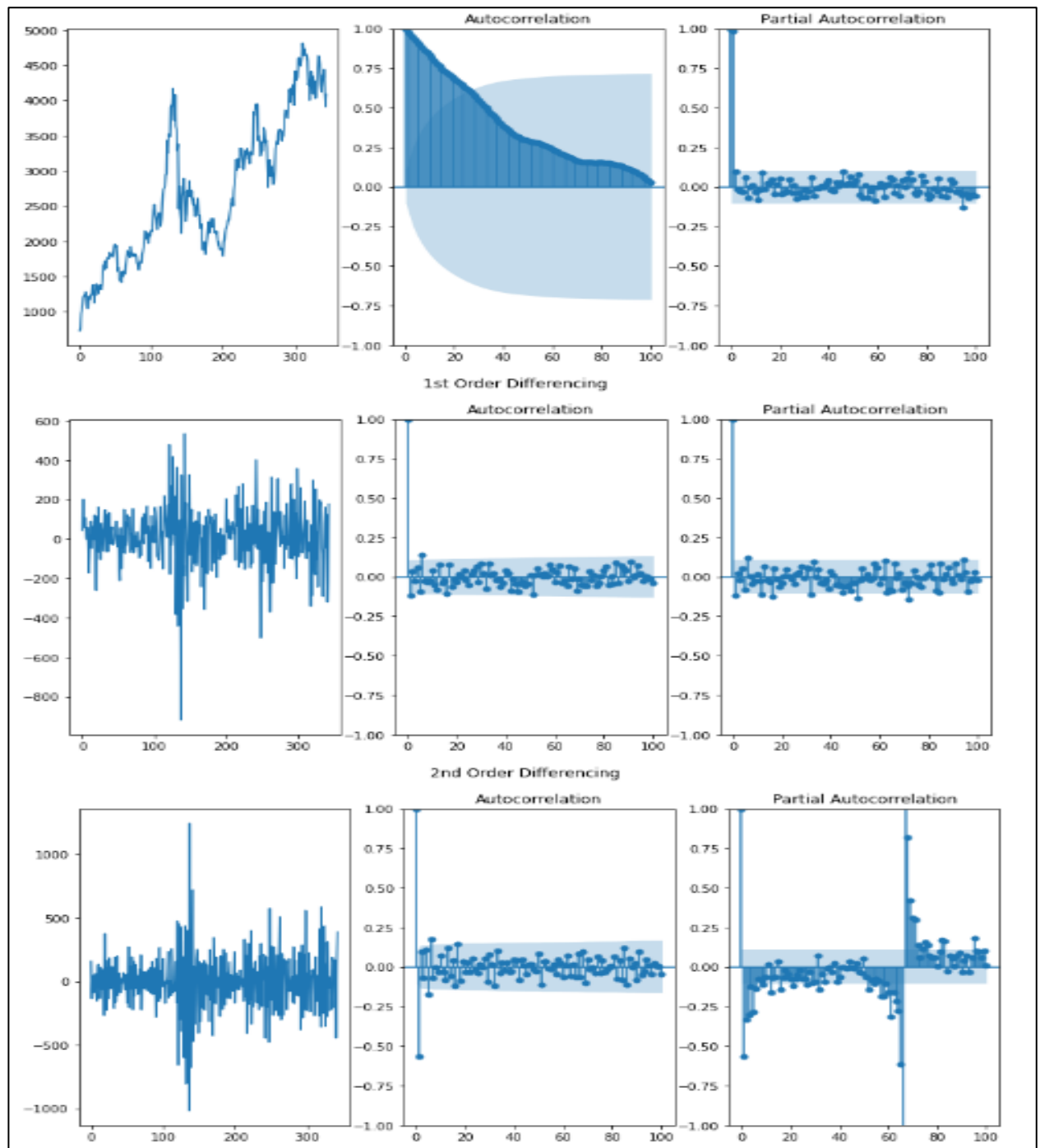


Рисунок 3.4 – Часткові автокореляційні та автокореляційні функції для першої та другої різниць ряду даних курсу криптовалюти Ethereum

Визначимо інші параметри - авторегресії та ковзного середнього. Для цього аналізуємо часткову автокореляційну функцію та спостерігаємо, на якому зсуві її значення зменшуються до зони значущості  $p\text{-value} = 0,05$ , враховуючи верхню та нижню межі цієї зони. Значення цього зсуву визначає порядок  $p$  у моделі авторегресії (рис. 3.5)

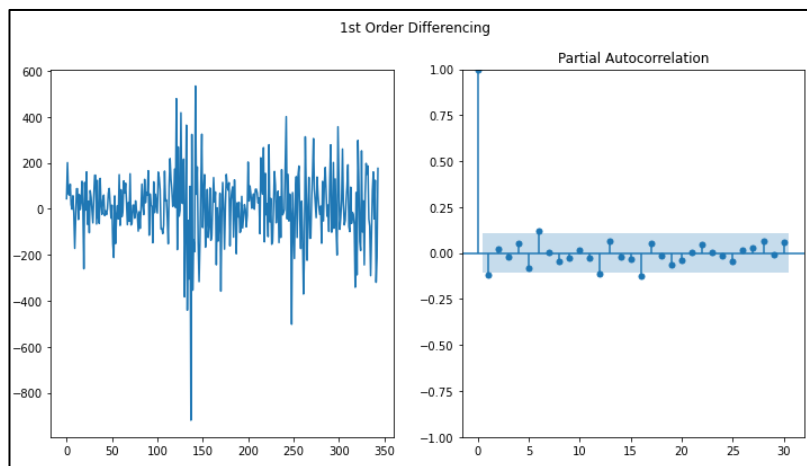


Рисунок 3.5 – Часткова автокореляційна функція курсу криптовалюти Ethereum для першої різниці

Як видно з рисунку 3.5,  $p = 0$ .

Потім проводимо обчислення автокореляційної функції та вивчаємо, на якому зсуві її значення зменшуються до зони значущості  $p\text{-value} = 0,05$ , враховуючи верхню та нижню межі цієї зони. Значення цього зсуву визначає порядок  $q$  у моделі ковзного середнього (рис. 3.6).

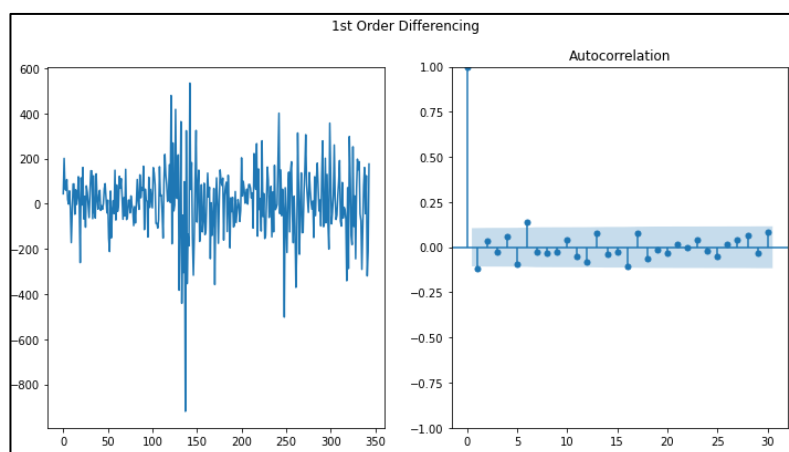


Рисунок 3.6 – Автокореляційна функція курсу криптовалюти Ethereum для першої різниці

Як видно з рисунку 3.6,  $q = 0$ .

Отже, аналіз показує, що оптимальною повинна бути модель:

$$\text{ARIMA}(p, q, d) = \text{ARIMA}(0, 1, 0). \quad (3.2)$$

На рисунках 3.7, 3.8 наведено результат моделювання валідаційних даних за моделлю (3.2).

SARIMAX Results					
Dep. Variable:	y	No. Observations:	345		
Model:	ARIMA(0, 1, 0)	Log Likelihood	-2212.315		
Date:	Wed, 08 Jun 2022	AIC	4426.630		
Time:	09:39:25	BIC	4430.471		
Sample:	0	HQIC	4428.160		
	- 345				
Covariance Type:	opg				
	coef	std err	z	P> z	[0.025 0.975]
sigma2	2.257e+04	946.557	23.845	0.000	2.07e+04 2.44e+04
Ljung-Box (L1) (Q):			4.66	Jarque-Bera (JB):	362.88
Prob(Q):			0.03	Prob(JB):	0.00
Heteroskedasticity (H):			3.11	Skew:	-0.70
Prob(H) (two-sided):			0.00	Kurtosis:	7.83

Рисунок 3.7 – Результат прогнозування за оптимальною моделлю ARIMA для курсу валюти (3.2)

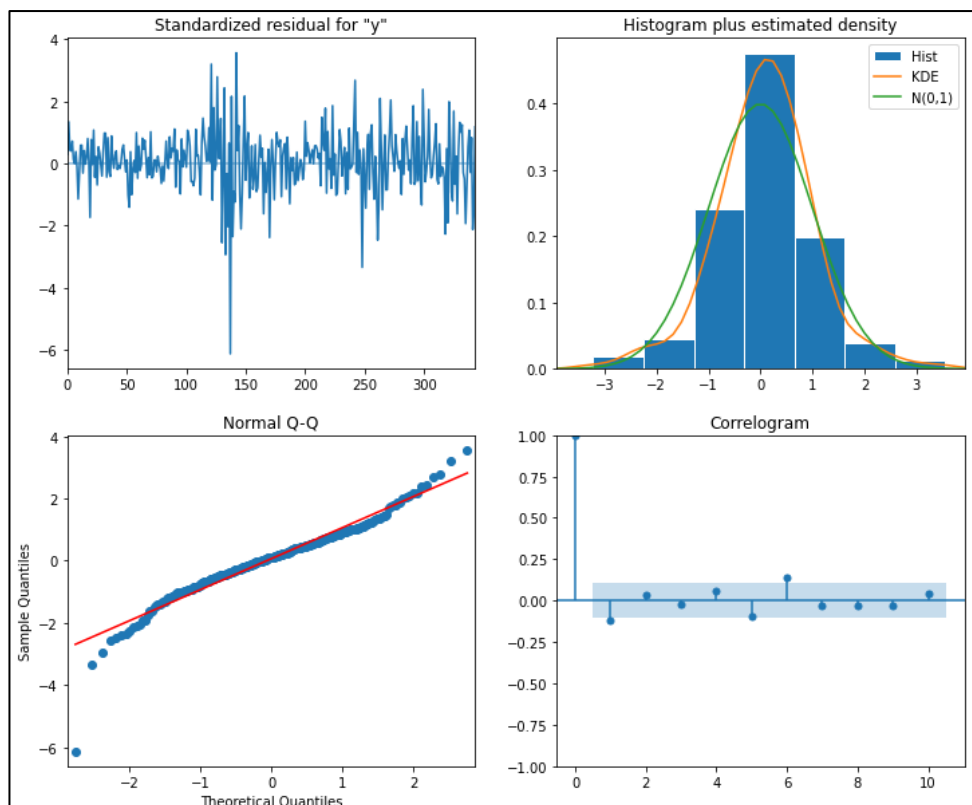


Рисунок 3.8 – Результат похибки прогнозу верифікованих даних оптимальної моделі ARIMA (3.2)

Помітно, що отримані результати є задовільними. Однак модель виявилася занадто простою, що свідчить про те, що метод ARIMA не дозволяє побудувати достатньо адекватну модель. Тому, ймовірно, необхідно розглядати інші підходи для досягнення кращих результатів.

### 3.5 Прогнозування криптовалюти методами машинного навчання

Плануємо побудувати кілька моделей, використовуючи методи машинного навчання. Під час цього процесу використаємо різні ознаки та виконаємо стандартизацію за допомогою методу "sklearn.preprocessing.StandardScaler".

Ідентифікуємо моделі, які зазначені на рисунку 3.9.

	name	model	param_grid
0	Linear Regression	LinearRegression()	{'fit_intercept': [True, False]}
1	KNeighbors Regressor	KNeighborsRegressor()	{'n_neighbors': [3, 5, 10, 20, 30], 'leaf_size...
2	Support Vector Machines	SVR()	{'kernel': ['linear', 'poly', 'rbf', 'sigmoid'...
3	Linear SVR	LinearSVR()	{'C': [1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0,...
4	Random Forest Regressor	RandomForestRegressor()	{'n_estimators': [40, 50, 60, 80], 'min_sample...
5	Bagging Regressor	BaggingRegressor()	{'max_features': [0.05], 'n_estimators': [3, 4...
6	XGB Regressor	XGBRegressor(base_score=None, booster=None, ca...	{'n_estimators': [50, 70, 90], 'learning_rate'...
7	MLP Regressor	MLPRegressor()	{'hidden_layer_sizes': [2, 3, 4], 'solver': [...

Рисунок 3.9 – Моделі машинного навчання та їх параметри

Linear Regression – це метод аналізу, що використовується для прогнозування значень однієї змінної на основі значень іншої змінної. Змінна, яку ми намагаємося передбачити, називається залежною змінною, тоді як змінна, за допомогою якої ми робимо прогноз, називається незалежною змінною [15].

Лінійна регресія визначає пряму лінію або поверхню, яка мінімізує відхилення між прогнозованими та фактичними значеннями вихідної змінної. Для цього використовується метод "найменших квадратів", який дозволяє знайти

лінію, найкраще відповідну набору парних даних. Потім визначаються значення залежної змінної ( $Y$ ) на основі незалежної змінної ( $X$ ) [23].

KNeighbors Regressor використовує набір даних, відомий як навчальний набір вибірок, де кожен запис має мітку або категорію. Коли подаються нові дані без міток, кожна їх ознака порівнюється з відповідною ознакою в навчальному наборі. Алгоритм отримує мітки класифікації даних, які мають найбільше схожі ознаки в навчальному наборі (найближчі сусіди). Зазвичай вибирають лише перші  $K$  найбільш схожих записів у навчальному наборі даних. Це основна ідея алгоритму  $K$ -найближчих сусідів [24].

Bagging Regressor - це метаоцінювач, який працює на основі іншого регресора. Він створює кілька випадкових підмножин вихідного набору даних і надає кожному базовому регресору можливість оцінити дані в цих підмножинах. Після цього він комбінує прогнози кожного базового регресора, часто використовуючи голосування або усереднення, щоб отримати остаточний прогноз. Bagging Regressor допомагає зменшити дисперсію оцінки, дозволяючи базовим регресорам будувати моделі випадковим чином, що призводить до більш стійких та точних результатів [25].

XGBoost (eXtreme Gradient Boosting) Regressor - це алгоритм машинного навчання, який використовує метод градієнтного бустінгу для задач регресії. Він базується на ідеї об'єднання слабких моделей (зазвичай дерев рішень) в сильну модель. Алгоритм працює шляхом послідовного навчання дерев рішень, коригуючи помилки попередніх моделей.

XGBoost володіє численними перевагами, такими як ефективність, висока швидкість роботи та висока точність прогнозування. Він також враховує регуляризацію, що допомагає уникнути перенавчання моделі.

У випадку XGBoost Regressor, алгоритм використовується для вирішення завдань регресії, де цільовою змінною є неперервна величина, наприклад, прогнозування ціни акцій, температури або іншого числового параметра.

MLP Regressor (Multi-Layer Perceptron Regressor) - це алгоритм машинного навчання, який використовує штучні нейронні мережі (ансамблі нейронів) для



вирішення задач регресії. Цей алгоритм входить до сімейства нейронних мереж та представляє собою тип штучних нейронних мереж, в яких інформація рухається вперед через кілька шарів нейронів.

Основна ідея MLP Regressor полягає в тому, щоб навчити модель вагам мережі, які забезпечують оптимальне відображення вхідних функцій на цільову змінну. Архітектура мережі має хоча б один прихований шар, який дозволяє виявлення складних нелінійних залежностей в даних.

MLP Regressor застосовується для завдань регресії, де цільовою змінною є неперервна величина. Він широко використовується для прогнозування значень у таких областях, як фінанси, економіка, медицина та інші, де важливо отримати числові прогнози.

Linear SVR (Support Vector Regression) - це метод машинного навчання, який використовує механізми опорних векторів для розв'язання задач регресії. Він є варіацією методу опорних векторів (SVM) і призначений для вирішення задач прогнозування значень, де цільова змінна є неперервною.

Основна ідея Linear SVR полягає в тому, щоб знаходити лінійну функцію, яка найкращим чином відображає залежність між вхідними ознаками і цільовою змінною. Модель намагається знайти гіперплощину, яка максимально відділяє точки даних у просторі так, щоб максимізувати точність прогнозування.

Відмінністю Linear SVR від класичного SVM для класифікації полягає в тому, що Linear SVR спрямований на знаходження лінійної регресійної функції, яка максимально відділяє точки даних від регресійної функції, що відображає залежність між вхідними ознаками і цільовою змінною.

Support Vector Machines (SVM) - це метод машинного навчання, який використовується як для задач класифікації, так і для регресії. Основна ідея SVM полягає в тому, щоб знаходити гіперплощину в просторі ознак, яка найкращим чином розділяє два класи (в задачі класифікації) або максимально підтримує точки даних (в задачі регресії).

```

Tuning model 'Linear Regression'
Best parameters: {'fit_intercept': False}

Tuning model 'KNeighbors Regressor'
Best parameters: {'leaf_size': 10, 'n_neighbors': 30}

Tuning model 'Support Vector Machines'
Best parameters: {'C': 7.0, 'kernel': 'sigmoid', 'tol': 0.0001}

Tuning model 'Linear SVR'
Best parameters: {'C': 1.0}

Tuning model 'Random Forest Regressor'
Best parameters: {'max_depth': 4, 'max_features': 'auto', 'min_samples_leaf': 50, 'min_samples_split': 30, 'n_estimators': 80}

Tuning model 'Bagging Regressor'
Best parameters: {'max_features': 0.05, 'n_estimators': 6, 'warm_start': False}

Tuning model 'XGB Regressor'
Best parameters: {'learning_rate': 0.01, 'max_depth': 3, 'n_estimators': 50}

Tuning model 'MLP Regressor'
Best parameters: {'hidden_layer_sizes': 2, 'learning_rate': 'adaptive', 'learning_rate_init': 0.001, 'max_iter': 1000, 'solver': 'sgd'}

```

Рисунок 3.10 – Оптимальні параметри ідентифікованих моделей, побудованих для багатьох ознак для верифікованих даних курсу криптовалюти

### 3.6 Вибір оптимальної моделі

Після виконання обчислень на перевірених даних відбувається відновлення значень прогнозу з «Close\_diff» на «Close». Подальше порівняння прогнозів з відомими цільовими значеннями «Close» у верифікованому датасеті проводиться за допомогою трьох метрик:  $r^2\_score$ , RMSE, MAPE.

