

Вінницький національний технічний університет
Факультет інтелектуальних інформаційних технологій та автоматизації
Кафедра системного аналізу та інформаційних технологій

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

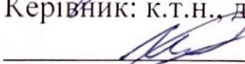
на тему:

“Інформаційна технологія прогнозування курсу криптовалюти Bitcoin у 2021-2022 роках”

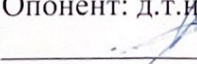
Виконав: студент 2 курсу, групи 2ІСТ-21м
спеціальності 126 – «Інформаційні системи
та технології»


Бурденюк А. В.

Керівник: к.т.н., доц. каф. САІТ



Козачко О. М.
«01» 12 2022 р.

Опонент: д.т.н., проф. каф. КСУ


Ковтун В.В.
«16» 12 2022 р.

Допущено до захисту

Завідувач кафедри САІТ


д.т.н., проф. Мокін В. Б.

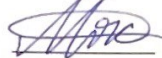
«05» 12 2022 р.

Вінниця ВНТУ – 2022 рік

Вінницький національний технічний університет
Факультет інтелектуальних інформаційних технологій та автоматизації
Кафедра системного аналізу та інформаційних технологій
Рівень вищої освіти – II-й (магістерський)
Галузь знань – 12 Інформаційні технології
Спеціальність – 126 Інформаційні системи та технології
Освітньо-професійна програма – Інформаційні технології аналізу даних та зображень

ЗАТВЕРДЖУЮ

Завідувач кафедри САІТ




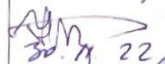
 д.т.н., проф. Мокін В. Б.

«16» _____ 09 2022 р.

ЗАВДАННЯ
НА МАГІСТЕРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ
Бурденюку Артему Васильовичу

1. Тема роботи: “Інформаційна технологія прогнозування курсу криптовалюти Bitcoin у 2021-2022 роках”,
керівник роботи: Козачко О. М., к.т.н., доц. каф. САІТ,
затверджені наказом закладу вищої освіти від «14» _____ 09 2022 року №205
2. Строк подання студентом роботи «01» _____ 12 2022 року
3. Вихідні дані до роботи:
Набір даних про курс Bitcoin на світовому ринку за період з 01.01.2021 р. по 01.01.2022 р.
4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)
 - 1) Загальна характеристика об’єкту досліджень;
 - 2) Вибір оптимальних налаштувань для розв’язання поставленої задачі;
 - 3) Розроблення інформаційної технології аналізу та прогнозування курсу криптовалюти Bitcoin у 2021 – 2022 роках;
 - 4) Економічна частина.
5. Перелік ілюстративного матеріалу:
 - Капіталізація ринку криптовалюти;
 - Графік курсу криптовалюти Bitcoin та дати аномальних змін;
 - OHLC-діаграма криптовалюти Bitcoin за даними сайту Yahoo;
 - Таблиця похибок моделей;
 - Блок-схема алгоритму інформаційної технології прогнозування курсу Bitcoin у 2021-2022 роках;
 - Графік тестових даних курсу криптовалюти Bitcoin та їх прогноз за оптимальною моделлю.

6. Консультанти розділів МКР

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
3	Мокін В.Б., д.т.н., проф. каф. САІТ	 01.11.22р.	 20.11.22р.
4	Буреннікова Н.В., д.е.н., проф. каф. ЕПВМ	 05.11.22р.	 30.11.22р.

7. Дата видачі завдання «16» 09 2022 року

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів МКР	Строк виконання етапів роботи	Примітка
1	Аналіз предметної області	09.2022	
2	Вибір оптимальних налаштувань для поставленої задачі.	10.2022	
3	Розроблення інформаційної технології аналізу та прогнозування курсу криптовалюти Bitcoin у 2021 – 2022 роках.	11.2022	
4	Економічна частина	11.2022	
5	Оформлення матеріалів до захисту МКР	11.2022	

Студент

Керівник роботи



Бурденюк А. В.

Козачко О. М.

АНОТАЦІЯ

УДК 33.336

Бурденюк А. В. Інформаційна технологія прогнозування курсу криптовалюти Bitcoin у 2021-2022 роках. Магістерська кваліфікаційна робота зі спеціальності 126 – інформаційні системи та технології, освітньо-професійна програма – інформаційні технології аналізу даних та зображень. Вінниця: ВНТУ, 2022. 85 с.

На укр. мові. Бібліогр.: 23 назв; рис.: 40; табл.: 9.

В магістерській кваліфікаційній роботі проведено аналіз предметної області, обґрунтовано доцільність та актуальність даного дослідження. Проаналізовано фактори які впливають на курс криптовалюти Bitcoin, виділено аномальні дати. Зроблено прогноз курсу криптовалюти Bitcoin за допомогою методів машинного навчання. Виділено оптимальну модель та відібрано важливі ознаки, які найбільше вплинули на прогноз. Розроблено інформаційну технологію прогнозування курсу криптовалюти Bitcoin у 2021-2022 роках.

Ілюстративна частина складається з 6 плакатів.

У розділі економічної частини розглянуто питання про доцільність розробки та впровадження інформаційної системи прогнозування курсу криптовалюти Bitcoin.

Ключові слова: інформаційна система, цифрова валюта, криптовалюта, нейронна мережа, машинне навчання, Bitcoin.

ABSTRACT

Burdeniuk A. V. Information technology for forecasting the Bitcoin cryptocurrency rate in 2021-2022. Master's qualification thesis on specialty 126 – information systems and technologies, educational and professional program – information technologies of data and image analysis. Vinnytsia: VNTU, 2022. 85 p.

Ukraine language Bibliography: 23 titles; Fig.: 40; tab: 9.