На рисунку 3.11 зображено результат відбору оптимальної моделі у випадку прогнозування курсу Ethereum на 10 діб.

	name_model	type_data	r2_score	rmse	mape
14	Random Forest Regressor	valid	-0.947526	151.154076	2.945693
11	KNeighbors Regressor	valid	-0.982406	152.501624	3.232935
10	Linear Regression	valid	-1.362067	166.465542	3.635112
12	Support Vector Machines	valid	-1.567659	173.558912	3.764183
16	XGB Regressor	valid	-1.668826	176.945012	3.905291
13	Linear SVR	valid	-1.702011	178.041709	3.940936
7	Prophet_14_days_12_order	valid	-2.480632	202.072527	4.007154
6	Prophet_14_days_3_order	valid	-2.626355	206.259228	4.01593
9	ARIMA_auto	valid	-1.809227	181.539691	4.039218
8	ARIMA_manual	valid	-1.873414	183.60197	4.087203
4	Prophet_7_days_3_order	valid	-2.811506	211.459154	4.517025
2	Prophet_5_days_3_order	valid	-3.001961	216.677911	4.627023
0	Prophet_4_days_3_order	valid	-3.013654	216.994213	4.665902
17	MLP Regressor	valid	-3.095436	219.193802	5.147478
5	Prophet_7_days_12_order	valid	-4.013708	242.525687	5.510833
3	Prophet_5_days_12_order	valid	-4.571395	255.658463	5.861905
1	Prophet_4_days_12_order	valid	-4.844546	261.850624	6.063803
15	Bagging Regressor	valid	-36.917089	666.95372	14.294553

Рисунок 3.11 – Похибки всіх моделей, що були побудовані для багатьох ознак для верифікованих даних курсу криптовалюти на 10 діб

$r2\_score$  – це оцінка, де найкращий можливий бал - 1,0, і вона може бути негативною, що вказує на те, що модель гірша за постійну модель, яка завжди передбачає середнє значення у. Якщо істинне значення у є непостійним,  $R^2$  може бути 0,0 [27].

RMSE – це стандартне відхилення залишків, що вимірює, наскільки далеко від лінії регресії розташовані точки даних. Ця метрика показує, наскільки розподілені залишки.

MSE – середня сума квадратної різниці між прогнозованим і фактичним значенням для всіх точок даних. Чим менше MSE, тим точніше прогнозування, і ідеальний результат досягається при значенні 0.

MAPE – це середня абсолютна помилка у відсотках.

Random Forest Regressor – це модель, яка використовує дерева рішень для регресії або класифікації. Вона використовує дерева, які візуально нагадують структуру дерева, розгалужуючись на основі значень змінних, поки не досягає листового вузла і не надає прогноз.

У випадку прогнозування курсу на 10 днів оптимальна модель Random Forest Regressor має низьку відносну похибку за метрикою MAPE, порівняно малу похибку RMSE, але дуже погану точність за метрикою r2\_score, яка має від'ємне значення. Це означає, що самі значення прогнозуються досить точно, але прогноз напрямку зростання чи падіння курсу є дуже неточним. Таким чином, оптимальна модель виявляється доброю лише за метрикою MAPE, тобто відносною похибкою.

На рисунку 3.12 зображено результат відбору оптимальної моделі у випадку прогнозування курсу Ethereum на 4 доби.

	name_model	type_data	r2_score	rmse	mape
11	KNeighbors Regressor	valid	-0.121617	22.468789	0.446885
14	Random Forest Regressor	valid	-0.312955	24.309851	0.59479
13	Linear SVR	valid	-0.744307	28.020044	0.609081
9	ARIMA_auto	valid	-2.462671	39.478707	0.826748
8	ARIMA_manual	valid	-4.799886	51.093664	1.147166
16	XGB Regressor	valid	-4.977551	51.870325	1.170538
12	Support Vector Machines	valid	-11.976779	76.425939	1.762409
15	Bagging Regressor	valid	-46.460514	146.158319	3.51471
7	Prophet_14_days_12_order	valid	-47.873985	148.318798	3.580896
17	MLP Regressor	valid	-58.98706	164.318366	3.588498
6	Prophet_14_days_3_order	valid	-51.97449	154.415436	3.724006
10	Linear Regression	valid	-65.480593	172.983532	4.081759
0	Prophet_4_days_3_order	valid	-87.956725	200.099827	4.877806
2	Prophet_5_days_3_order	valid	-88.498487	200.708224	4.910621
4	Prophet_7_days_3_order	valid	-98.091198	211.190734	5.178781
3	Prophet_5_days_12_order	valid	-103.481647	216.858938	5.307153
5	Prophet_7_days_12_order	valid	-109.523133	223.040594	5.469506
1	Prophet_4_days_12_order	valid	-110.656693	224.181463	5.470795

Рисунок 3.12 – Похибка усіх моделей, побудованих для багатьох ознак для валідаційних даних курсу цифрової валюти на 4 доби

Для прогнозування курсу Ethereum протягом 4 діб "KNeighbors Regressor" виявився найкращою моделлю. Похибки моделі за трьома метриками свідчать про середню точність за  $r2\_score$ , а також малу похибку RMSE і низьку відносну похибку, що вказує на відносно добре передбачення напрямку змін курсу.

### **3.7 Висновок до розділу 3**

У третьому розділі здійснено аналіз параметрів курсу криптовалюти як часового ряду, спрогнозовано курс криптовалюти з використанням Facebook Prophet та ARIMA як часового ряду, також спрогнозовано курс за допомогою ряду методів машинного навчання та обрано оптимальну модель для прогнозування курсу криптовалюти на 10 діб – «Random Forest Regressor» з похибкою MAPE 2.95% та на 4 доби – «KNeighbors Regressor» з похибкою MAPE 0.44%.

## 4 ТЕСТУВАННЯ ТА АНАЛІЗ РЕЗУЛЬТАТІВ РОБОТИ ПРОГРАМИ ПРОГНОЗУВАННЯ КУРСУ ETHEREUM У 2023 РОЦІ

### 4.1 Розроблення інформаційної технології для подальшого прогнозування

На рисунку 4.1 зображено алгоритм розв'язання поставленої задачі для тестових даних:

1. Загальна характеристика поставленої задачі.
2. Виконати синтез ознак. З аномальних дат потрібно сформувати датасет. Для тестових даних синтезувати дані ознаки: `upper_shadow`, `low2mean`, `median_price`, `low2median`, `high2median`, `high2mean`, `lower_shadow`, `open2close` і ін.
3. Генерування цільової ознаки.
4. Попереднє оброблення даних, тобто за допомогою метода «`sklearn.preprocessing.StandardScaler`» зробити стандартизацію ознак.
5. Для різних типів моделей створити тренувальний, валідаційний і тестовий датасет.
6. Здійснення вибору і тренування моделей на «`train`» датасеті і за кожною ідентифікованою моделлю «`valid`» даних виконання прогнозу.
7. Вибір найоптимальнішої моделі за різними метриками.
8. За розширеним датасетом, до якого входить валідаційний і тренувальний, здійснити тренування оптимальної моделі. За ідентифікованою моделлю «`test`» даних зробити прогнозування.
9. На основі прогнозу тестових даних, потрібно здійснити оцінку точності прогнозування.

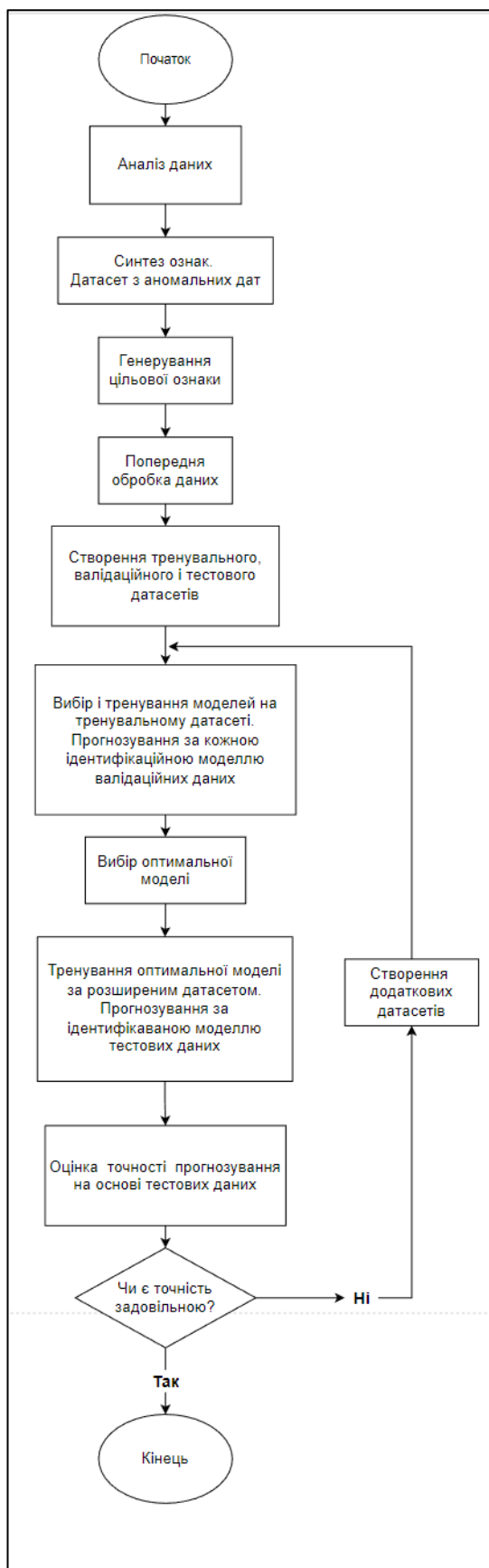


Рисунок 4.1 – Блок-схема виконання алгоритму інформаційної технології аналізу та прогнозування курсу Ethereum у 2023 році

## 4.2 Реалізація нейронної мережі

Відповідно до запропонованої у підрозділі 4.1 інформаційної технології вибрана за кожною метрикою оптимальна модель була заново ідентифікована вже на розширеному датасеті «train+valid», після чого зроблено прогнозування вже тестових даних та порівняння з реальними даними (рис. 4.2-4.7).

	name_model	type_data	r2_score	rmse	mape
11	KNeighbors Regressor	valid	-6.628865	5.516704	9.75295
8	ARIMA_manual	valid	-7.790947	5.921989	10.417483
14	Random Forest Regressor	valid	-7.682764	5.885437	10.451046
2	Prophet_5_days_3_order	valid	-8.614675	6.193227	10.852221
0	Prophet_4_days_3_order	valid	-9.270049	6.400825	11.071546

Рисунок 4.2 – Похибки і оптимальні параметри моделі Random Forest Regressor, за метриками r2\_score, RMSE та MAPE

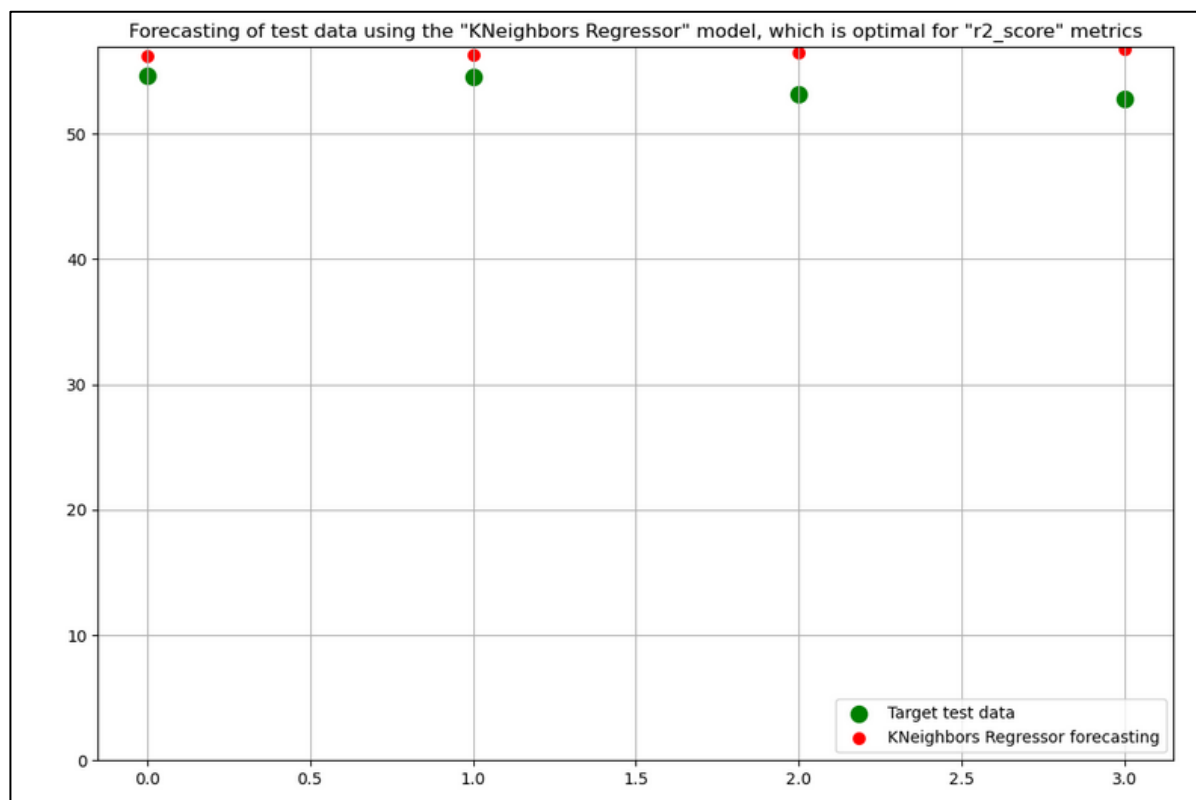


Рисунок 4.3 – Реальні тестові дані курсу криптовалюти Ethereum та їх прогноз за моделлю Random Forest Regressor, оптимальної за метрикою r2\_score



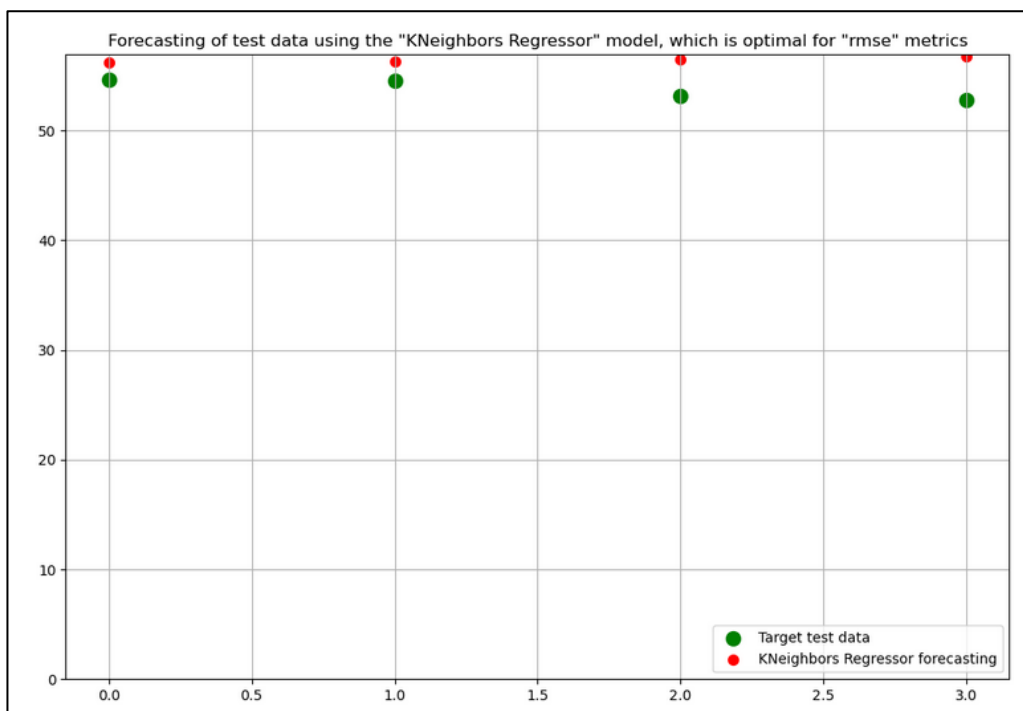


Рисунок 4.4 – Реальні тестові дані курсу криптовалюти Ethereum та їх прогноз за моделлю Random Forest Regressor, оптимальної за метрикою RMSE

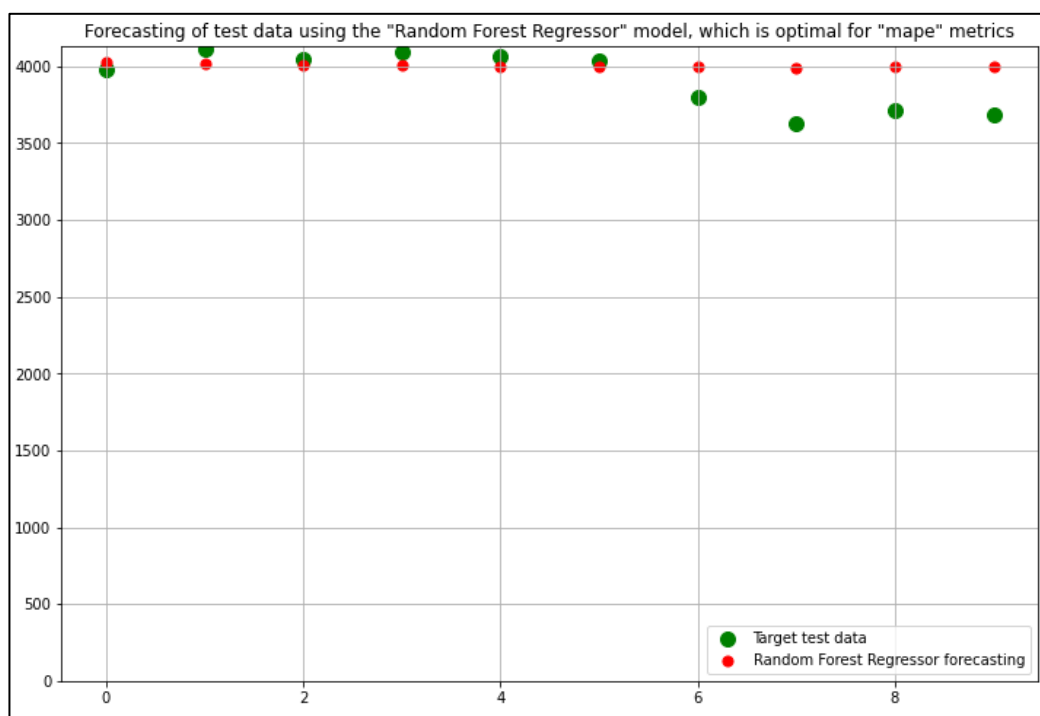


Рисунок 4.5 - Реальні тестові дані курсу криптовалюти Ethereum та їх прогноз за моделлю Random Forest Regressor, оптимальної за метрикою MAPE

	name_model	r2_score	rmse	mape	params
1	KNeighbors Regressor	-6.628865	5.516704	9.75295	{'leaf_size': 10, 'n_neighbors': 30}