In the master's qualification work, an analysis of the subject area was carried out, the expediency and relevance of this research was substantiated. Factors affecting the exchange rate of the Bitcoin cryptocurrency have been analyzed, anomalous dates have been highlighted. A forecast of the Bitcoin cryptocurrency exchange rate was made using machine learning methods. The optimal model was selected and the important features that had the greatest impact on the forecast were selected. Information technology for forecasting the exchange rate of the Bitcoin cryptocurrency in 2021-2022 has been developed.

The illustrative part consists of 6 posters.

In the section of the economic part, the issue of the feasibility of developing and implementing an information system for forecasting the exchange rate of the Bitcoin cryptocurrency is considered.

Keywords: information system, digital currency, cryptocurrency, neural network, machine learning, Bitcoin.

ЗМІСТ

ВСТУП	4
1 ЗАГАЛЬНА ХАРАКТЕРИСТИКА ОБ'ЄКТУ ДОСЛІДЖЕНЬ	6
1.1 Аналіз предметної області.....	6
1.2 Переваги та недоліки криптовалют.....	10
1.3 Огляд існуючих технологій прогнозування цифрових валют.....	12
1.4 Формалізація постановки задачі.....	13
1.5 Висновки	17
2 ВИБІР ОПТИМАЛЬНИХ НАЛАШТУВАНЬ ДЛЯ РОЗВ'ЯЗАННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ.....	18
2.1 Огляд технології Kaggle середовища.....	18
2.2 Синтез нових ознак	20
2.3 Розвідувальний аналіз даних.....	23
2.4 Побудова моделей методами машинного навчання.....	39
2.5 Висновки	44
3 РОЗРОБЛЕННЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ТА ПРОГНОЗУВАННЯ ЗА НЕЮ КУРСУ BITCOIN.....	45
3.1 Розроблення інформаційної технології прогнозування курсу Bitcoin.....	45
3.2 Прогнозування тестових даних.....	47
3.3 Висновки	50
4 ЕКОНОМІЧНА ЧАСТИНА	52
4.1 Комерційний та технологічний аудит науково-технічної розробки	52
4.2 Прогнозування витрат на виконання науково-дослідної (дослідно- конструкторської) роботи.....	55
4.3 Розрахунок економічної ефективності науково-технічної розробки за її можливої комерціалізації потенційним інвестором	60
4.4 Розрахунок ефективності вкладених інвестицій і періоду їх окупності ...	63
4.5 Висновки	66
ВИСНОВКИ.....	67

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	69
Додаток А (обов'язковий). Технічне завдання	71
Додаток Б (обов'язковий). Протокол перевірки кваліфікаційної роботи на наявність текстових запозичен	74
Додаток В (довідниковий). Лістинг програми	75
Додаток Г (обов'язковий). Ілюстративна частина	79

ВСТУП

Актуальність теми. На сьогоднішній день, прогнози стали частиною нашого світу. Економіка складається з прогнозів: політичної обстановки, появи та популярності технологій, врожаю на аграрному секторі, погоди, поточної ситуації з видобутком корисних копалин для різноманітних галузей, та прогнозів вартості різноманітних активів та валют. Всі ці види поточної інформації та передбачень впливають на попит економічного ринку. А цей попит відображається у серцевині економіки – торгах на біржах, які є відображенням капіталістичного устрою людської цивілізації у 21-му столітті.

На відміну від інших електронних платіжних систем криптовалюта спочатку з'являється без участі реальних грошей. Щоб стати власником певної суми коїнів цілком достатньо підключитися до сервісу їх створення, стати учасником єдиної мережі Майнінг і дочекатися свого «заробітку». У цьому полягає ключова відмінність криптовалюта від реальних грошей.

Мета і завдання роботи. Метою дослідження є підвищення точності прогнозування курсу криптовалюти Bitcoin у 2021-2022 роках за рахунок використання методів машинного навчання.

Для досягнення мети дослідження нам потрібно розв'язати такі задачі:

- розглянути проблеми прогнозування курсу криптовалюти Bitcoin;
- провести розвідувальний аналіз даних та визначити оптимальну модель для вирішення задачі прогнозування курсу криптовалюти Bitcoin;
- розробити інформаційну технологію прогнозування курсу Bitcoin.

Об'єктом дослідження магістерської кваліфікаційної роботи є процес прогнозування ціни біткоїна на криптовалютному ринку у 2021-2022 роках.

Предметом дослідження. Предметом дослідження магістерської кваліфікаційної роботи є методи та технологія прогнозування курсу цифрових валют на основі методів машинного навчання.

Методи дослідження. У роботі використано методи розвідувального аналізу для дослідження даних, методи аналізу часових рядів для прогнозування

та методи і технології машинного навчання для тренування та вибору оптимальної моделі.

Новизна одержаних результатів. Дістала подальший розвиток інформаційна технологія прогнозування курсу криптовалюти Bitcoin у 2021-2022 роках, яка забезпечує підвищення точності прогнозування за рахунок використання методів машинного навчання.

Апробація результатів магістерської кваліфікаційної роботи. Результати роботи доповідались на всеукраїнській науково-практичній інтернет-конференції «Молодь в науці: дослідження, проблеми, перспективи» (Вінниця, 2022-2023 рр.).

Публікації результатів магістерської кваліфікаційної роботи. Опубліковано тези на Всеукраїнській науково-практичній інтернет-конференції «Молодь в науці: дослідження, проблеми, перспективи» (Вінниця, 2022-2023 рр.) [1].

1 ЗАГАЛЬНА ХАРАКТЕРИСТИКА ОБ'ЄКТУ ДОСЛІДЖЕНЬ

1.1 Аналіз предметної області

За даними CoinMarketCap – Forbes, у світі криптовалют на сьогоднішній день існує понад 1100 видів цифрових грошей. А загальна капіталізація цього ринку – 133 мільярди доларів. DW збрала факти про найпопулярніші види криптовалют і про те, в які з них, крім біткоіна, найчастіше вкладають гроші [2].

Альткоіни – альтернатива біткоінам. Більшість альткоінів (альтернативні цифрові валюти) є лише варіацією біткоіну. Творці нових валют беруть вже існуючий код біткоіну і покращують його на власний розсуд. Виходить розгалуження криптовалют, з єдиним "стартом" у вигляді біткоіну. Подібний процес називається "хардфорк". Хардфорк відбувається, коли користувачі валюти не можуть дійти одного рішення щодо роботи цифрових грошей.

Наприклад, так сталося з Bitcoin у серпні. Відмінна риса біткоіну – швидкість та дешевизна транзакцій. Проте криптовалюта стала такою популярною, що система перестала долати обсяг роботи. У результаті користувачі біткоіну могли чекати по кілька годин, а то й днів, щоб пройшла операція [2].

Користувачі біткоіну вирішили, що настав час покращувати систему. Але криптовалюта не має єдиного власника, який міг би внести нові зміни. Щоб змінити децентралізовану криптовалюту, недостатньо рішення більшості. Потрібно, щоб 99 відсотків користувачів погодилися з новими правилами і почали їх дотримуватися. І хоча рішення змінити систему біткоіну було прийнято, користувачі не змогли дійти загального плану дій. У результаті біткоін розколовся на Bitcoin і Bitcoin Cash, і на ринку стало ще одну криптовалюту більше. Крім Bitcoin Cash, серед альтернативних біткоіну цифрових грошей, популярні ще 5 видів [2].

Ethereum (ефіріум, ефір) – валюта "розумних контрактів". Невідомо чому автор "Ефіру" – канадський програміст російського походження, Віталій

Бутерін, назвав його так. Можливо через те, що на відміну від того ж біткоїну, кількість ефіру необмежена. А можливо через те, що ефір – перша криптовалюта, в якій використовується технологія "розумних контрактів", а отже він може "набути форми" будь-якого об'єкта [3].

Ripple – найшвидша криптовалюта. Більшість нових криптовалют з'являються завдяки невеликим змінам коду біткоїну. Як, наприклад, Ethereum.

Ripple створювався, щоб збільшити швидкість банківських операцій та заощадити на них. Калькулятор на офіційному сайті Ripple показує, що якщо оборот банку 5 мільйонів доларів, а кількість транзакцій складає 300 тисяч штук на рік, можна заощадити 3,4 долара на кожній транзакції. У рік банк може заощадити понад мільйон доларів [3].

Технологію Ripple вже використовують Bank of America, HSBC. На відміну від біткоїну та "ефіру", криптовалюту Ripple не можна "майнути" (добувати). Це централізована система, де всі цифрові гроші вже існують і належать до однієї компанії – Ripple Lab. На даний момент існує понад 38 мільярдів одиниць криптовалюти.

І якщо системи біткоїну та "ефіру" вимагають порожніх математичних обчислень, щоб "гальмувати" створення нових блоків, то Ripple Lab видає криптовалюту за корисні обчислення – опрацювання наукових даних різних університетів. Ripple займає 5,7 відсотка від усього ринку криптовалюту, а це – 7 мільярдів доларів. Коштує одна криптовалюта Ripple менша за долар – 18 центів.

Litecoin – цифрові гроші для швидких транзакцій. Займає 2 відсотки загальної вартості ринку криптовалют, це 2,7 мільярда доларів. Litecoin з'явився в 2011 році, завдяки колишньому інженеру Google – Чарльзу Лі. Litecoin, як і, як і Ethereum, хардфорк від біткоїна [2].

Dash – анонімна криптовалюта. Dash посідає 6-те місце у списку криптовалют. Усього існує понад 7,5 мільйонів одиниць цієї криптовалюти, а коштує вона 327 доларів за один деш. Це ще одна криптовалюта, яка з'явилася у

2014 році внаслідок зміни коду біткоїну. Основна відмінність від біткоїну – повна анонімність [3].

Біткоїн анонімне доти, доки не буде виявлено власника гаманця. Тобто всі транзакції біткоїну і так видно, але який толк, якщо відправник і одержувач невідомі? Якщо якимось чином власник біткоїн-гаманця стане відомий, то можна буде простежити всі переміщення його коштів на біткоїн-рахунку, навіть якщо він 5 років тому купив каву [2].

За це відповідають оператори – ще одна відмінність від біткоїну. Оператори, так само як і "майнери", обробляють інформацію на своїх комп'ютерах та отримують за це кошти. На форумах, присвячених видобутку криптовалюти, кажуть, що оператори заробляють у два рази більше, ніж "майнери".

Nem – валюта з унікальним кодом. Криптовалюта Nem з'явилася наприкінці 2015 року. На відміну від більшості криптовалют вона має свій унікальний код. Але найголовніша відмінність – Nem працює використовуючи технологію алгоритма POI [3].

Наприклад, Litecoin і Dash працюють, використовуючи алгоритм proof-of-work або POW (доказ роботи. – Ред). Це ніби ви влаштувалися на роботу художником, а платили б вам виключно за написані картини. Також і система біткоїн-подібних валют нагороджує своїх користувачів за правильні обчислення. Однак у цій системі вже є мінуси – оскільки завдання ускладнюються, обчислювальні потужності також мають збільшуватися. У результаті витрачається безліч ресурсів для пошуку відповіді на кожне завдання, яке не несе жодної користі суспільству.

Алгоритм POI, який використовується в Nem, поєднує концепції цих двох алгоритмів. POI не тільки винагороджує тих, у кого більше залишок на рахунку, але й враховує, наскільки часто відбуваються угоди з іншими користувачами. Кожному користувачеві надається рейтинг довіри. Чим вищий він, тим більше шансів отримати винагороду [3].

Криптовалюта займає 1,5 відсотка загальної вартості ринку криптовалют, а це 2 мільярди доларів. Існує 8,9 мільярда одиниць криптовалюти Nem, а її курс сягнув 0,18 долара.

Серед усіх криптовалют за капіталізацією станом на 29.01.2020 р. ця криптовалюта знаходиться на першому місці (рис. 1.1) [1].