Optimal model is 'KNeighbors Regressor' with parameters {'leaf\_size': 10, 'n\_neighbors': 30}

Tuning model 'KNeighbors Regressor'

KNeighbors Regressor

Рисунок 4.6 – Похибка і оптимальні параметри моделі KNeighbors Regressor, оптимальної за усіма трьома метриками, побудованої для тестових даних курсу цифрової валюти

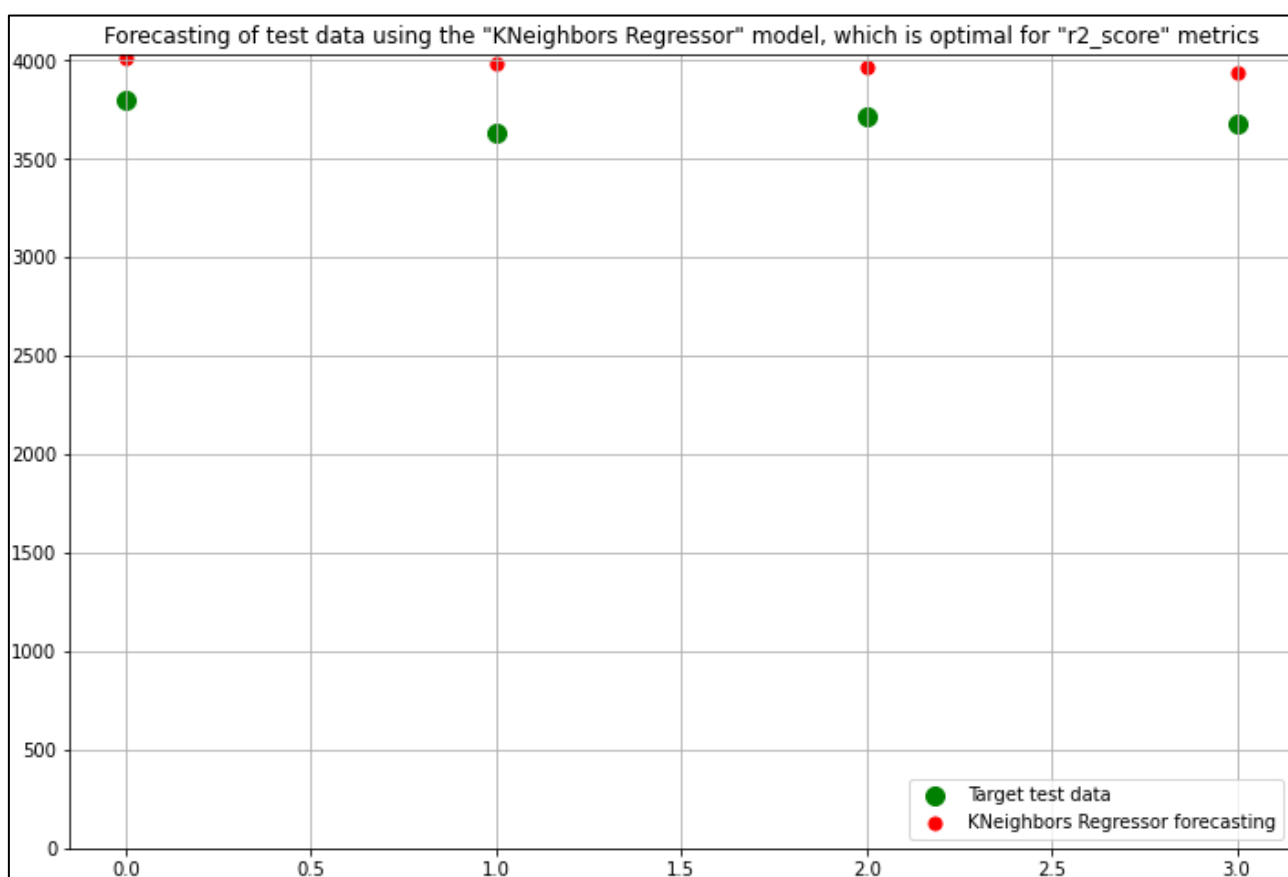


Рисунок 4.7 – Реальні тестові дані курсу криптовалюти Ethereum та їх прогноз за моделлю KNeighbors Regressor, оптимальної за усіма трьома метриками

Проаналізувавши дані з рисунків 4.5-4.7, можна зробити висновок, що прогноз на перших 4 доби виявився досить точним.

### 4.3 Тестування та аналіз результатів роботи програми прогнозування курсу криптовалюти ETHEREUM

Після успішного вирішення завдання машинного навчання з використанням різних ознак, стандартно проводиться аналіз важливості кожної ознаки. Після цього розумно повернутися до етапу "Feature engineering" для оптимізації та вдосконалення ознак, включаючи створення нових та вилучення з меншою важливістю. Однак основна увага при цьому приділяється видаленню менш суттєвих ознак та створенню нових, які можуть бути ще ефективнішими для точного прогнозування. За отримання обіцяючих результатів можна перейти до опису цінності кожної ознаки в кінцевому рішенні, що подробиць описано у підрозділі 4.2.

Зробимо оцінку важливості ознак отриманих оптимальною моделі «Random Forest Regressor» за кількома способами.

SHAP – це метод, який пов'язує розподіл кредитів з локальними тлумаченнями, використовуючи стандартні значення Шеплі з теорії ігор та пов'язані з ними доповнення. Його перевагами є зручне візуалізування та можливість оцінки для конкретної підвибірки даних [29].

Scikit-learn (Sklearn) - це бібліотека, широко використовувана у Python для машинного навчання. Вона надає підтримку для зменшення розмірності, виконання регресії, класифікації, вибору певної моделі, попередньої обробки даних і кластерного аналізу.

ELI5 - це пакет для мови програмування Python. Використовуючи ELI5, можна налаштовувати класифікатори машинного навчання і пояснювати їх прогнозування. Крім того, ELI5 надає підтримку для різних пакетів машинного навчання і фреймворків, таких як Sklearn, CatBoost, XGBoost, Keras та інші [30].

Використовуючи бібліотеку SHAP для оптимальної моделі, створеної за допомогою методу машинного навчання "Random Forest Regressor" для прогнозування курсу цифрової валюти Ethereum, була побудована діаграма важливості ознак, яку можна побачити на рисунку 4.8.

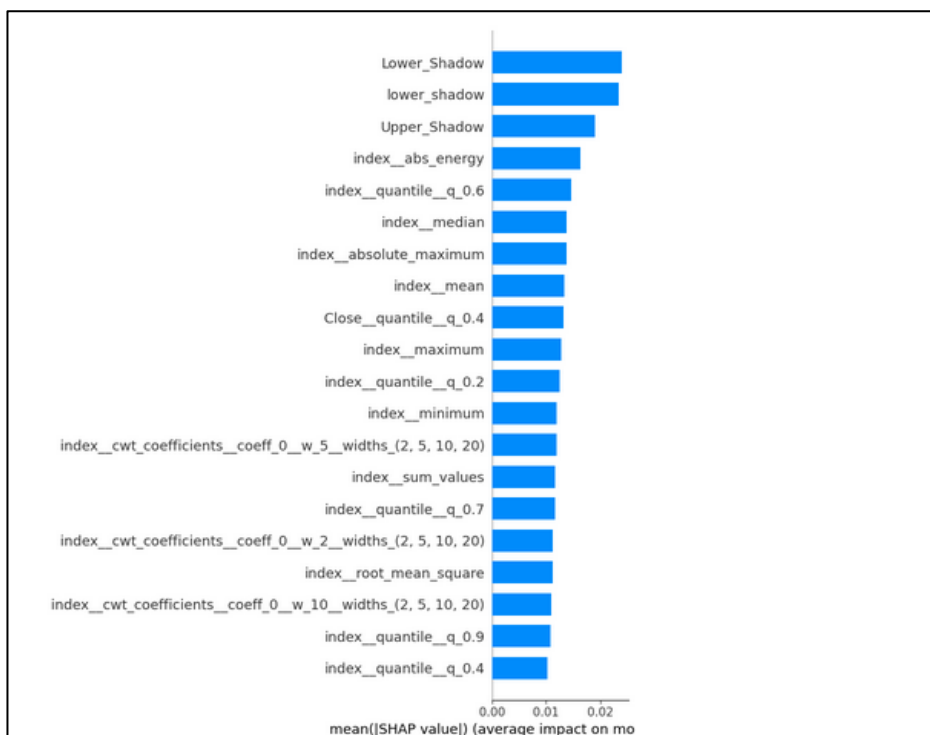


Рисунок 4.8 – Діаграма важливості ознак.

Використовуючи бібліотеку SHAP для оптимальної моделі, побудованої методом машинного навчання за допомогою «Random Forest Regressor» для прогнозування курсу цифрової валюти Ethereum, створено розгорнуту діаграму важливості ознак, яку можна побачити на рисунку 4.9.

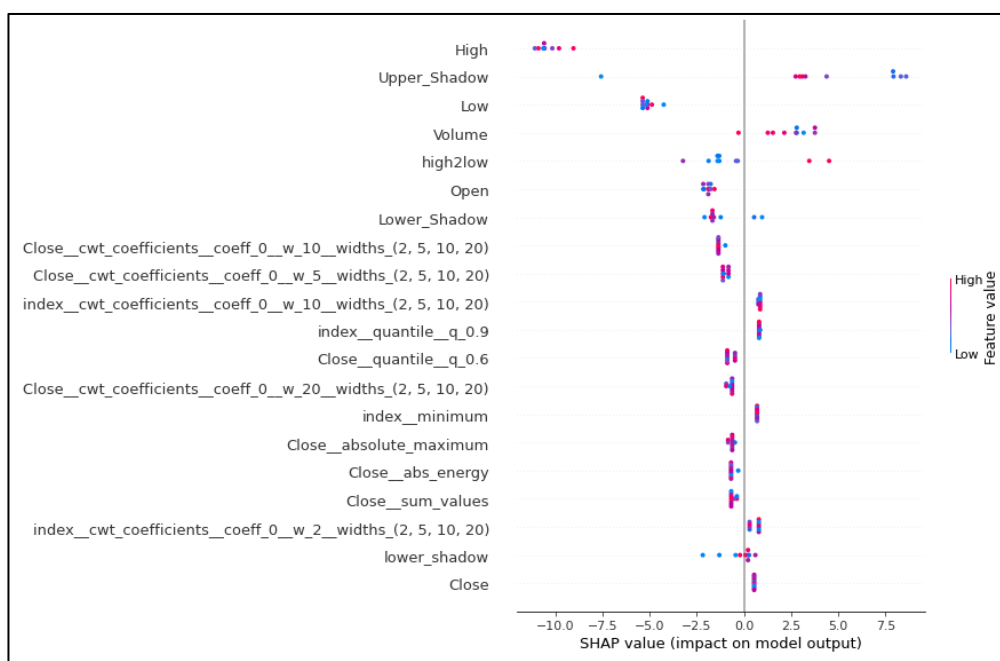


Рисунок 4.9 – Деталізована діаграма важливості ознак.

Використовуючи бібліотеку SHAP для оптимальної моделі, створеної методом машинного навчання за допомогою «Random Forest Regressor» для прогнозування курсу цифрової валюти Ethereum, побудовано діаграму типу "Force plot". На цьому графіку відображена важливість ознак, які впливають на прогнозовані значення на першу дату тестового датасету, і це представлено на рисунку 4.10.

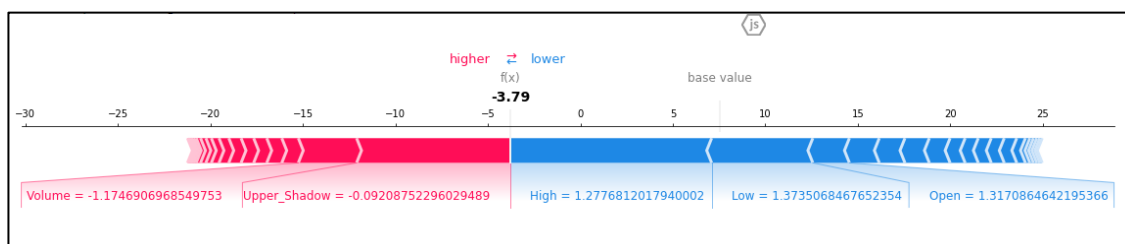


Рисунок 4.10 – Діаграма типу «Force plot» із важливістю ознак.

За допомогою бібліотеки ELI5 для оптимальної моделі, побудованої за методом машинного навчання "Random Forest Regressor" для моделювання курсу цифрової валюти Ethereum, створено діаграму перестановок із важливістю ознак. Цю діаграму представлено на рисунку 4.11.

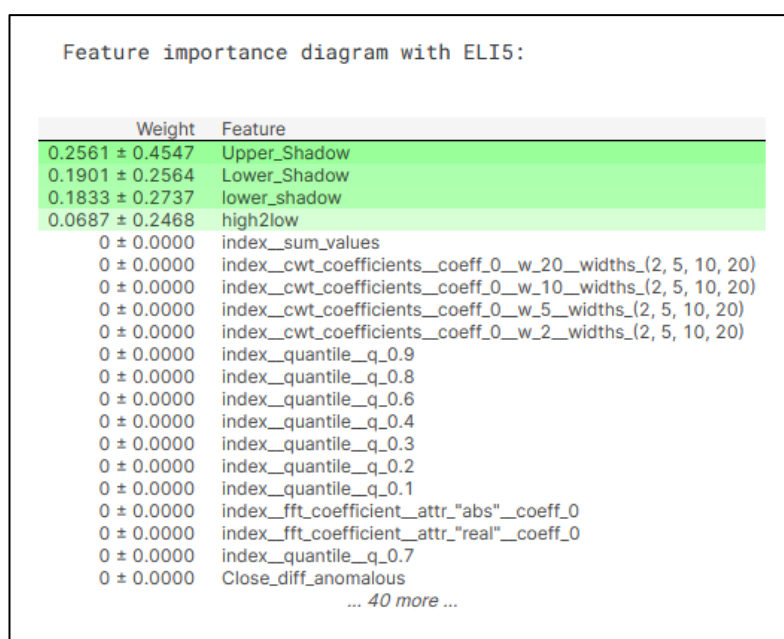


Рисунок 4.11 – Діаграма перестановок із важливістю ознак.

На основі поданих діаграм визначено, що найбільший вплив на прогноз курсу цифрової валюти Ethereum у моделі "KNeighbors Regressor" мають ознаки типу "High", "Upper\_Shadow" та "Low", в той час як ознаки "lower\_shadow" та "Close" мають менш значущий внесок. Діаграму важливості ознак для цієї моделі представлено на рисунку 4.12.

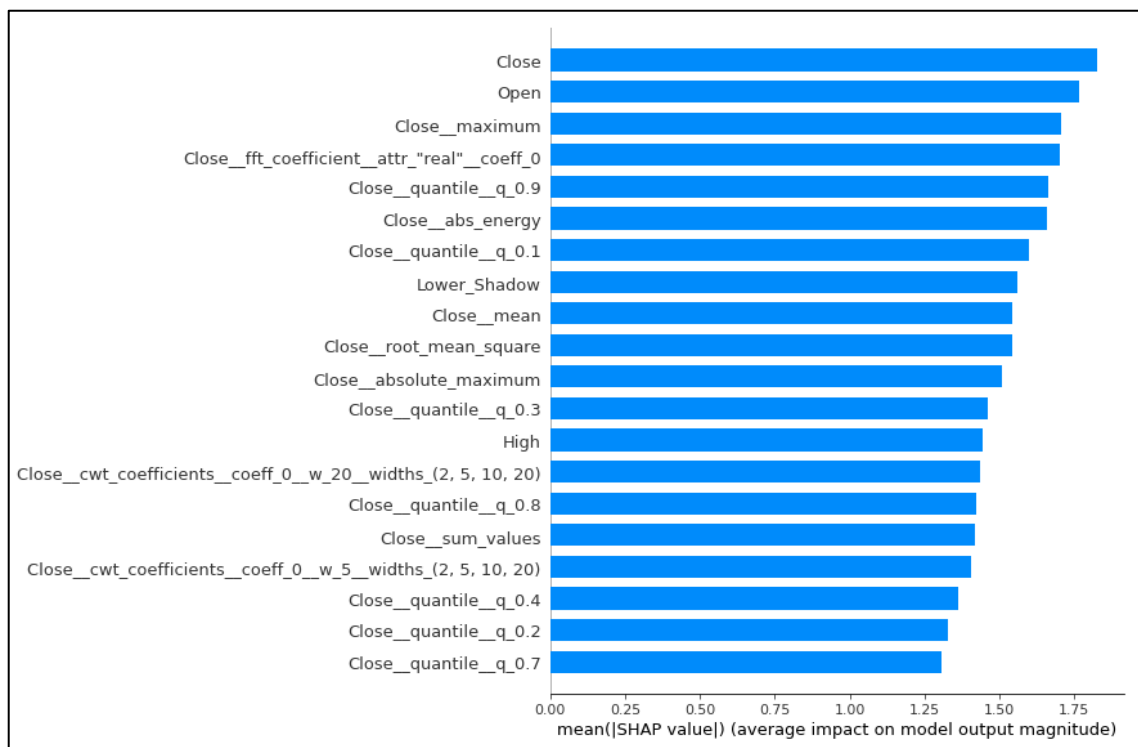


Рисунок 4.12 – Діаграма важливості ознак.

Також, за допомогою бібліотеки SHAP для оптимальної моделі, побудованої за методом машинного навчання «KNeighbors Regressor» для моделювання курсу цифрової валюти Ethereum, побудовано деталізовану діаграму важливості ознак і наведено на рисунку 4.13.