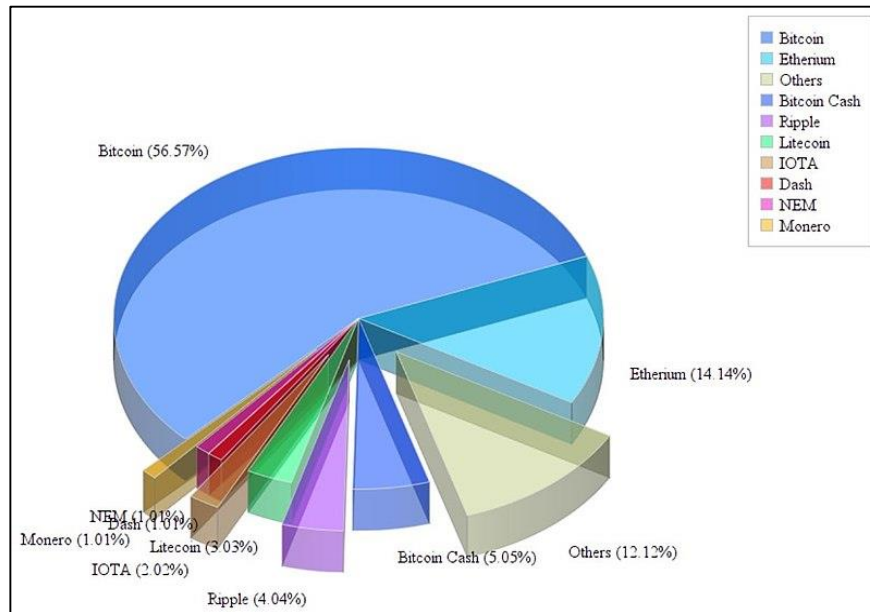


Рисунок 1.1 –Bitcoin за капіталізацією серед 10 топових криптовалют за даними сайту Wikinews станом на 29.01.2022 р.

Наприклад, так сталося з Bitcoin у серпні. Відмінна риса біткоіну – швидкість та дешевизна транзакцій. Проте криптовалюта стала такою популярною, що система перестала долати обсяг роботи. У результаті користувачі біткоіну могли чекати по кілька годин, а то й днів, щоб пройшла операція [4].

Користувачі біткоіну вирішили, що настав час покращувати систему. Але криптовалюта не має єдиного власника, який міг би вести нові зміни. Щоб змінити децентралізовану криптовалюту, недостатньо рішення більшості. Потрібно, щоб 99 відсотків користувачів погодилися з новими правилами і почали їх дотримуватися. І хоча рішення змінити систему біткоіну було прийнято, користувачі не змогли дійти загального плану дій. У результаті біткоїн

розколовся на Bitcoin і Bitcoin Cash, і на ринку стало ще одну криптовалюту більше. Крім Bitcoin Cash, серед альтернативних біткоїну цифрових грошей, популярні ще 5 видів [4].

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



Рисунок 1.2 – Курс Bitcoin за даними сайту Yahoo

1.2 Переваги та недоліки криптовалют

До переваг криптовалют можна віднести наступне [5]:

- Наявність криптовалют – готових до використання електронних грошей;

- Ви можете в будь-який момент перевірити дійсність транзакції;
- Код відкритий. Завдяки цій функції кожен може майнити віртуальні монети;
- Анонім. На відміну від традиційних електронних валют, де транзакції легко відстежуються, неможливо отримати інформацію про власника гаманця криптовалюти. Надаються лише обмежені дані про номери гаманців і суми рахунків;

- Незламність. Віртуальну валюту можна зламати, підробити або виконати інші подібні операції – вона надійно захищена;

– У більшості випадків комісія стягується повністю на добровільній основі;

– У більшості випадків криптовалюта випускається в обмеженому обсязі, що привертає до себе більше уваги.

До недоліків криптовалюти відносять як правило наступні етапи:

– Складність аналізу курсу криптовалюти;

– Високі ризики відміни або заборони в різних країнах;

– Високий ризик втрати. Якщо втрати ключ доступу або пароль до свого гаманця цифрова валюта стане недоступна;

– Відсутні зафіксовані організації які займаються торгівлею.

Таким чином, з таблиці 1.1 відображені переваги та недоліки криптовалют в порівнянні зі банківськими грошима.

Таблиця 1.1 – Переваги та недоліки

Переваги	Недоліки
Можливість видобутку криптовалюти кожному бажаючому в будь-якому куточку світу	Відсутність гарантій збереження криптовалюти
Анонімність транзакцій. Інформація про власника крипто гамане відсутня (є тільки номер гаманця)	Висока мінливість криптовалют
Відсутність єдиного цифрового банку.	Національні регулятори можуть вжити проти нього негативних заходів.
Криптовалюта не схильна до інфляції (імітується обмежена кількість монет).	Втрата пароля до електронного крипто гаманець призводить до безповоротної втрати всіх, хто знаходиться в ньому криптомонет.
Захищеність криптовалют: криптовалюту не можна скопіювати.	Зі зростанням рівня складності стає не вигідно майнити криптовалюту на пристрої окремого користувача.

1.3 Огляд існуючих технологій прогнозування цифрових валют

Основною задачею науковців, які займаються дослідженнями в даній сфері – це визначення системи критеріїв та показників, які чинять вплив, проте цього замало. Так, розглядаючи тему прогнозування криптовалют варто звернутися до теоретичних аспектів, що освітлюються у працях науковців.

Навчання з підкріпленням є окремим випадком навчання під наглядом, де вчитель є «оточенням». Машини (у цьому контексті часто називають «агентами») не мають попередніх знань про середовище, але мають здатність виконувати довільні дії в середовищі. Середовище реагує на ці дії та надає агенту дані, які дозволяють йому реагувати та навчатися. Фактично агент і середовище утворюють систему зворотного зв'язку[3].

Термін навчання з учителем, розуміється, що людина або компанія яка навчає систему, розділяє дані, приводить самій програмі або машині або програмі конкретні приклади та очікує або спостерігає за її результатом. (табл.1.2).

Таблиця 1.2 – Переваги та недоліки методології дослідження криптовалют

Метод	Переваги	Недоліки
Системний підхід	<ul style="list-style-type: none"> - Опис системи; - Розкриває взаємозв'язки з іншими системами; - Описує процеси всередині системи; - Надає рекомендації щодо покращення процесів; - Визначає основні фактори, що впливають на систему. 	<ul style="list-style-type: none"> - Не надає кількісного аналізу системи; - Потребує більшого аналізу при зміні компонентів системи; - Важко застосувати при описі нових структур
Нейронні мережі	<ul style="list-style-type: none"> - Дає хорошу модель після проведення багатьох ітерацій дослідження - Досить проста модель в реалізації 	<ul style="list-style-type: none"> - Не враховуються фактори, які впливають на показник.

Отже, прогнозування курсу Bitcoin є складною задачею, яка потребує дослідження.

Станом на сьогодні немає єдиного підходу до роботи з криптовалютами як в Україні, так і за кордоном. Криптовалюта як така незаборонена, проте її статус не визначений законодавством. Криптовалюта знаходиться в правовому вакуумі. Вона не є платіжним засобом, відтак не внесене до національного Реєстру платіжних систем. Відповідно до статті 99 Конституції України [1], статті 192 Цивільного процесуального кодексу України [2] та статті 3 Закону України «Про платіжні системи та переказ коштів в Україні» [3], єдиним законним платіжним засобом, який приймається усіма фізичними та юридичними особами без будь-яких обмежень на території країни для проведення переказів та розрахунків є гривня як грошова одиниця України.

Розумні вкладення в криптовалюти здатні принести великі дивіденди. Для того, щоб заробити на криптовалютному трейдингу необхідно вміти правильно спрогнозувати майбутній рух цін.

Для цих цілей можна використовувати NeuroShell або Trader.

NeuroShell це спеціалізований нейропакет, призначений спеціально для прогнозування фінансових ринків. Така політика розробників зумовила дружній для сприйняття інтерфейс і можливість працювати з нейронними мережами без навичок програмування.

Trader є система дозволяє прогнозувати рух курсів валют на валютній біржі. В якості вихідних використовується інформація за попередніми результатами торгів (часовий ряд) – максимальна, мінімальна ціна, ціна закриття і обсяг угод за день [4].

1.4 Формалізація постановки задачі

Отже, всі датасети з даними з криптовалютних ринків містять базові ознаки.

OHLCV – це агрегована форма ринкових даних, що розрізняють Open, High, Low, Close та Volume. Дані OHLCV включають 5 точок даних: Відкриття і Закриття представляють перший і останній рівень ціни протягом заданого

інтервалу. High та Low представляють найвищу та найнижчу досягнуту ціну протягом цього інтервалу. Обсяг – це загальна сума торгівлі за цей період. Ці дані найчастіше представлені на свічковому графіку, що дозволяє трейдерам проводити технічний аналіз внутрішньо денних значень. Ми надаємо дані OHLCV у деталях від 1 секунди до 1 дня.

- «Open» – Ціна відкриття інтервалу часу у валюті котирування (для BTC/USD ціна буде USD).
- «High» – Найвища ціна, досягнута протягом інтервалу часу, у валюті котирування.
- «Low» – Найнижча ціна, досягнута протягом інтервалу часу, у валюті котирування.
- «Close» – Ціна закриття інтервалу часу, у валюті котирування.
- «Volume» – Quantity of asset bought or sold, displayed in base currency.

Базові ознаки Bitcoin подані на рисунку 1.3:

Date	High	Low	Open	Close	Volume
2021-01-01	6612.573730	6202.373535	6437.319336	6606.776367	40346426266
2021-01-02	7088.247559	6595.918457	6606.776367	6793.627512	47660646124
2021-01-03	7003.220703	6673.335938	6797.396484	6733.387207	38976504903
2021-01-04	6878.953613	6696.484863	6738.382812	6867.527344	33185988584
2021-01-05	6883.414062	6715.929199	6862.537598	6791.129395	29510409856
...
2021-12-28	50679.859375	47414.210938	50679.859375	47588.855469	33430376883
2021-12-29	48119.742188	46201.496094	47623.871094	46444.710938	30049226299
2021-12-30	47879.964844	46060.312500	46490.605469	47178.125000	26686491018
2021-12-31	48472.527344	45819.953125	47169.371094	46306.445312	36974172400
2022-01-01	47827.312500	46288.484375	46311.746094	47686.812500	24582667004

Рисунок 1.3 – Базові ознаки датасету по криптовалюти Bitcoin за даними сайту

Yahoo

API на Yahoo містить ще «Adj Close», але, ми можемо знехтувати їм або видалити оскільки параметри «Close» дорівнюють їм.

Саме за даною ознакою, відображає курс біткоїну або інших криптовалют на ринку.

Перш за все важливо приділити увагу формуванню тренувальних датасетів, валідаційних датасетів для відбору оптимальної моделі та тестових для надання оцінки точності прогнозу.

Відповідно до поставленої в індивідуальному завданні задачі, під час розв'язання задачі методами машинного навчання цільову ознаку «Target» слід обчислювати по курсу в кінці кожної доби «Close» за такою формулою:

$$Target_t = Close_{t-d}, \quad (1.1)$$

В даній формулі t – це час або днів; d – їхня кількість, на які перш за все вчать модель робити прогноз.

Іншою особливістю моделей часових рядів є те, що як тестовий, так і перевірений набори даних вибираються в кінці спостережуваних даних. Основна мета набору даних перевірки — дозволити нам вибрати найкращу модель, яка найкраще прогнозує майбутні значення, а не минулі значення.