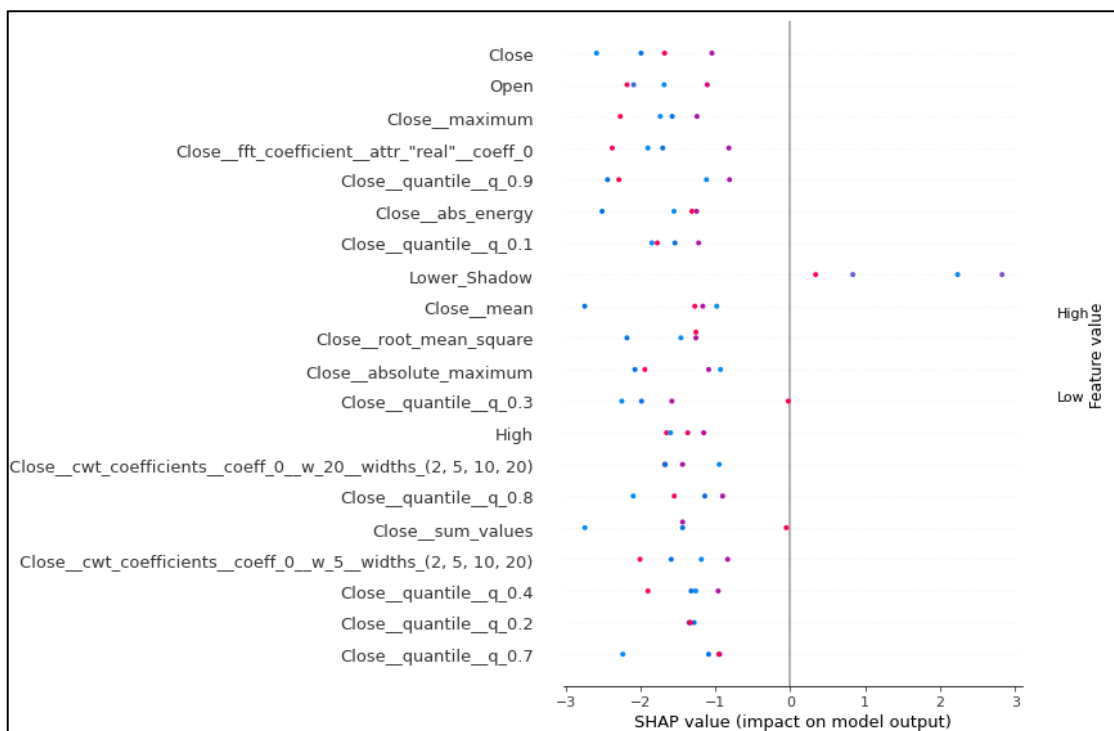


Рисунок 4.13 – Деталізована діаграма важливості ознак.

Використовуючи модуль «inspection» з бібліотеки Sklearn для оптимальної моделі побудованої за методом машинного навчання «KNeighbors Regressor» для моделювання курсу цифрової валюти Ethereum, побудовано діаграму значення коефіцієнтів із важливістю ознак і наведено на рисунку 4.14.

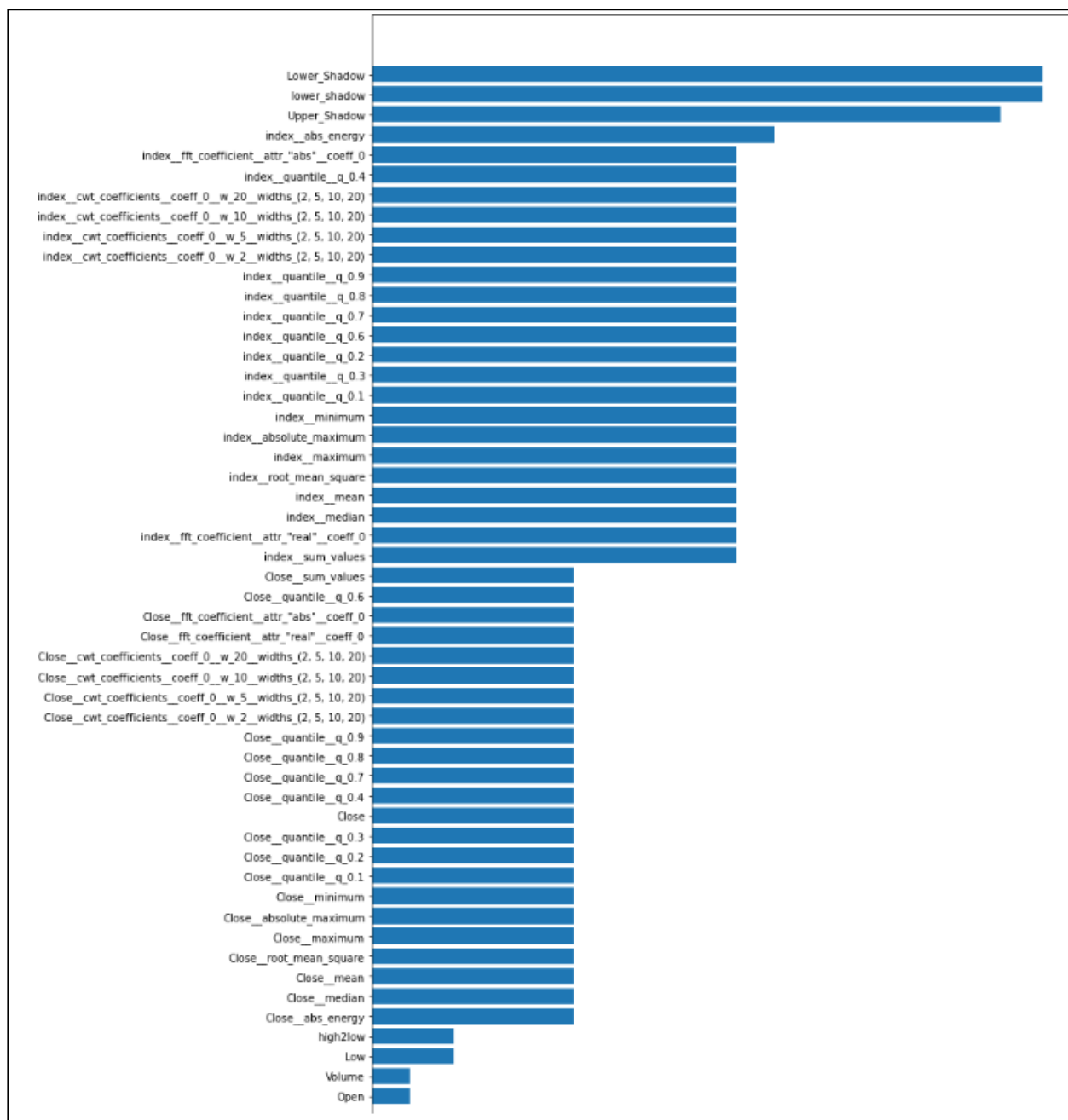


Рисунок 4.14 – Діаграма значення коефіцієнтів із важливістю ознак.

За допомогою бібліотеки SHAP для оптимальної моделі, побудованої за методом машинного навчання «KNeighbors Regressor» для моделювання курсу цифрової валюти Ethereum, побудовано діаграму типу «Force plot» із важливістю ознак, які вплинули на прогноз значення у першу дату тестового датасету і наведено на рисунку 4.15.



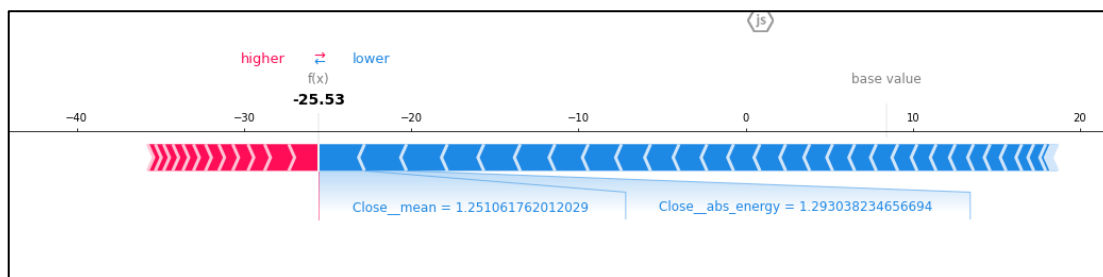


Рисунок 4.15 – Діаграма типу «Force plot» із важливістю ознак.

За допомогою бібліотеки ELI5 для оптимальної моделі, побудованої за методом машинного навчання «KNeighbors Regressor» для моделювання курсу цифрової валюти Ethereum, побудовано діаграму перестановок із цінністю ознак і наведено на рисунку 4.16.

Weight	Feature
0.1043 ± 0.0283	lower_shadow
0.0608 ± 0.0775	Lower_Shadow
0.0444 ± 0.0171	index_sum_values
0.0444 ± 0.0171	index_quantile_q_0.3
0.0409 ± 0.0140	index_quantile_q_0.7
0.0409 ± 0.0140	index_quantile_q_0.1
0.0404 ± 0.0426	index_fft_coefficient_attr_"abs"_coeff_0
0.0369 ± 0.0401	index_abs_energy
0.0369 ± 0.0401	index_fft_coefficient_attr_"real"_coeff_0
0.0306 ± 0.0229	Upper_Shadow
0.0299 ± 0.0299	index_quantile_q_0.2
0.0299 ± 0.0299	index_root_mean_square
0.0299 ± 0.0299	index_cwt_coefficients_coeff_0_w_20_widths_(2, 5, 10, 20)
0.0299 ± 0.0299	index_cwt_coefficients_coeff_0_w_5_widths_(2, 5, 10, 20)
0.0294 ± 0.0378	Close_abs_energy
0.0259 ± 0.0443	index_minimum
0.0259 ± 0.0443	Close_absolute_maximum
0.0259 ± 0.0443	index_quantile_q_0.8
0.0259 ± 0.0443	index_maximum
0.0259 ± 0.0443	index_absolute_maximum
... 40 more ...	

Рисунок 4.16 – Діаграма перестановок із цінністю ознак.

Проаналізувавши всі діаграми наведені вище, можна дійти до висновку, що для оптимальної моделі «KNeighbors Regressor» і має найбільшу цінність ознаки: «Close», «Open», «Close\_maximum», а найменшу цінність: «Close\_quantile\_q\_0.2», «Close\_quantile\_q\_0.7».

Для доведення факту досягнення поставленої в роботі мети – підвищення точності прогнозування курсу криптовалюти Ethereum – було проведено тестування роботи розробленої програми та програми-аналога, яка базується на LSTM (Long Short-Term Memory) модель у Deep Learning на прикладі прогнозу 2 діб криптовалюти Ethereum(тестова вибірка).

Створення мережі обчислювальним шляхом, щоб ми могли досліджувати різні мережеві архітектури, ми використали підмножину з 2000 балів (приблизно два дні), надано 50 вхідних даних прогнозувати спотову ціну на наступні 50 періодів.

Результати тестування подані у табл. 4.1.

Таблиця 4.1 – Параметри, отримані при тестуванні розробленої програми та програми-аналога

Програмний засіб	К-ть діб прогнозування	Набір перевірки з 2000 балів становив	Середній абсолютний відсоток похибки моделей, %	Середня квадратична похибка та середня абсолютна похибка для моделей
Модель LSTM	2	$4,2 \times 10^{-5}$	3,76%	(RMSE) 0.557 (MAE) 0.362
Розроблена програма	2	$1,7 \times 10^{-5}$	3,05%	(RMSE) 0.423 (MAE) 0.320

Обидві моделі відстежують фактичне значення даних досить точно. На малюнку 4.17 показано MSE для кожної моделі кожна точка. Хоча помилки виглядають схожими, розроблена програма схожа явно точніше, ніж модель LSTM.

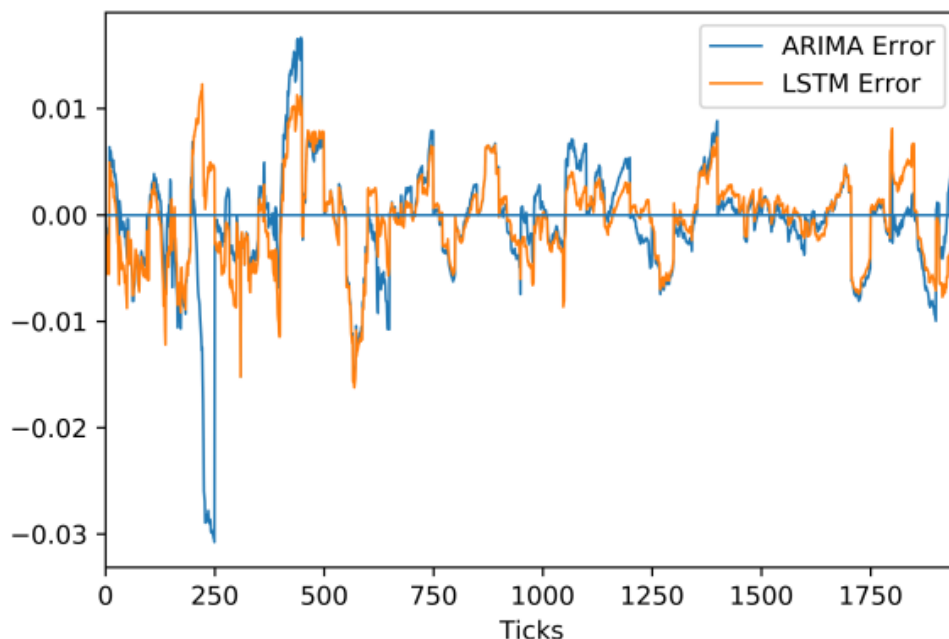


Рисунок 4.17 – Прогнозування набору перевірки з 2000 балів за допомогою моделі ARIMA і LSTM

Набір перевірки 2000 балів був  $1.7 \times 10^{-5}$  для розробленої програми і  $4.2 \times 10^{-5}$  для LSTM. Крім того, середній абсолютний відсоток (MAPE) для моделей ARIMA і LSTM становила 3,76% і 3,05% відповідно. У таблиці 4.2 ми повідомляємо про середню квадратичну похибку або (RMSE) і середньої абсолютної похибки (MAE). В обох випадках модель ARIMA точніша, ніж модель LSTM.

Таблиця 4.2 – Середньоквадратична похибка (RMSE) та Середня абсолютна похибка (MAE) моделей

	LSTM	ARIMA
Середньоквадратична похибка (RMSE)	0.557	0.423
Середня абсолютна похибка (MAE)	0.362	0.320

Використання розробленої програми на основі ARIMA з 2000 точок даних про історичні спотові ціни, ми спостерігали зменшення похибки приблизно піввідсотка за тик для цього типу даних. Наші передбачення можуть дозволити користувачеві точно визначити і оцінити в середньому на півцента ближче до фактичної ціни пропозиції. Після збільшення вибірки користувач із 12-годинним виділенням `thes3.2xlargeinstancetype` може зменшити очікувану вартість приблизно на 5%.

Таким чином, після проведення порівняння розробленої програми прогнозування курсу криптовалюти Ethereum з програмою-аналогом, можна зробити висновок, що розроблена програма має переваги у достовірності прогнозування курсу криптовалюти Ethereum.

#### **4.5 Висновок до розділу 4**

У четвертому розділі розроблено інформаційну технологію аналізу та прогнозування курсу криптовалюти Ethereum. Відповідно до розробленої інформаційної технології здійснено прогнозування тестових даних та порівняння результатів прогнозування з реальними даними. Створена технологія для прогнозування курсу криптовалют, заснована на комплексному відборі ознак, відрізняється системним підходом до цього процесу. Проведено аналіз внутрішніх та зовнішніх факторів, що можуть впливати на криптовалютний ринок. Результати показали, що окрім базових ознак, важливо враховувати вторинні, які отримані шляхом математичних операцій чи алгоритмічних трансформацій. Систематизовано та формалізовано характеристики криптовалют, що сприяє покращенню вибору ознак. Запропоновано математичну модель для формування факторів, яка базується на генерації вторинних ознак та відборі найрелевантніших. Технологія включає синтез нових ознак на основі існуючих, забезпечуючи уникнення перенавчання моделі та покращення її узагальнюючих властивостей. Проведено аналіз важливості ознак при прогнозуванні на 10 діб, який показав, що найбільш важливими є ознаки

«Lower Shadow», «Upper\_Shadow» та «lower shadow», найменш важливими є ознаки «lower\_shadow» та «Close». Найбільш ефективна модель буде «KNeighbors Regressor» з  $r^2\_score$  -6.628865,  $rmase$  5.516704 та  $mape$  9.75295. В аналізі важливості ознак при прогнозуванні на 4 доби для оптимальної моделі «KNeighbors Regressor» найбільшу цінність мають ознаки: «Close», «Open», «Close\_maximum», а найменшу: «Close\_quantile\_q\_0.2», «Close\_quantile\_q\_0.7».

Крім цього, розроблена програма має також вищу достовірність прогнозування курсу криптовалюти ніж програма аналог. Набір перевірки 2000 балів був  $1.7 \times 10^{-5}$  для розробленої програми і  $4.2 \times 10^{-5}$  для LSTM. Крім того, середній абсолютний відсоток (MAPE) для моделей ARIMA і LSTM становила 3,76% і 3,05% відповідно. . Після збільшення вибірки користувач із 12-годинним виділенням `thes3.2xlargeinstancetype` може зменшити очікувану вартість приблизно на 5%.

Тобто мета магістерської кваліфікаційної роботи досягнута – достовірність прогнозування курсу криптовалюти Ethereum підвищена на піввідсотка, що може допомогти користувачеві приймати рішення на спотовому ринку.

## 5 ЕКОНОМІЧНА ЧАСТИНА

Науково-технічна розробка має право на існування та впровадження, якщо вона відповідає вимогам часу, як в напрямку науково-технічного прогресу та і в плані економіки. Тому для науково-дослідної роботи необхідно оцінювати економічну ефективність результатів виконаної роботи.

Магістерська кваліфікаційна робота з розробки та дослідження «Інформаційна технологія прогнозування курсу криптовалюти ETHEREUM» відноситься до науково-технічних робіт, які орієнтовані на виведення на ринок (або рішення про виведення науково-технічної розробки на ринок може бути прийнято у процесі проведення самої роботи), тобто коли відбувається так звана комерціалізація науково-технічної розробки. Цей напрямок є пріоритетним, оскільки результатами розробки можуть користуватися інші споживачі, отримуючи при цьому певний економічний ефект. Але для цього потрібно знайти потенційного інвестора, який би взявся за реалізацію цього проекту і переконати його в економічній доцільності такого кроку.

Для наведеного випадку нами мають бути виконані такі етапи робіт:

- 1) проведено комерційний аудит науково-технічної розробки, тобто встановлення її науково-технічного рівня та комерційного потенціалу;
- 2) розраховано витрати на здійснення науково-технічної розробки;
- 3) розрахована економічна ефективність науково-технічної розробки у випадку її впровадження і комерціалізації потенційним інвестором і проведено обґрунтування економічної доцільності комерціалізації потенційним інвестором.

### **5.1 Проведення комерційного та технологічного аудиту науково-технічної розробки**

Метою проведення комерційного і технологічного аудиту дослідження за темою «Інформаційна технологія прогнозування курсу криптовалюти

ETHEREUM» є оцінювання науково-технічного рівня та рівня комерційного потенціалу розробки, створеної в результаті науково-технічної діяльності.

Оцінювання науково-технічного рівня розробки та її комерційного потенціалу рекомендується здійснювати із застосуванням 5-ти бальної системи оцінювання за 12-ма критеріями, наведеними в табл. 5.1.

Таблиця 5.1 – Рекомендовані критерії оцінювання науково-технічного рівня і комерційного потенціалу розробки та бальна оцінка

Бали (за 5-ти бальною шкалою)					
	0	1	2	3	4
Технічна здійсненність концепції					
1	Достовірність концепції не підтверджена	Концепція підтверджена експертними висновками	Концепція підтверджена розрахунками	Концепція перевірена на практиці	Перевірено працездатність продукту в реальних
Ринкові переваги (недоліки)					
2	Багато аналогів на малому ринку	Мало аналогів на малому ринку	Кілька аналогів на великому ринку	Один аналог на великому ринку	Продукт не має аналогів на великому ринку
3	Ціна продукту значно вища за ціни аналогів	Ціна продукту дещо вища за ціни аналогів	Ціна продукту приблизно дорівнює цінам аналогів	Ціна продукту дещо нижче за ціни аналогів	Ціна продукту значно нижче за ціни аналогів
4	Технічні та споживчі властивості продукту значно гірші,	Технічні та споживчі властивості продукту трохи гірші, ніж в	Технічні та споживчі властивості продукту на рівні аналогів	Технічні та споживчі властивості продукту трохи кращі, ніж в	Технічні та споживчі властивості продукту значно кращі,
5	Експлуатаційні витрати значно вищі, ніж в аналогів	Експлуатаційні витрати дещо вищі, ніж в аналогів	Експлуатаційні витрати на рівні експлуатаційних витрат аналогів	Експлуатаційні витрати трохи нижчі, ніж в аналогів	Експлуатаційні витрати значно нижчі, ніж в аналогів
Ринкові перспективи					
6	Ринок малий і не має позитивної	Ринок малий, але має позитивну	Середній ринок з позитивною динамікою	Великий стабільний ринок	Великий ринок з позитивною динамікою
7	Активна конкуренція великих компаній на	Активна конкуренція	Помірна конкуренція	Незначна конкуренція	Конкурентів немає
Практична здійсненність					

8	Відсутні фахівці як з технічної, так і з комерційної реалізації ідеї	Необхідно наймати фахівців або витратити значні кошти та час на навчання наявних	Необхідне незначне навчання фахівців та збільшення їх штату	Необхідне незначне навчання фахівців	Є фахівці з питань як з технічної, так і з комерційної реалізації ідеї
9	Потрібні значні фінансові ресурси, які відсутні. Джерела фінансування ідеї відсутні	Потрібні незначні фінансові ресурси. Джерела фінансування відсутні	Потрібні значні фінансові ресурси. Джерела фінансування є	Потрібні незначні фінансові ресурси. Джерела фінансування є	Не потребує додаткового фінансування
10	Необхідна розробка нових матеріалів	Потрібні матеріали, що використовуються у військово-промисловому комплексі	Потрібні дорогі матеріали	Потрібні досяжні та дешеві матеріали	Всі матеріали для реалізації ідеї відомі та давно використовуються у
11	Термін реалізації ідеї більший за 10 років	Термін реалізації ідеї більший за 5 років. Термін окупності інвестицій більше 10-ти	Термін реалізації ідеї від 3-х до 5-ти років. Термін окупності інвестицій більше 5-ти років	Термін реалізації ідеї менше 3-х років. Термін окупності інвестицій від 3-х до 5-ти	Термін реалізації ідеї менше 3-х років. Термін окупності інвестицій менше 3-х років
12	Необхідна розробка регламентних документів та отримання великої кількості дозвільних документів на виробництво та реалізацію	Необхідно отримання великої кількості дозвільних документів на виробництво та реалізацію продукту, що вимагає значних коштів	Процедура отримання дозвільних документів для виробництва та реалізації продукту вимагає незначних коштів та часу	Необхідно тільки повідомлення відповідним органам про виробництво та реалізацію продукту	Відсутні будь-які регламентні обмеження на виробництво та реалізацію продукту

Результати оцінювання науково-технічного рівня та комерційного потенціалу науково-технічної розробки потрібно звести до таблиці.



Таблиця 5.2 – Результати оцінювання науково-технічного рівня і комерційного потенціалу розробки експертами

Критерії	Експерт (ПІБ, посада)		
	1	2	3
	Бали:		
1. Технічна здійсненність концепції	4	4	4
2. Ринкові переваги (наявність аналогів)	3	3	3
3. Ринкові переваги (ціна продукту)	4	4	3
4. Ринкові переваги (технічні властивості)	3	3	4
5. Ринкові переваги (експлуатаційні витрати)	2	2	3
6. Ринкові перспективи (розмір ринку)	3	3	3
7. Ринкові перспективи (конкуренція)	3	3	3
8. Практична здійсненність (наявність фахівців)	5	5	5
9. Практична здійсненність (наявність фінансів)	2	3	2
10. Практична здійсненність (необхідність нових матеріалів)	4	5	5
11. Практична здійсненність (термін реалізації)	3	4	5
12. Практична здійсненність (розробка документів)	4	5	4
Сума балів	40	44	44
Середньоарифметична сума балів $СБ_c$	42,7		

За результатами розрахунків, наведених в таблиці 5.2, зробимо висновок щодо науково-технічного рівня і рівня комерційного потенціалу розробки. При цьому використаємо рекомендації, наведені в табл. 5.3.

Таблиця 5.3 – Науково-технічні рівні та комерційні потенціали розробки

Середньоарифметична сума балів $СБ_c$ , розрахована на основі висновків експертів	Науково-технічний рівень та комерційний потенціал розробки
41...48	Високий
31...40	Вище середнього
21...30	Середній
11...20	Нижче середнього
0...10	Низький

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою «Інформаційна технологія прогнозування курсу криптовалюти ETHEREUM» становить 42,7 бала, що, відповідно до таблиці 4.3, свідчить про

комерційну важливість проведення даних досліджень (рівень комерційного потенціалу розробки високий).

## 5.2 Розрахунок узагальненого коефіцієнта якості розробки

Окрім комерційного аудиту розробки доцільно також розглянути технічний рівень якості розробки, розглянувши її основні технічні показники. Ці показники по-різному впливають на загальну якість проектної розробки.

Узагальнений коефіцієнт якості ( $B_n$ ) для нового технічного рішення розрахуємо за формулою:

$$B_n = \sum_{i=1}^k \alpha_i \cdot \beta_i \quad (5.1)$$

де  $k$  – кількість найбільш важливих технічних показників, які впливають на якість нового технічного рішення;

$\alpha_i$  – коефіцієнт, який враховує питому вагу  $i$ -го технічного показника в загальній якості розробки. Коефіцієнт  $\alpha_i$  визначається експертним шляхом і при цьому має

виконуватись умова  $\sum_{i=1}^k \alpha_i = 1$ ;

$\beta_i$  – відносне значення  $i$ -го технічного показника якості нової розробки.

Відносні значення  $\beta_i$  для різних випадків розраховуємо за такими формулами: для показників, зростання яких вказує на підвищення в лінійній залежності якості нової розробки:

$$\beta_i = \frac{I_{ni}}{I_{ai}}, \quad (5.2)$$

де  $I_{ni}$  та  $I_{na}$  – чисельні значення конкретного  $i$ -го технічного показника якості відповідно для нової розробки та аналога;

для показників, зростання яких вказує на погіршення в лінійній залежності якості нової розробки:

$$\beta_i = \frac{I_{ai}}{I_{ni}} ; \quad (5.3)$$

Використовуючи наведені залежності можемо проаналізувати та порівняти техніко-економічні характеристики аналогу та розробки на основі отриманих наявних та проектних показників, а результати порівняння зведемо до таблиці 5.4.

Таблиця 5.4 – Порівняння основних параметрів розробки та аналога.

Показники (параметри)	Одиниця вимірювання	Аналог	Проектований пристрій	Відношення параметрів нової розробки до аналога	Питома вага показника
Швидкодія	с	26	12	2,16	0,25
Швидкість доступу до даних	с	580	260	2,23	0,3
Кількість обробки запитів на хвилину	од	60	100	1,67	0,15
Місце на диску	Мб	1000	145	6,9	0,05
Потреба в оперативній пам'яті	Гб	16	4	4	0,25

Узагальнений коефіцієнт якості ( $B_n$ ) для нового технічного рішення складе:

$$B_n = \sum_{i=1}^k \alpha_i \cdot \beta_i = 2,16 \cdot 0,25 + 2,23 \cdot 0,3 + 1,67 \cdot 0,15 + 6,9 \cdot 0,05 + 4 \cdot 0,25 = 2,80.$$

Отже за технічними параметрами, згідно узагальненого коефіцієнту якості розробки, науково-технічна розробка переважає існуючі аналоги приблизно в 2,80 рази.

### 5.3 Розрахунок витрат на проведення науково-дослідної роботи

Витрати, пов'язані з проведенням науково-дослідної роботи на тему «Інформаційна технологія прогнозування курсу криптовалюти ETHEREUM», під час планування, обліку і калькулювання собівартості науково-дослідної роботи групуємо за відповідними статтями.

#### 5.3.1 Витрати на оплату праці

До статті «Витрати на оплату праці» належать витрати на виплату основної та додаткової заробітної плати керівникам відділів, лабораторій, секторів і груп, науковим, інженерно-технічним працівникам, конструкторам, технологам, креслярам, копіювальникам, лаборантам, робітникам, студентам, аспірантам та іншим працівникам, безпосередньо зайнятим виконанням конкретної теми, обчисленої за посадовими окладами, відрядними розцінками, тарифними ставками згідно з чинними в організаціях системами оплати праці.

#### Основна заробітна плата дослідників

Витрати на основну заробітну плату дослідників ( $Z_o$ ) розраховуємо у відповідності до посадових окладів працівників, за формулою

$$Z_o = \sum_{i=1}^k \frac{M_{ni} \cdot t_i}{T_p}, \quad (5.4)$$

де  $k$  – кількість посад дослідників залучених до процесу досліджень;

$M_{ni}$  – місячний посадовий оклад конкретного дослідника, грн;

$t_i$  – число днів роботи конкретного дослідника, дн.;

$T_p$  – середнє число робочих днів в місяці,  $T_p=21$  дні.

$$Z_o = 21350,00 \cdot 45 / 21 = 45750,00 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці.

Таблиця 5.5 – Витрати на заробітну плату дослідників

Найменування посади	Місячний посадовий оклад, грн	Оплата за робочий день, грн	Число днів роботи	Витрати на заробітну плату, грн
Керівник проекту	21350,00	1016,67	45	45750,00
Консультант (аналітик ринку криптовалют)	25500,00	1214,29	10	12142,86
Інженер-дослідник (програміст)	17000,00	809,52	42	34000,00
Фахівець 1-ї категорії	7300,00	347,62	42	14600,00
Всього				106492,86

### Основна заробітна плата робітників

Витрати на основну заробітну плату робітників ( $Z_p$ ) за відповідними найменуваннями робіт НДР на тему «Інформаційна технологія прогнозування курсу криптовалюти ETHEREUM» розраховуємо за формулою:

$$Z_p = \sum_{i=1}^n C_i \cdot t_i, \quad (5.5)$$

де  $C_i$  – погодинна тарифна ставка робітника відповідного розряду, за виконану відповідну роботу, грн/год;

$t_i$  – час роботи робітника при виконанні визначеної роботи, год.

Погодинну тарифну ставку робітника відповідного розряду  $C_i$  можна визначити за формулою:

$$C_i = \frac{M_M \cdot K_i \cdot K_c}{T_p \cdot t_{зм}}, \quad (5.6)$$

де  $M_M$  – розмір прожиткового мінімуму працездатної особи, або мінімальної місячної заробітної плати (в залежності від діючого законодавства), приймемо  $M_M=6700,00$  грн;

$K_i$  – коефіцієнт міжкваліфікаційного співвідношення для встановлення тарифної ставки робітнику відповідного розряду (табл. Б.2, додаток Б);

$K_c$  – мінімальний коефіцієнт співвідношень місячних тарифних ставок робітників першого розряду з нормальними умовами праці виробничих об'єднань і підприємств до законодавчо встановленого розміру мінімальної заробітної плати.

$T_p$  – середнє число робочих днів в місяці, приблизно  $T_p = 21$  дн;

$t_{зм}$  – тривалість зміни, год.

$$C_l = 6700,00 \cdot 1,10 \cdot 1,65 / (21 \cdot 8) = 72,38 \text{ грн.}$$

$$Z_{pl} = 72,38 \cdot 6,75 = 488,59 \text{ грн.}$$

Таблиця 5.6 – Величина витрат на основну заробітну плату робітників

Найменування робіт	Тривалість роботи, год	Розряд роботи	Тарифний коефіцієнт	Погодинна тарифна ставка, грн	Величина оплати на робітника грн
Встановлення обладнання для проведення досліджень	6,75	2	1,10	72,38	488,59
Інсталяція програмного забезпечення розробки програмного забезпечення	5,30	3	1,35	88,83	470,82
Встановлення цифрових обчислювальних систем та мережевого доступу	5,00	4	1,50	98,71	493,53
Відлагодження програмних модулів аналітичного дослідження	8,00	5	1,70	111,87	894,93
Підготовка тестового дослідження наявної бази даних	4,60	3	1,35	88,83	408,64

Формування бази даних результатів випробування	11,00	3	1,35	88,83	977,18
Всього					3733,69

Додаткова заробітна плата дослідників та робітників

Додаткову заробітну плату розраховуємо як 10 ... 12% від суми основної заробітної плати дослідників та робітників за формулою:

$$Z_{\text{доо}} = (Z_o + Z_p) \cdot \frac{H_{\text{доо}}}{100\%}, \quad (5.7)$$

де  $H_{\text{доо}}$  – норма нарахування додаткової заробітної плати. Прийmemo 12%.

$$Z_{\text{доо}} = (106492,86 + 3733,69) \cdot 12 / 100\% = 13227,19 \text{ грн.}$$

### 5.3.2 Відрахування на соціальні заходи

Нарахування на заробітну плату дослідників та робітників розраховуємо як 22% від суми основної та додаткової заробітної плати дослідників і робітників за формулою:

$$Z_n = (Z_o + Z_p + Z_{\text{доо}}) \cdot \frac{H_{\text{зн}}}{100\%} \quad (5.8)$$

де  $H_{\text{зн}}$  – норма нарахування на заробітну плату. Приймаємо 22%.

$$Z_n = (106492,86 + 3733,69 + 13227,19) \cdot 22 / 100\% = 27159,82 \text{ грн.}$$

### 5.3.3 Сировина та матеріали

До статті «Сировина та матеріали» належать витрати на сировину, основні та допоміжні матеріали, інструменти, пристрої та інші засоби і предмети праці,

які придбані у сторонніх підприємств, установ і організацій та витрачені на проведення досліджень за темою «Інформаційна технологія прогнозування курсу криптовалюти ETHEREUM».

Витрати на матеріали ( $M$ ), у вартісному вираженні розраховуються окремо по кожному виду матеріалів за формулою:

$$M = \sum_{j=1}^n H_j \cdot C_j \cdot K_j - \sum_{j=1}^n B_j \cdot C_{ej}, \quad (5.9)$$

де  $H_j$  – норма витрат матеріалу  $j$ -го найменування, кг;

$n$  – кількість видів матеріалів;

$C_j$  – вартість матеріалу  $j$ -го найменування, грн/кг;

$K_j$  – коефіцієнт транспортних витрат, ( $K_j = 1,1 \dots 1,15$ );

$B_j$  – маса відходів  $j$ -го найменування, кг;

$C_{ej}$  – вартість відходів  $j$ -го найменування, грн/кг.

$$M_1 = 2,0 \cdot 201,00 \cdot 1 - 0 \cdot 0 = 402,00 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці.

Таблиця 5.7 – Витрати на матеріали

Найменування матеріалу, марка, тип, сорт	Ціна за 1 кг, грн	Норма витрат, кг	Величина відходів, кг	Ціна відходів, грн/кг	Вартість витраченого матеріалу, грн
Папір канцелярський офісний (A4)	201,00	2,0	0	0	402,00
Папір для заміток (A5)	122,00	3,0	0	0	366,00
Начиння канцелярське	187,00	3,0	0	0	561,00
Органайзер офісний	235,00	3,0	0	0	705,00
Картридж для принтера	2200,00	2,0	0	0	4400,00



## Продовження таблиці 5.7

Диск оптичний	32,00	4,0	0	0	128,00
USB-пам'ять Microtech 64 GB	195,00	2,0	0	0	390,00
Тека для паперів	89,90	3,0	0	0	269,70
Всього					7221,70

## 5.3.4 Розрахунок витрат на комплектуючі

Витрати на комплектуючі ( $K_6$ ), які використовують при проведенні НДР на тему «Інформаційна технологія прогнозування курсу криптовалюти ETHEREUM», розраховуємо, згідно з їхньою номенклатурою, за формулою:

$$K_6 = \sum_{j=1}^n H_j \cdot C_j \cdot K_j \quad (5.10)$$

де  $H_j$  – кількість комплектуючих  $j$ -го виду, шт.;

$C_j$  – покупна ціна комплектуючих  $j$ -го виду, грн;

$K_j$  – коефіцієнт транспортних витрат, ( $K_j = 1, 1 \dots 1, 15$ ).

$$K_6 = 1 \cdot 1689,00 \cdot 1 = 1689,00 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці.

Таблиця 5.8 – Витрати на комплектуючі

Найменування комплектуючих	Кількість, шт.	Ціна за штуку, грн	Сума, грн
Зовнішній жорсткий диск	1	1689,00	1689,00
Концентратор	1	1110,00	1110,00
Кабель для передачі даних USB	1	485,00	485,00
Всього			3284,00

## 5.3.5 Спецустаткування для наукових (експериментальних) робіт

Витрати на придбання спецустаткування відсутні

### 5.3.6 Програмне забезпечення для наукових (експериментальних) робіт

До статті «Програмне забезпечення для наукових (експериментальних) робіт» належать витрати на розробку та придбання спеціальних програмних засобів і програмного забезпечення, (програм, алгоритмів, баз даних) необхідних для проведення досліджень, також витрати на їх проектування, формування та встановлення.