Тестовими даними зазвичай є останні d (forecasting_days) значення в серії спостережень. Верифікація (для перевірки) – це те, що стежить за вами, коли ви проходите повз. Всі попередні значення формують навчальний (навчальний) набір даних.

Після визначення найкращої моделі, яка забезпечує найкращі прогнози щодо даних перевірки для даного показника (або множини), її слід повторно обчислити на розширеному наборі даних навчання, який є комбінацією наборів даних навчання та перевірки. Потім на основі цієї вдосконаленої найкращої моделі ми створюємо прогнози даних на основі тестового набору даних.

Курс криптовалюти можна прогнозувати методами машинного навчання з використанням дерев рішень, нейронних мереж та інших моделей, які

враховують вплив багатьох ознак, або – як часовий ряд, для чого використати моделі авторегресії та проінтегрованого ковзного середнього (англ. скорочено: «ARIMA») та їх аналоги або Facebook Prophet з різними параметрами сезонності та коефіцієнтів Фур'є для врахування ступеню їх не лінійності (рис. 1.4).

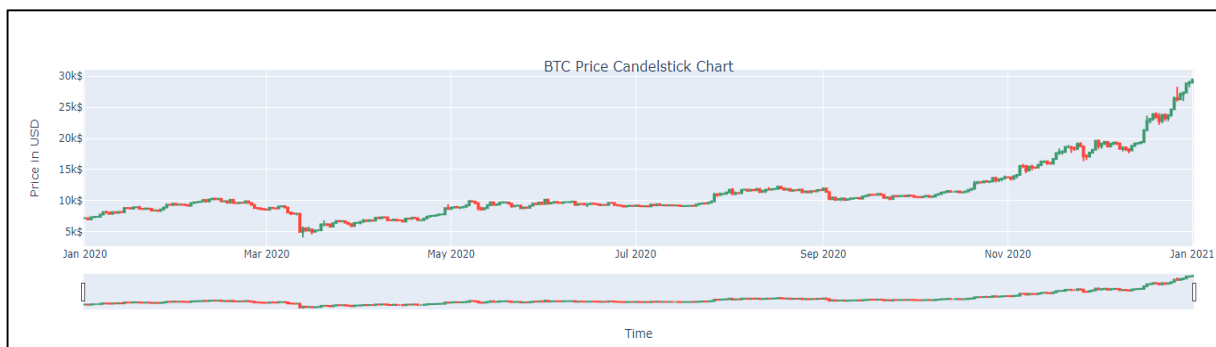


Рисунок 1.4 – Прогнозування часового ряду

Етапи прогнозу:

1. Очищення та попередня обробка.

Спочатку ми очищаємо дані, перетворюючи дату в правильний формат і перевіряючи значення null, як ми це робили раніше. Наступна команда дає нам кількість нульових значень у кожному стовпці. Ми виявили, що лише стовпець обсягу торгівлі містить нульові значення.

2. Аналіз.

Спочатку ми робимо графіки розподілу для всіх функцій за допомогою оцінки щільності ядра, яка показала нам, що дані, здавалося, були сильно спотвореними. Для цього ми візуалізуємо перекоси кожної з цих ознак.

Дослідницька група Facebook розробила простішу реалізацію прогнозування часових рядів за допомогою своєї нової бібліотеки під назвою Prophet (Провісник). Рідко можна зустріти аналітика, здатного надавати якісні прогнозні дані. Це одна з причин, через яку дослідницька група Facebook вигадала легко доступний спосіб використання передових концепцій для прогнозування часових рядів.

Ми можемо легко познайомитися з цією бібліотекою, тому що вона використовує API, схожу на Scikit-Learn. Основна мета команди Prophet – дати можливість експертам та нефаківцям робити якісні прогнози, що задовольняють існуючі запити.

На мій погляд, Prophet найкраще працює із наборами даних, на які сильно впливає сезонність, наприклад, комунальні платежі, кількість відвідувачів ресторанів тощо.

1.5 Висновки

У даному розділі був проведений тематичний аналіз з огляду на актуальність цифрових валют на світовому ринку. Проаналізована історія від народження віртуальної валюти до сьогодення. Порівняння програмних засобів для прогнозування обмінних курсів криптовалют висвітлює їх сильні та слабкі сторони. Розглянуто проблеми прогнозування курсу криптовалюти Bitcoin здійснено формалізацію постановки задачі та вибрано оптимальні інформаційні технології.

2 ВИБІР ОПТИМАЛЬНИХ НАЛАШТУВАНЬ ДЛЯ РОЗВ'ЯЗАННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ

2.1 Огляд технології Kaggle середовища

Для вирішення поставленого завдання було запропоновано, скористатись інтернет середовищем Kaggle.

Kaggle – популярна платформа для змагань з Data Science від Google. Користувачі (люди та організації) можуть публікувати на ній свої набори даних, створювати та досліджувати моделі машинного навчання, змагатися один з одним [5].

Типова схема Kaggle-змагання: організатор конкурсу (як правило, велика компанія) публікує свої дані та опис проблеми, терміни, критерії правильного вирішення та приз, який отримає переможець. А учасники намагаються застосувати до даних різні методи, моделі та алгоритми, щоб вирішити проблему. Автор найкращого рішення отримує приз. Решта учасників – бали залежно від місця в таблиці з результатами. Із цих балів складається загальний рейтинг на платформі. Оцінка робіт є прозорою: платформа автоматично перевіряє рішення за критеріями, виставленими компанією-організатором [6].