Балансову вартість програмного забезпечення розраховуємо за формулою:

$$B_{npz} = \sum_{i=1}^k C_{inprz} \cdot C_{npz.i} \cdot K_i, \quad (5.11)$$

де  $C_{inprz}$  – ціна придбання одиниці програмного засобу даного виду, грн;

$C_{npz.i}$  – кількість одиниць програмного забезпечення відповідного найменування, які придбані для проведення досліджень, шт.;

$K_i$  – коефіцієнт, що враховує інсталяцію, налагодження програмного засобу тощо, ( $K_i = 1, 10 \dots 1, 12$ );

$k$  – кількість найменувань програмних засобів.

$$B_{npz} = 4890,00 \cdot 1 \cdot 1,05 = 5134,50 \text{ грн.}$$

Отримані результати зведемо до таблиці:

Таблиця 5.9 – Витрати на придбання програмних засобів по кожному виду

Найменування програмного засобу	Кількість, шт	Ціна за одиницю, грн	Вартість, грн
Програмне забезпечення розробки Kaggle	1	4890,00	5134,50
Всього			5134,50

### 5.3.7 Амортизація обладнання, програмних засобів та приміщень

В спрощеному вигляді амортизаційні відрахування по кожному виду обладнання, приміщень та програмному забезпеченню тощо, розраховуємо з використанням прямолінійного методу амортизації за формулою:

$$A_{обл} = \frac{Ц_б}{T_в} \cdot \frac{t_{вик}}{12}, \quad (5.12)$$

де  $Ц_б$  – балансова вартість обладнання, програмних засобів, приміщень тощо, які використовувались для проведення досліджень, грн;

$t_{вик}$  – термін використання обладнання, програмних засобів, приміщень під час досліджень, місяців;

$T_в$  – строк корисного використання обладнання, програмних засобів, приміщень тощо, років.

$$A_{обл} = (52399,00 \cdot 2) / (3 \cdot 12) = 2911,06 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці.

Таблиця 5.10 – Амортизаційні відрахування по кожному виду обладнання

Найменування обладнання	Балансова вартість, грн	Строк корисного використання, років	Термін використання обладнання, місяців	Амортизаційні відрахування, грн
Персональний комп'ютер Ноутбук HP Pavilion 15-e001	52399,00	3	2	2911,06
Робоче місце інженера-розробника (дослідника програмного забезпечення)	6500,00	5	2	216,67

Продовження таблиці 5.10

Графічні пристрої виводу інформації	7500,00	4	2	312,50
Офісна оргтехніка	8499,00	4	2	354,13
Приміщення лабораторії розробки та дослідження	410000,00	25	2	2733,33
ОС Windows 11	5800,00	2	2	483,33
Прикладний пакет Microsoft Office 2021 Professional Plus	5100,00	2	2	425,00
Мережеве обладнання передачі цифрових даних	7599,00	5	2	253,30
Всього				7689,31

## 5.3.8 Паливо та енергія для науково-виробничих цілей

Витрати на силову електроенергію ( $B_e$ ) розраховуємо за формулою:

$$B_e = \sum_{i=1}^n \frac{W_{yi} \cdot t_i \cdot C_e \cdot K_{eni}}{\eta_i}, \quad (5.13)$$

де  $W_{yi}$  – встановлена потужність обладнання на визначеному етапі розробки, кВт;

$t_i$  – тривалість роботи обладнання на етапі дослідження, год;

$C_e$  – вартість 1 кВт-години електроенергії, грн; (вартість електроенергії визначається за даними енергопостачальної компанії), прийmemo  $C_e = 7,50$  грн;

$K_{eni}$  – коефіцієнт, що враховує використання потужності,  $K_{eni} < 1$ ;

$\eta_i$  – коефіцієнт корисної дії обладнання,  $\eta_i < 1$ .

$$B_e = 0,40 \cdot 480,0 \cdot 7,50 \cdot 0,95 / 0,97 = 1440,00 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці.

Таблиця 5.11 – Витрати на електроенергію

Найменування обладнання	Встановлена потужність, кВт	Тривалість роботи, год	Сума, грн
Персональний комп'ютер Ноутбук HP Pavilion 15- e001	0,40	480,0	1440,00
Робоче місце інженера- розробника (дослідника програмного забезпечення)	0,08	480,0	288,00
Графічні пристрої виводу інформації	0,12	2,3	2,07
Офісна оргтехніка	0,32	3,0	7,20
Мережеве обладнання передачі цифрових даних	0,15	400,0	450,00
Всього			2187,27

### 5.3.9 Службові відрядження

До статті «Службові відрядження» дослідної роботи на тему «Інформаційна технологія прогнозування курсу криптовалюти ETHEREUM» належать витрати на відрядження штатних працівників, працівників організацій, які працюють за договорами цивільно-правового характеру, аспірантів, зайнятих розробленням досліджень, відрядження, пов'язані з проведенням випробувань машин та приладів, а також витрати на відрядження на наукові з'їзди, конференції, наради, пов'язані з виконанням конкретних досліджень.

Витрати за статтею «Службові відрядження» розраховуємо як 20...25% від суми основної заробітної плати дослідників та робітників за формулою:

$$B_{cv} = (Z_o + Z_p) \cdot \frac{H_{cv}}{100\%}, \quad (5.14)$$

де  $H_{cv}$  – норма нарахування за статтею «Службові відрядження», прийmemo  $H_{cv} = 20\%$ .

$$B_{cv} = (106492,86 + 3733,69) \cdot 20 / 100\% = 22045,31 \text{ грн.}$$

5.3.10 Витрати на роботи, які виконують сторонні підприємства, установи і організації

Витрати за статтею «Витрати на роботи, які виконують сторонні підприємства, установи і організації» розраховуємо як 30...45% від суми основної заробітної плати дослідників та робітників за формулою:

$$B_{cn} = (Z_o + Z_p) \cdot \frac{H_{cn}}{100\%}, \quad (5.15)$$

де  $H_{cn}$  – норма нарахування за статтею «Витрати на роботи, які виконують сторонні підприємства, установи і організації», прийmemo  $H_{cn} = 0\%$ .

$$B_{cn} = (106492,86 + 3733,69) \cdot 0 / 100\% = 0,00 \text{ грн.}$$

5.3.11 Інші витрати

До статті «Інші витрати» належать витрати, які не знайшли відображення у зазначених статтях витрат і можуть бути віднесені безпосередньо на собівартість досліджень за прямими ознаками.

Витрати за статтею «Інші витрати» розраховуємо як 50...100% від суми основної заробітної плати дослідників та робітників за формулою:

$$I_e = (Z_o + Z_p) \cdot \frac{H_{ie}}{100\%}, \quad (5.16)$$

де  $H_{ie}$  – норма нарахування за статтею «Інші витрати», прийmemo  $H_{ie} = 65\%$ .

$$I_6 = (106492,86 + 3733,69) \cdot 65 / 100\% = 71647,26 \text{ грн.}$$

### 5.3.12 Накладні (загальновиробничі) витрати

До статті «Накладні (загальновиробничі) витрати» належать: витрати, пов'язані з управлінням організацією; витрати на винахідництво та раціоналізацію; витрати на підготовку (перепідготовку) та навчання кадрів; витрати, пов'язані з набором робочої сили; витрати на оплату послуг банків; витрати, пов'язані з освоєнням виробництва продукції; витрати на науково-технічну інформацію та рекламу та ін.

Витрати за статтею «Накладні (загальновиробничі) витрати» розраховуємо як 100...150% від суми основної заробітної плати дослідників та робітників за формулою:

$$B_{нзв} = (Z_o + Z_p) \cdot \frac{H_{нзв}}{100\%}, \quad (5.17)$$

де  $H_{нзв}$  – норма нарахування за статтею «Накладні (загальновиробничі) витрати», прийmemo  $H_{нзв} = 112\%$ .

$$B_{нзв} = (106492,86 + 3733,69) \cdot 112 / 100\% = 123453,74 \text{ грн.}$$

Витрати на проведення науково-дослідної роботи на тему «Інформаційна технологія прогнозування курсу криптовалюти ETHEREUM» розраховуємо як суму всіх попередніх статей витрат за формулою:

$$B_{заг} = Z_o + Z_p + Z_{дод} + Z_n + M + K_6 + B_{спец} + B_{прз} + A_{обл} + B_e + B_{св} + B_{сп} + I_6 + B_{нзв}. \quad (4.18)$$

$$B_{заг} = 106492,86 + 3733,69 + 13227,19 + 27159,82 + 7221,70 + 3284,00 + 0,00 + 5134,50 + 7689,31 + 2187,27 + 22045,31 + 0,00 + 71647,26 + 123453,74 = 393276,65 \text{ грн.}$$

Загальні витрати  $ZB$  на завершення науково-дослідної (науково-технічної) роботи та оформлення її результатів розраховується за формулою:

$$3B = \frac{B_{заг}}{\eta}, \quad (5.19)$$

де  $\eta$  - коефіцієнт, який характеризує етап (стадію) виконання науково-дослідної роботи, прийmemo  $\eta=0,95$ .

$$3B = 393276,65 / 0,95 = 413975,42 \text{ грн.}$$

#### **5.4 Розрахунок економічної ефективності науково-технічної розробки при її можливій комерціалізації потенційним інвестором**

В ринкових умовах узагальнюючим позитивним результатом, що його може отримати потенційний інвестор від можливого впровадження результатів цієї чи іншої науково-технічної розробки, є збільшення у потенційного інвестора величини чистого прибутку.

Результати дослідження проведені за темою «Інформаційна технологія прогнозування курсу криптовалюти ETHEREUM» передбачають комерціалізацію протягом 4-х років реалізації на ринку.

*1) Розробка інформаційної системи на основі нових алгоритмів, програмних або технічних засобів.*

В цьому випадку основу майбутнього економічного ефекту будуть формувати:

$\Delta N$  – збільшення кількості споживачів яким надається відповідна інформаційна послуга у періоди часу, що аналізуються;

Показник	1-й рік	2-й рік	3-й рік	4-й рік
Збільшення кількості споживачів, осіб	350	740	900	620

$N$  – кількість споживачів яким надавалась відповідна інформаційна послуга у році до впровадження результатів нової науково-технічної розробки, прийmemo 7700 осіб;

$C_6$  – вартість послуги у році до впровадження інформаційної системи, прийmemo 1228,00 грн;



$\pm \Delta C_o$  – зміна вартості послуги від впровадження результатів, прийmemo 415,93 грн.

Можливе збільшення чистого прибутку у потенційного інвестора  $\Delta \Pi_i$  для кожного із 4-х років, протягом яких очікується отримання позитивних результатів від можливого впровадження та комерціалізації науково-технічної розробки, розраховуємо за формулою:

$$\Delta \Pi_i = (\pm \Delta C_o \cdot N + C_o \cdot \Delta N)_i \cdot \lambda \cdot \rho \cdot \left(1 - \frac{\mathcal{G}}{100}\right), \quad (5.20)$$

де  $\lambda$  – коефіцієнт, який враховує сплату потенційним інвестором податку на додану вартість. У 2023 році ставка податку на додану вартість складає 20%, а коефіцієнт  $\lambda = 0,8333$ ;

$\rho$  – коефіцієнт, який враховує рентабельність інноваційного продукту).  
Прийmemo  $\rho = 45\%$ ;

$\mathcal{G}$  – ставка податку на прибуток, який має сплачувати потенційний інвестор, у 2023 році  $\mathcal{G} = 18\%$ ;

Збільшення чистого прибутку 1-го року:

$$\Delta \Pi_1 = (415,93 \cdot 7700,00 + 1643,93 \cdot 350) \cdot 0,83 \cdot 0,45 \cdot (1 - 0,18/100\%) = 1157089,38 \text{ грн.}$$

Збільшення чистого прибутку 2-го року:

$$\Delta \Pi_2 = (415,93 \cdot 7700,00 + 1643,93 \cdot 1090) \cdot 0,83 \cdot 0,45 \cdot (1 - 0,18/100\%) = 1529668,44 \text{ грн.}$$

Збільшення чистого прибутку 3-го року:

$$\Delta \Pi_3 = (415,93 \cdot 7700,00 + 1643,93 \cdot 1990) \cdot 0,83 \cdot 0,45 \cdot (1 - 0,18/100\%) = 1982805,13 \text{ грн.}$$

Збільшення чистого прибутку 4-го року:

$$\Delta \Pi_4 = (415,93 \cdot 7700,00 + 1643,93 \cdot 2610) \cdot 0,83 \cdot 0,45 \cdot (1 - 0,18/100\%) = 2294965,97 \text{ грн.}$$

Приведена вартість збільшення всіх чистих прибутків  $\Pi_{\text{П}}$ , що їх може отримати потенційний інвестор від можливого впровадження та комерціалізації науково-технічної розробки:

$$ППП = \sum_{i=1}^T \frac{\Delta\Pi_i}{(1+\tau)^i}, \quad (5.21)$$

де  $\Delta\Pi_i$  – збільшення чистого прибутку у кожному з років, протягом яких виявляються результати впровадження науково-технічної розробки, грн;

$T$  – період часу, протягом якого очікується отримання позитивних результатів від впровадження та комерціалізації науково-технічної розробки, роки;

$\tau$  – ставка дисконтування, за яку можна взяти щорічний прогнозований рівень інфляції в країні,  $\tau=0,22$ ;

$t$  – період часу (в роках) від моменту початку впровадження науково-технічної розробки до моменту отримання потенційним інвестором додаткових чистих прибутків у цьому році.

$$\begin{aligned} ППП &= 1157089,38/(1+0,22)^1 + 1529668,44/(1+0,22)^2 + 1982805,13/(1+0,22)^3 + \\ &+ 2294965,97/(1+0,22)^4 = 948433,92 + 1027726,71 + 1091944,44 + 1035945,54 = \\ &= 4104050,61 \text{ грн.} \end{aligned}$$

Величина початкових інвестицій  $PV$ , які потенційний інвестор має вкласти для впровадження і комерціалізації науково-технічної розробки:

$$PV = k_{инв} \cdot 3B, \quad (5.22)$$

де  $k_{инв}$  – коефіцієнт, що враховує витрати інвестора на впровадження науково-технічної розробки та її комерціалізацію, приймаємо  $k_{инв}=2$ ;

$3B$  – загальні витрати на проведення науково-технічної розробки та оформлення її результатів, приймаємо 413975,42 грн.

$$PV = k_{инв} \cdot 3B = 2 \cdot 413975,42 = 827950,84 \text{ грн.}$$

Абсолютний економічний ефект  $E_{абс}$  для потенційного інвестора від можливого впровадження та комерціалізації науково-технічної розробки становитиме:

$$E_{abc} = III - PV \quad (5.23)$$

де  $III$  – приведена вартість зростання всіх чистих прибутків від можливого впровадження та комерціалізації науково-технічної розробки, 4104050,61 грн;

$PV$  – теперішня вартість початкових інвестицій, 827950,84 грн.

$$E_{abc} = III - PV = 4104050,61 - 827950,84 = 3276099,77 \text{ грн.}$$

Внутрішня економічна дохідність інвестицій  $E_e$ , які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:

$$E_e = T_{жс} \sqrt[1 + \frac{E_{abc}}{PV}]{} - 1, \quad (5.24)$$

де  $E_{abc}$  – абсолютний економічний ефект вкладених інвестицій, 3276099,77 грн;

$PV$  – теперішня вартість початкових інвестицій, 827950,84 грн;

$T_{жс}$  – життєвий цикл науково-технічної розробки, тобто час від початку її розробки до закінчення отримання позитивних результатів від її впровадження, 4 роки.

$$E_e = T_{жс} \sqrt[1 + \frac{E_{abc}}{PV}]{} - 1 = (1 + 3276099,77/827950,84)^{1/4} - 1 = 0,49.$$

Мінімальна внутрішня економічна дохідність вкладених інвестицій  $\tau_{min}$ :

$$\tau_{min} = d + f, \quad (5.25)$$

де  $d$  – середньозважена ставка за депозитними операціями в комерційних банках; в 2023 році в Україні  $d = 0,1$ ;

$f$  – показник, що характеризує ризикованість вкладення інвестицій, прийmemo 0,25.

$\tau_{\min} = 0,1+0,25 = 0,35 < 0,49$  свідчить про те, що внутрішня економічна дохідність інвестицій  $E_g$ , які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки вища мінімальної внутрішньої дохідності. Тобто інвестувати в науково-дослідну роботу за темою «Інформаційна технологія прогнозування курсу криптовалюти ETHEREUM» доцільно.

Період окупності інвестицій  $T_{ок}$  які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:

$$T_{ок} = \frac{1}{E_g}, \quad (5.26)$$

де  $E_g$  – внутрішня економічна дохідність вкладених інвестицій.

$$T_{ок} = 1 / 0,49 = 2,03 \text{ р.}$$

$T_{ок} < 3$ -х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

## 5.5 Висновок до розділу 5

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою «Інформаційна технологія прогнозування курсу криптовалюти ETHEREUM» становить 42,7 бала, що, свідчить про комерційну важливість проведення даних досліджень (рівень комерційного потенціалу розробки високий).

При оцінюванні за технічними параметрами, згідно узагальненого коефіцієнту якості розробки, науково-технічна розробка переважає існуючі аналоги приблизно в 2,80 рази.

Також термін окупності становить 2,03 р., що менше 3-х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати

потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

Отже можна зробити висновок про доцільність проведення науково-дослідної роботи за темою «Інформаційна технологія прогнозування курсу криптовалюти ETHEREUM» та доведена працездатність розробленої технології, розглянуто приклад її застосування на базі криптовалюти Ethereum за щодобовими даними 2021—2023 років, проведені дослідження та комп'ютерні експерименти показали ефективність та працездатність запропонованої технології.

## ВИСНОВКИ

В ході аналізу предметної області та вивчення існуючих реалізацій прогнозування курсу криптовалют виявлено ключові вимоги та функціональність, які повинен надавати розроблюваний додаток. Окреслено основні принципи організації та взаємодії з користувачем.

Було розроблено алгоритм роботи прогнозування курсу криптовалюти Ethereum. Представлено UML-діаграму, що ілюструє взаємодію компонентів модуля.

Відповідно до розробленої інформаційної технології здійснено прогнозування тестових даних та порівняння результатів прогнозування з реальними даними.

Проведено аналіз внутрішніх та зовнішніх факторів, що можуть впливати на криптовалютний ринок. Було розглянуто важливі аспекти, такі як точність розрахунків та швидкість виконання операцій.

Проведено перевірку розробленої програми для прогнозування курсу криптовалюти Ethereum за визначеними критеріями. Були враховані такі аспекти, як функціональність, надійність, ефективність та інші. За допомогою правильно розробленого алгоритму, програма для прогнозування курсу криптовалюти Ethereum може ефективно виконувати розрахунки та забезпечувати точність результатів.

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою «Інформаційна технологія прогнозування курсу криптовалюти ETHEREUM» становить 42,7 бала, що, свідчить про комерційну важливість проведення даних досліджень (рівень комерційного потенціалу розробки високий).

При оцінюванні за технічними параметрами, згідно узагальненого коефіцієнту якості розробки, науково-технічна розробка переважає існуючі аналоги приблизно в 2,80 рази.

Також термін окупності становить 2,03 р., що менше 3-х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

Всі завдання магістерської кваліфікаційної роботи виконані, мету повністю досягнуто.

**ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ**

1. Саміленко В. О., Паночишин Ю. М. «Вибір оптимальної моделі для прогнозування курсу криптовалюти Ethereum у 2023 році», в Матеріали конференції «Молодь в науці: дослідження, проблеми, перспективи (МН-2024)», Вінниця, 2023, [Електронний ресурс]. Режим доступу: <https://conferences.vntu.edu.ua/index.php/mn/mn2024/paper/viewFile/19665/16281>
2. Mokin V.B. Kaggle Dataset «Forecasting Top Cryptocurrencies» [Електронний ресурс]. URL: <https://www.kaggle.com/datasets/vbmokin/forecasting-top-cryptocurrencies>
3. Jumaili M. L. F., Karim S. M. Comparison of tow two cryptocurrencies: Bitcoin and ETH: *Journal of Physics : Conference Series*. – IOP Publishing, 2021. №. 1. P.163.
4. Sebastião H., Godinho P. Forecasting and trading cryptocurrencies with machine learning under changing market conditions : *Financial Innovation*. 2021. №. 1. P. 1-30.
5. Мерінова С. В., Половенко Л. П. Роль криптовалюти у цифровій економіці : *Науковий вісник ХДУ*. Серія «Економічні науки». 2021. №. 42. С. 80-87.
6. Mokin V.B. Kaggle Notebook «Cryptocurrency: Years data comparing with AutoViz» [Електронний ресурс]. URL: <https://www.kaggle.com/code/vbmokin/cryptocurrency-years-data-comparing-with-autoviz>
7. Sihombing S., Nasution M. R., Sadalia I. Analisis Fundamental Cryptocurrency terhadap Fluktuasi Harga: Studi Kasus Tahun 2019-2020 : *Jurnal Akuntansi, Keuangan, Dan Manajemen*. 2021. №. 3. P. 213-224.
8. Mokin V.B. Kaggle Notebook «Crypto – BTC : Advanced Analysis & Forecasting» [Електронний ресурс]. URL: <https://www.kaggle.com/vbmokin/crypto-btc-advanced-analysis-forecasting>



9. Mind: Новини криптовалюти «Ефіріум досягнув нового рекорду» [Електронний ресурс]. URL: <https://mind.ua/news/20225614-efirium-dosyagnuv-novogo-rekordu-vishche-3400>
10. UA.News: Новини криптовалюти «За декаду Bitcoin втратив \$100 млрд: причини та прогноз» [Електронний ресурс]. URL: <https://ua.news/ua/za-dekadu-bitcoin-poteryal-100-mlrd-prychynu-y-prognoz/>
11. Корреспондент: Новини криптовалюти «Курс біткоїна: останні новини» [Електронний ресурс]. URL: <https://ua.korrespondent.net/business/financial/3912444-kurs-bitkoina-ostanni-novyny>
12. Live Updates: Ukraine Government Turns to Crypto to Crowdfund Millions of Dollars [Електронний ресурс]. URL: <https://www.elliptic.co/blog/live-updates-millions-in-crypto-crowdfunded-for-the-ukrainian-military>
13. Time series forecasting (Part 1 of 3): Understanding the fundamentals [Електронний ресурс]. URL: <https://medium.com/data-science-at-microsoft/time-series-forecasting-part-1-of-3-understanding-the-fundamentals-13b52eda3e5>
14. Statsmodels.tsa Documentation. Statsmodels.tsa.stattools.adfuller [Електронний ресурс]. URL: <https://www.statsmodels.org/dev/generated/statsmodels.tsa.stattools.adfuller.html>
15. Linear regression [Електронний ресурс]. URL: <https://www.ibm.com/topics/linear-regression#:~:text=Resources-,What%20is%20linear%20regression%3F,is%20called%20the%20independent%20variable.>
16. Shahbazi Z., Byun Y. C. Knowledge Discovery on Cryptocurrency Exchange Rate Prediction Using Machine Learning Pipelines : *Sensors*. 2022. Т. 22. №. 5. P. 174.
17. Nils J. Nilsson Robotics Laboratory Department of Computer Science Stanford University Stanford, CA 94305, 2005. 179 с.
18. Understanding Feature Importance and How to Implement it in Python | by Terence Shin | Towards Data Science [Електронний ресурс]. URL:

<https://towardsdatascience.com/understanding-feature-importance-and-how-to-implement-it-in-python-ff0287b20285?gi=ec79fea147cf>

19. Statsmodels.tsa Documentation. Statsmodels.tsa.stattools.adfuller [Электронный ресурс]. URL:

<https://www.statsmodels.org/dev/generated/statsmodels.tsa.stattools.adfuller.html>

20. Statsmodels.tsa Documentation. Statsmodels.tsa.seasonal\_decompose [Электронный ресурс]. URL:

[https://www.statsmodels.org/dev/generated/statsmodels.tsa.seasonal.seasonal\\_decompose.html](https://www.statsmodels.org/dev/generated/statsmodels.tsa.seasonal.seasonal_decompose.html)

21. Facebook Prophet [Электронный ресурс]. URL: <https://habr.com/ru/company/ods/blog/323730/>

22. ARIMA model [Электронный ресурс]. URL: <https://otexts.com/fpp2/arima.html>

23. Rodriguez-Galiano V. et al. Machine learning predictive models for mineral prospectivity: *An evaluation of neural networks, random forest, regression trees and support vector machines* : Ore Geology Reviews. 2015. Т. 71. P. 804-818.

24. KNeighborsRegressor [Электронный ресурс]. URL: <https://fda.readthedocs.io/en/stable/modules/ml/autosummary/sklda.ml.regression.KNeighborsRegressor.html>

25. sklearn.ensemble.BaggingRegressor [Электронный ресурс]. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.BaggingRegressor.html>

26. sklearn.metrics.r2\_score — scikit-learn 1.1.1 documentation [Электронный ресурс]. URL: [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2\\_score.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html)

27. Metrics r2\_score [Электронный ресурс]. URL: [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2\\_score.html?highlight=r2\\_score#sklearn.metrics.r2\\_score](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html?highlight=r2_score#sklearn.metrics.r2_score)

28. Random Forest Regression [Электронный ресурс]. URL: <https://towardsdatascience.com/random-forest-regression-5f605132d19d>

29. Feng D. C. et al. Interpretable XGBoost-SHAP machine-learning model for shear strength prediction of squat RC walls : *Journal of Structural Engineering*. 2021. Т. 147. №. 11. Р. 173.

30. ELI5 [Електроний ресурс]. URL: <https://eli5.readthedocs.io/en/latest/index.html>

31. Методичні вказівки до виконання економічної частини магістерських кваліфікаційних робіт / Уклад. : В. О. Козловський, О. Й. Лесько, В. В. Кавецький. – Вінниця : ВНТУ, 2021. – 42 с.

32. Методичні вказівки до виконання економічної частини магістерських кваліфікаційних робіт / Уклад. : В. О. Козловський, О. Й. Лесько, В. В. Кавецький. – Вінниця : ВНТУ, 2021. – 42 с.

## **ДОДАТКИ**

Додаток А (обов'язковий) Протокол перевірки кваліфікаційної роботи на наявність текстових запозичень

ПРОТОКОЛ ПЕРЕВІРКИ КВАЛІФІКАЦІЙНОЇ РОБОТИ  
НА НАЯВНІСТЬ ТЕКСТОВИХ ЗАПОЗИЧЕНЬ

Назва роботи: Інформаційна технологія прогнозування курсу криптовалюти ETHEREUM»

Тип роботи: магістерська кваліфікаційна робота

(БДР, МКР)

Підрозділ кафедра комп'ютерних наук, ФІТА

(кафедра, факультет)

Показники звіту подібності Unicheck

Оригінальність 94,1% Схожість 5,9%

Аналіз звіту подібності (відмітити потрібне):

- Запозичення, виявлені у роботі, оформлені коректно і не містять ознак плагіату.
- Виявлені у роботі запозичення не мають ознак плагіату, але їх надмірна кількість викликає сумніви щодо цінності роботи і відсутності самостійності її виконання автором. Роботу направити на розгляд експертної комісії кафедри.
- Виявлені у роботі запозичення є недобросовісними і мають ознаки плагіату та/або в ній містяться навмисні спотворення тексту, що вказують на спроби приховування недобросовісних запозичень.

Ознайомлені з повним звітом подібності, який був згенерований системою Unicheck щодо роботи.

Автор роботи



Саміленко В.О.

Керівник роботи



Паночишин Ю.М.

Опис прийнятого рішення

Магістерську кваліфікаційну роботу допущено до захисту

Особа, відповідальна за перевірку



Озеранський В.С.

**Додаток Б (обов'язковий)****Лістинг програми**

```
# ## 1. Import libraries <a class="anchor" id="1"></a>
#
# [Back to Table of Contents](#0.1)

# In[1]:

# Import libraries
import random
import os
import numpy as np
import pandas as pd
import requests
import pandas_datareader as web

# Date
import datetime as dt
from datetime import date, timedelta, datetime

# EDA
import matplotlib.pyplot as plt
from matplotlib.pylab import rcParams
import plotly.express as px
import plotly.graph_objects as go
from plotly.offline import init_notebook_mode
init_notebook_mode(connected=True)

# FE
from tsfresh import extract_features, select_features, extract_relevant_features
from tsfresh.utilities.dataframe_functions import impute
from sklearn.inspection import permutation_importance
```

```
import eli5
from eli5.sklearn import PermutationImportance
import shap

# Time Series - EDA and Modelling
import statsmodels.api as sm
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.stattools import adfuller
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.tsa.arima_model import ARIMA

# Metrics
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error, mean_absolute_percentage_error

# Modeling and preprocessing
from sklearn.preprocessing import StandardScaler, MinMaxScaler
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.linear_model import LinearRegression
from sklearn.svm import SVR, LinearSVR
from sklearn.neighbors import KNeighborsRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.ensemble import BaggingRegressor, AdaBoostRegressor
from sklearn.neural_network import MLPRegressor
from prophet import Prophet
import xgboost as xgb
from xgboost import XGBRegressor
import lightgbm as lgb
from lightgbm import LGBMRegressor

import warnings
warnings.filterwarnings("ignore")

# In[2]:
```

```
# Automatic building ARIMA for Time Series
get_ipython().system('pip install pmdarima')
import pmdarima as pm

# In[3]:

# Set random state
def fix_all_seeds(seed):
    np.random.seed(seed)
    random.seed(seed)
    os.environ['PYTHONHASHSEED'] = str(seed)

random_state = 42
fix_all_seeds(random_state)

# **TASK:** It is proposed to experiment with forecasting_days

# In[4]:

# Set main parameters
cryptocurrency = 'ETH'
target = 'Close'
forecasting_days = 10 # forecasting_days > 1

# **TASK:** It is proposed to experiment with date_start and date_end
# In[88]:

def get_model_opt(name_model, params):
    # Model tuning for the name_model

    print(name_model)
    if name_model=='Linear Regression':
        model = LinearRegression(**params)
```



```

elif name_model=='KNeighbors Regressor':
    model = KNeighborsRegressor(**params)

elif name_model=='Support Vector Machines':
    model = SVR(**params)

elif name_model=='Linear SVR':
    model = LinearSVR(**params)

elif name_model=='Random Forest Regressor':
    model = RandomForestRegressor(**params)

elif name_model=='Bagging Regressor':
    model = BaggingRegressor(**params)

elif name_model=='MLP Regressor':
    model = MLPRegressor(**params)

elif name_model=='XGB Regressor':
    model = xgb.XGBRegressor(**params)

else: model = None

return model
# In[90]:

def model_training_forecasting(result, df, y, test, ytest,
                              name_model, type_model, params, type_test='1'):
    # Model training for df and y
    # Forecasting ypred
    # type_model = 'Prophet' or "ARIMA" or 'Other ML'
    # type_test = '1' (with find optimal parameters by GridSearchCV)
    # type_test = '2' (with optimal parameters - without GridSearchCV)

```

```
# return params and metrics in the dataframe result

if type_model=='Prophet':
    season_days_optimal = params[0]
    fourier_order_seasonality_optimal = params[1]
    model_opt = None
    _, ypred = prophet_modeling(result,
                                cryptocurrency,
                                df,
                                test,
                                holidays_df,
                                season_days_optimal,
                                fourier_order_seasonality_optimal,
                                forecasting_days,
                                f'{type_model}_optimal',
                                'test')

elif type_model=='ARIMA':
    season_days_optimal = params[0]
    fourier_order_seasonality_optimal = params[1]
    model_opt = None


# Training ARIMA optimal model for training+valid dataset
df['y'] = y
model_opt = arima_fit(df, 'y', order=(params[0],params[1],params[2]))
```

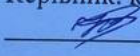
## Додаток В (обов'язковий)

## ІЛЮСТРАТИВНА ЧАСТИНА

«Інформаційна технологія прогнозування курсу криптовалюти  
ETHEREUM»

Виконав: студент 2-го курсу,  
групи ЗКН-22м  
спеціальності 122 «Комп'ютерні науки»  
(шифр і назва напрямку підготовки, спеціальності)

 Саміленко В.О.  
(прізвище та ініціали)

Керівник: к.т.н., доцент каф. КН  
 Паночішин Ю.М.  
(прізвище та ініціали)

« 07 » 12 2022 р.

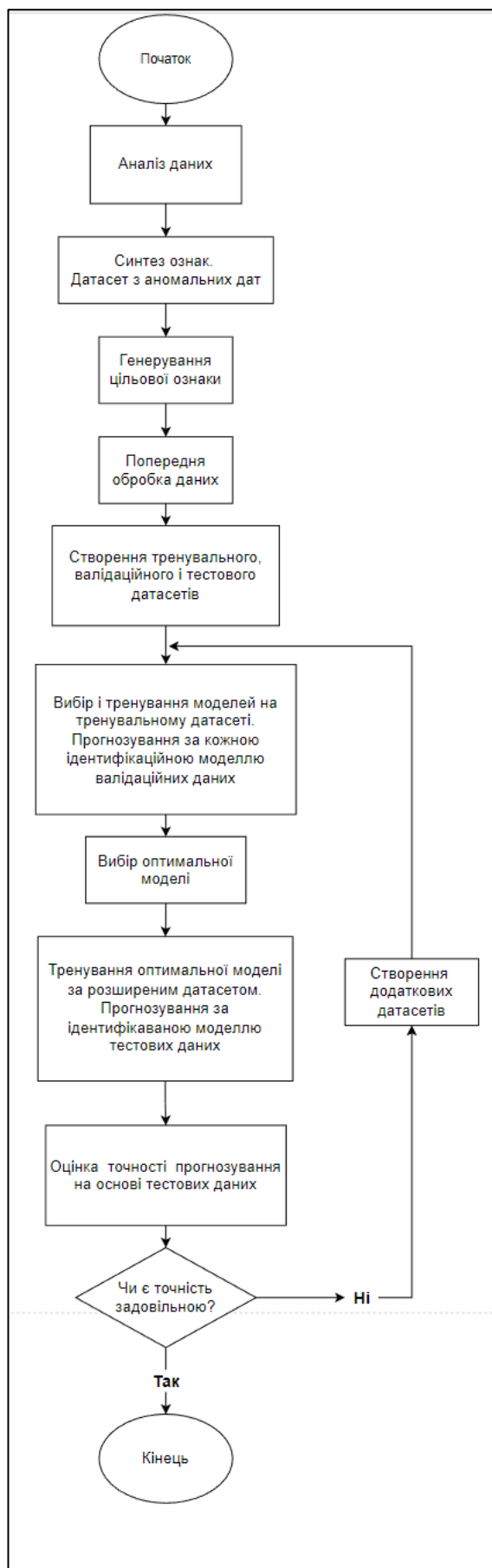


Рисунок В.1 – Блок-схема виконання алгоритму інформаційної технології аналізу та прогнозування курсу Ethereum у 2023 році