На рисунку 2.1 зображено зовнішній вигляд інтерфейсу користувача в Kaggle

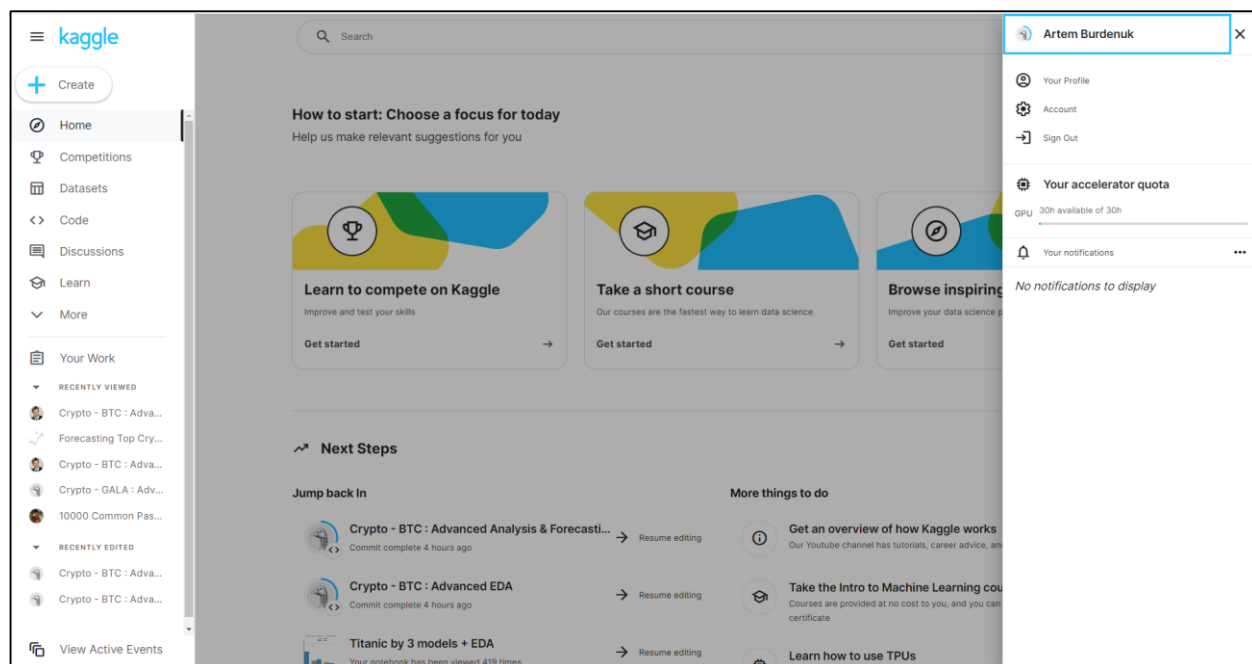


Рисунок 2.1 – Зовнішній вигляд інтерфейсу користувача в Kaggle

Середовище організоване як публічна веб-платформа, на якій користувачі та організації можуть публікувати набори даних, досліджувати та створювати моделі, взаємодіяти з іншими фахівцями з даних та інженерами з машинного навчання, організовувати конкурси з дослідження даних та брати участь у них. У системі розміщено набори відкритих даних, надаються хмарні інструменти для обробки даних та машинного навчання. Також реалізовані навчальні ресурси, є розділ для розміщення вакансій роботодавцями, де також можлива організація конкурсів для відбору найкращих кандидатів [5].

Окрім публічних конкурсів також організовуються закриті змагання, у яких беруть участь лише фахівці з певним рейтингом Kaggle. Крім того, пропонується безкоштовний інструмент для вчителів інформатики для проведення академічних змагань з машинного навчання (Kaggle In Class).

На рисунку 2.2 зображено датасет, на основі якого було виконано дану роботу.

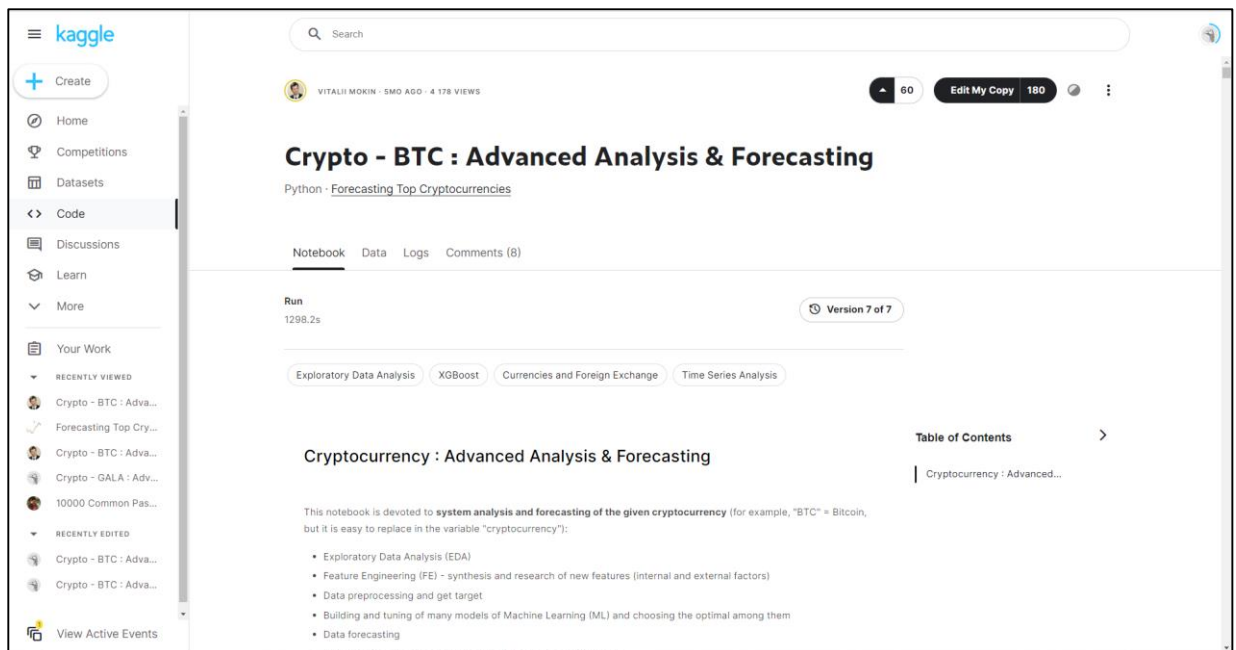


Рисунок 2.2 – Датасет у середовищі Kaggle

Усього проведено понад сотню публічних конкурсів з машинного навчання, серед них змагання щодо покращення розпізнавання жестів для Microsoft Kinect, конкурс із удосконалення системи пошуку бозона Хігса в ЦЕРНі.