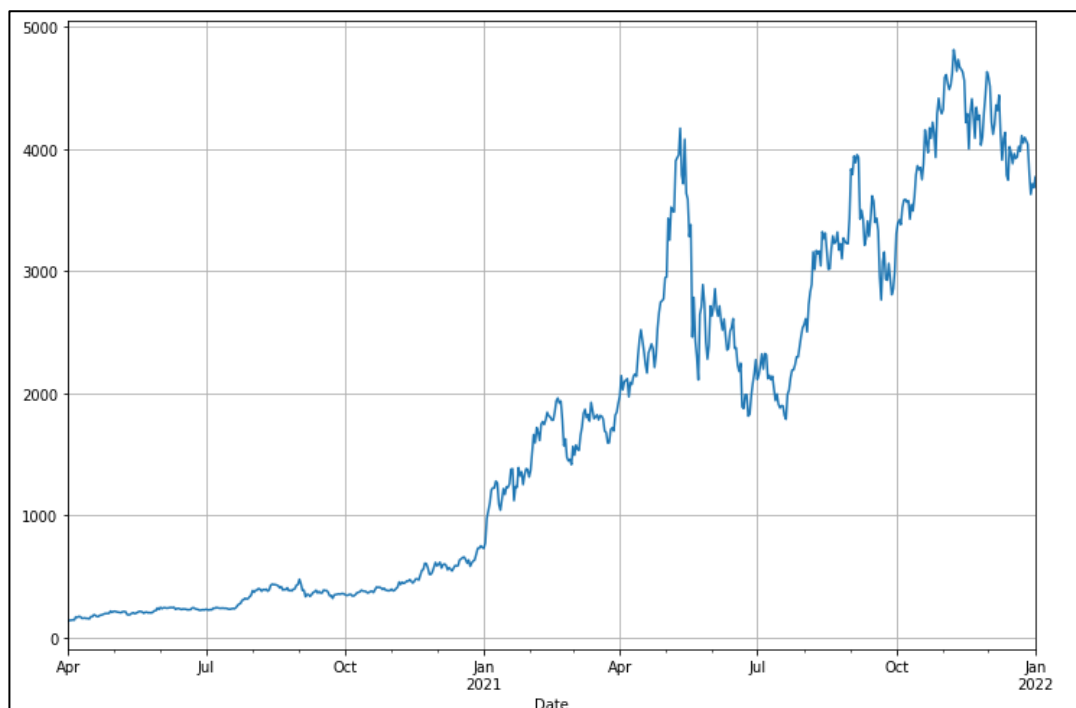


Рисунок В.2 – Курс Ethereum за даними сайту Yahoo

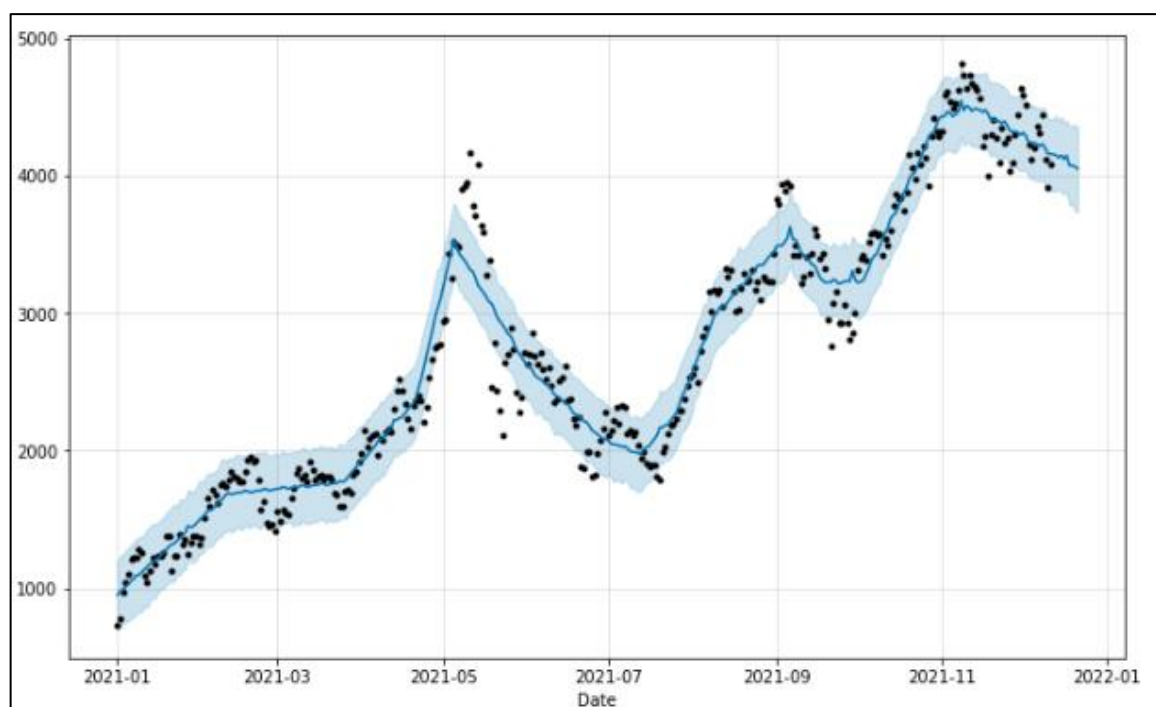


Рисунок В.3 – Результат прогнозування модель Facebook Prophet

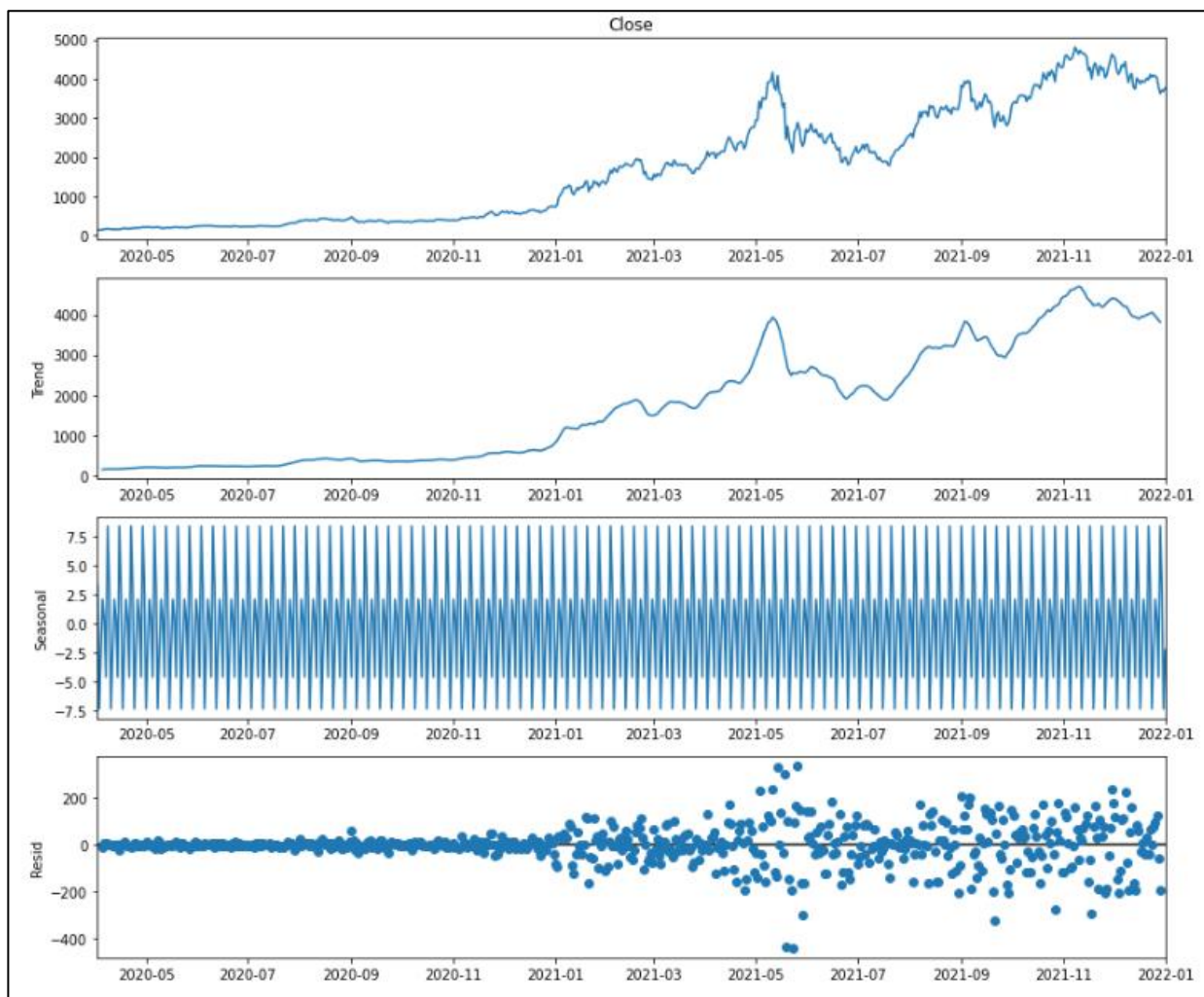


Рисунок В.4 – Декомпозиція часового ряду на «Trend», «Seasonal» та «Resid», курсу криптовалюти Ethereum

SARIMAX Results						
Dep. Variable:	y		No. Observations:	345		
Model:	ARIMA(0, 1, 0)		Log Likelihood	-2212.315		
Date:	Wed, 08 Jun 2022		AIC	4426.630		
Time:	09:39:25		BIC	4430.471		
Sample:	0		HQIC	4428.160		
	- 345					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
sigma2	2.257e+04	946.557	23.845	0.000	2.07e+04	2.44e+04
Ljung-Box (L1) (Q):			4.66	Jarque-Bera (JB):	362.88	
Prob(Q):			0.03	Prob(JB):	0.00	
Heteroskedasticity (H):			3.11	Skew:	-0.70	
Prob(H) (two-sided):			0.00	Kurtosis:	7.83	

Рисунок В.5 – Результат прогнозування модель ARIMA

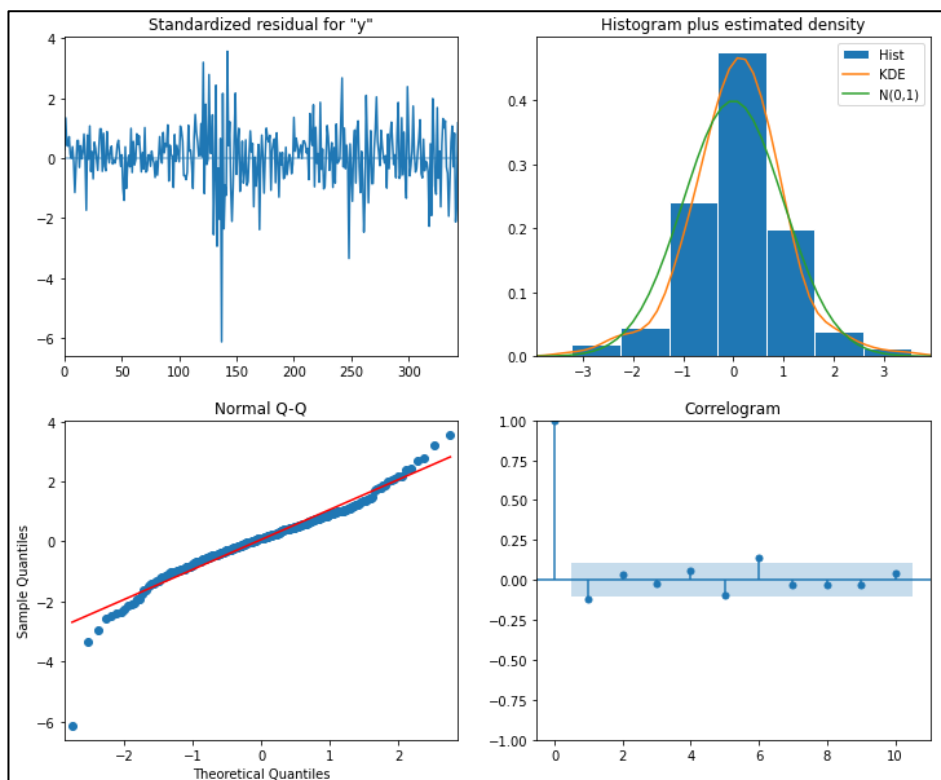


Рисунок В.6 – Похибки прогнозу верифікованих даних оптимальної моделі

### ARIMA

	name_model	type_data	r2_score	rmse	mape
14	Random Forest Regressor	valid	-0.947526	151.154076	2.945693
11	KNeighbors Regressor	valid	-0.982406	152.501624	3.232935
10	Linear Regression	valid	-1.362067	166.465542	3.635112
12	Support Vector Machines	valid	-1.567659	173.558912	3.764183
16	XGB Regressor	valid	-1.668826	176.945012	3.905291
13	Linear SVR	valid	-1.702011	178.041709	3.940936
7	Prophet_14_days_12_order	valid	-2.480632	202.072527	4.007154
6	Prophet_14_days_3_order	valid	-2.626355	206.259228	4.01593
9	ARIMA_auto	valid	-1.809227	181.539691	4.039218
8	ARIMA_manual	valid	-1.873414	183.60197	4.087203
4	Prophet_7_days_3_order	valid	-2.811506	211.459154	4.517025
2	Prophet_5_days_3_order	valid	-3.001961	216.677911	4.627023
0	Prophet_4_days_3_order	valid	-3.013654	216.994213	4.665902
17	MLP Regressor	valid	-3.095436	219.193802	5.147478
5	Prophet_7_days_12_order	valid	-4.013708	242.525687	5.510833
3	Prophet_5_days_12_order	valid	-4.571395	255.658463	5.861905
1	Prophet_4_days_12_order	valid	-4.844546	261.850624	6.063803
15	Bagging Regressor	valid	-36.917089	666.95372	14.294553

Рисунок В.7 – Похибки всіх моделей, що були побудовані для багатьох ознак для верифікованих даних курсу криптовалюти на 10 днів

	name_model	type_data	r2_score	rmse	mape
11	KNeighbors Regressor	valid	-0.121617	22.468789	0.446885
14	Random Forest Regressor	valid	-0.312955	24.309851	0.59479
13	Linear SVR	valid	-0.744307	28.020044	0.609081
9	ARIMA_auto	valid	-2.462671	39.478707	0.826748
8	ARIMA_manual	valid	-4.799886	51.093664	1.147166
16	XGB Regressor	valid	-4.977551	51.870325	1.170538
12	Support Vector Machines	valid	-11.976779	76.425939	1.762409
15	Bagging Regressor	valid	-46.460514	146.158319	3.51471
7	Prophet_14_days_12_order	valid	-47.873985	148.318798	3.580896
17	MLP Regressor	valid	-58.98706	164.318366	3.588498
6	Prophet_14_days_3_order	valid	-51.97449	154.415436	3.724006
10	Linear Regression	valid	-65.480593	172.983532	4.081759
0	Prophet_4_days_3_order	valid	-87.956725	200.099827	4.877806
2	Prophet_5_days_3_order	valid	-88.498487	200.708224	4.910621
4	Prophet_7_days_3_order	valid	-98.091198	211.190734	5.178781
3	Prophet_5_days_12_order	valid	-103.481647	216.858938	5.307153
5	Prophet_7_days_12_order	valid	-109.523133	223.040594	5.469506
1	Prophet_4_days_12_order	valid	-110.656693	224.181463	5.470795

Рисунок В.8 – Похибка усіх моделей, побудованих для багатьох ознак для валідаційних даних курсу цифрової валюти на 4 доби

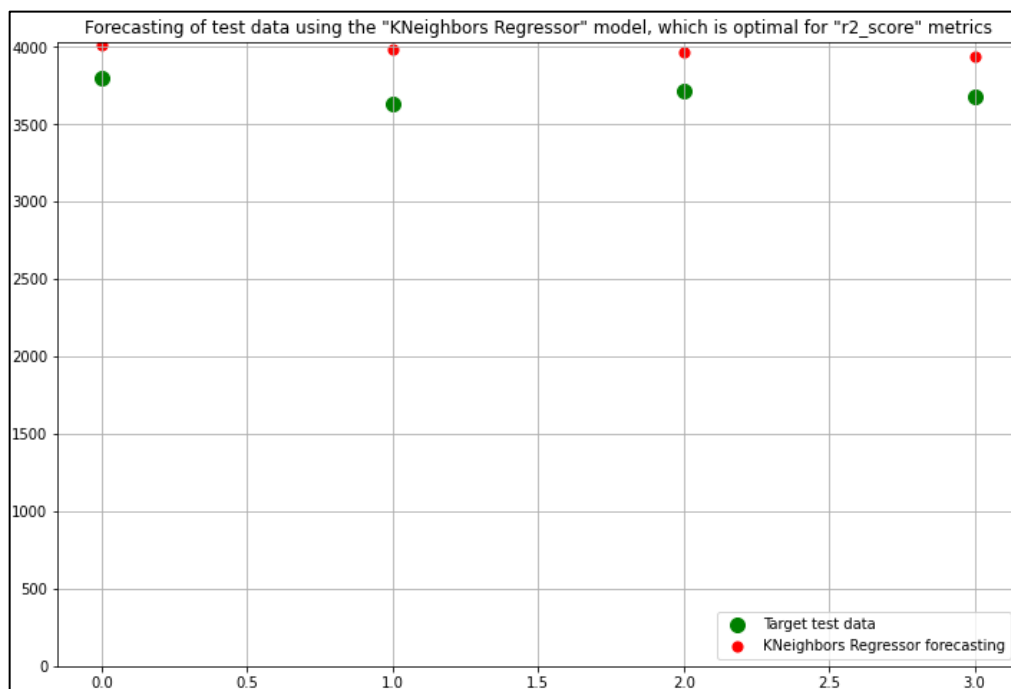


Рисунок В.9 – Результат прогнозування на 4 доби оптимальної моделі Machine Learning



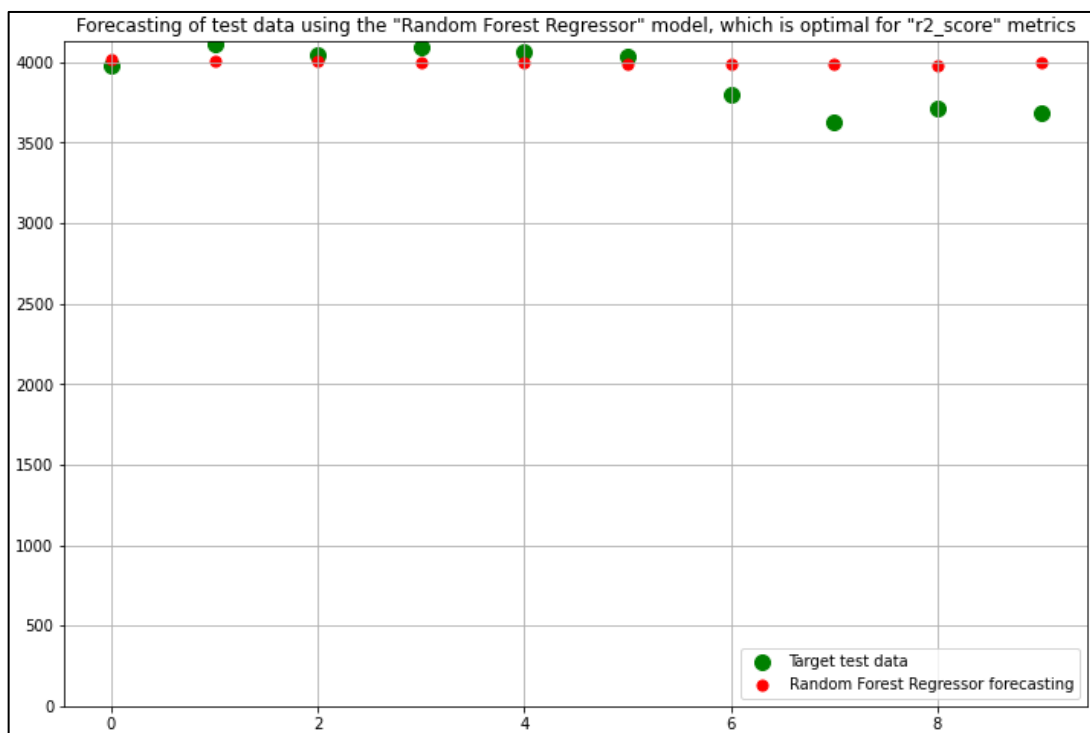


Рисунок В.10 – Результат прогнозування на 10 днів оптимальної моделі Machine Learning