Результати деяких конкурсів вилилися у великі проекти, серед них технології досліджень ВІЛ, шахові рейтинги, прогнозування завантаженості магістралей. Серед учасників конкурсів – Джеффри Хінтон та Джордж Даль, які виграли з використанням глибоких нейронних мереж змагання корпорації Merck; їх результат, поряд з перемогою учня Хінтона Влада Мніх на конкурсі Adzuna, послужив визнанням глибокого навчання як універсальної техніки, яка надалі широко поширилася серед інших учасників змагань. На основі результатів, отриманих на змаганнях Kaggle, було опубліковано кілька наукових праць [5].

2.2 Синтез нових ознак

Розглянемо ковзне середнє (MA). Технічні показники ковзного середнього є одними з найбільш простих інструментів, що використовуються в технічному аналізі. Вони є середньою ціною активу за певний період. Вони вказують на те,

що актив має тенденцію вгору в позитивному (бичачого) напрямку або рухається вниз у негативному (ведмежому) напрямку [6].

Їх також називають відстаючими індикаторами, оскільки вони слідують тренду і забезпечують відкладений зворотний зв'язок після того, як рух ціни вже відбувся. Найпопулярнішими є прості та експоненційні ковзні середні, які також є найпоширенішими моделями, що використовуються в криптовалютах.

Розглянемо індикатор дивергенції ковзної середньої (або осцилятор) є хорошим простим індикатором імпульсу для купівлі або продажу та одним з найпопулярніших інструментів, які використовуються крипто-трейдерами.

Дивергенція MACD відноситься до двох базових ковзних середніх, що розходяться, тоді як конвергенція відноситься до двох базових ковзних середніх, що наближаються один до одного [7].

Розглянемо MACD і RSI є цінними інструментами, за допомогою яких технічні трейдери можуть аналізувати графіки цін на актив, шукаючи закономірності, які вказують, коли купувати або продавати криптовалюту.

Індикатор відносної сили — це ще один індикатор імпульсу технічного аналізу, який оцінює слабкість або силу криптовалюти за останніми змінами ціни, щоб визначити умови перекупленості або перепроданості активу. Індикатор також підходить для виявлення дивергенції та попередження трейдерів про можливий розворот тренду [6].

Смуги Боллінджера є технічним індикатором ціни та волатильності криптоактиву з плином часу і беруть свою назву від Джона Боллінджера, технічного трейдера, який їх розробив. Вони складаються з трьох основних компонентів: простої ковзної середньої (зазвичай 20-періодної SMA), верхньої смуги та нижньої смуги, зазвичай на відстані двох стандартних відхилень від SMA. Стандартне відхилення обчислює різницю між групою цінностей або цін від середньої вартості/ціни [7].

Бібліотека TSFRESH допомагає автоматично синтезувати наші ознаки для заданих часових рядів: місяць, дисперсія середнє та ексцес.

На рисунку 2.3 – 2.4 відображені результат бібліотеки TSFRESH до ознаки «Close» після застосування фільтру, який видаляє ознаки середньоквадратичного відхилення яких дорівнює нулю, тобто які є константами для усього ряду спостережень.

```
['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',
```

Рисунок 2.3 – Нові ознаки для ознаки «Close» криптовалюти біткоїн, згенеровано бібліотекою TSFRESH


```

'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',

```

Рисунок 2.4 – Нові ознаки для ознаки «Close» криптовалюти біткоїн, згенеровані бібліотекою TSFRESH (продовження)

Отже, було вибрано 57 ознак, у т.ч. 5 базових та 50 ознак з TSFRESH, на основі базової інформації крипторинків.

2.3 Розвідувальний аналіз даних

Проведемо розвідувальний аналіз даних курсу криптовалюти Bitcoin за 2021-2022 рр.

На рисунку 2.5 – 2.9 наведено основні статистичні характеристики цієї випадкової величини.

Dataset statistics		Variable types	
Number of variables	7	Date Time	1
Number of observations	638	Numeric	6
Missing cells	0		
Missing cells (%)	0.0%		
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	35.0 KiB		
Average record size in memory	56.2 B		

Рисунок 2.5 – Результат розвідувального аналізу даних курсу Bitcoin за даними сайту Yahoo

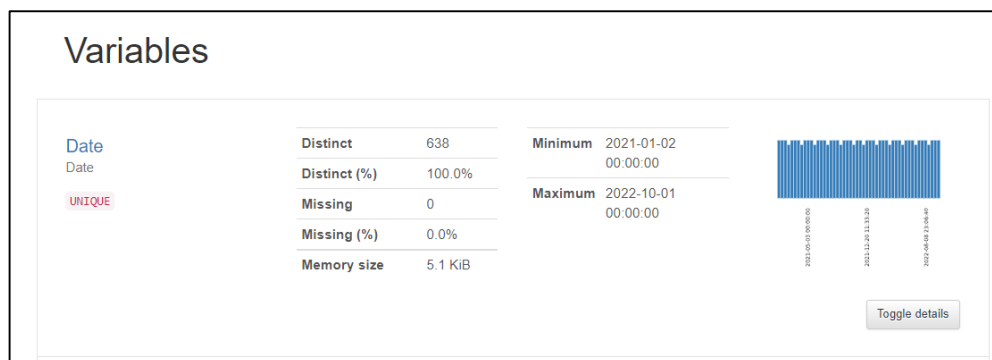


Рисунок 2.6 – Результат розвідувального аналізу даних курсу Bitcoin за даними сайту Yahoo

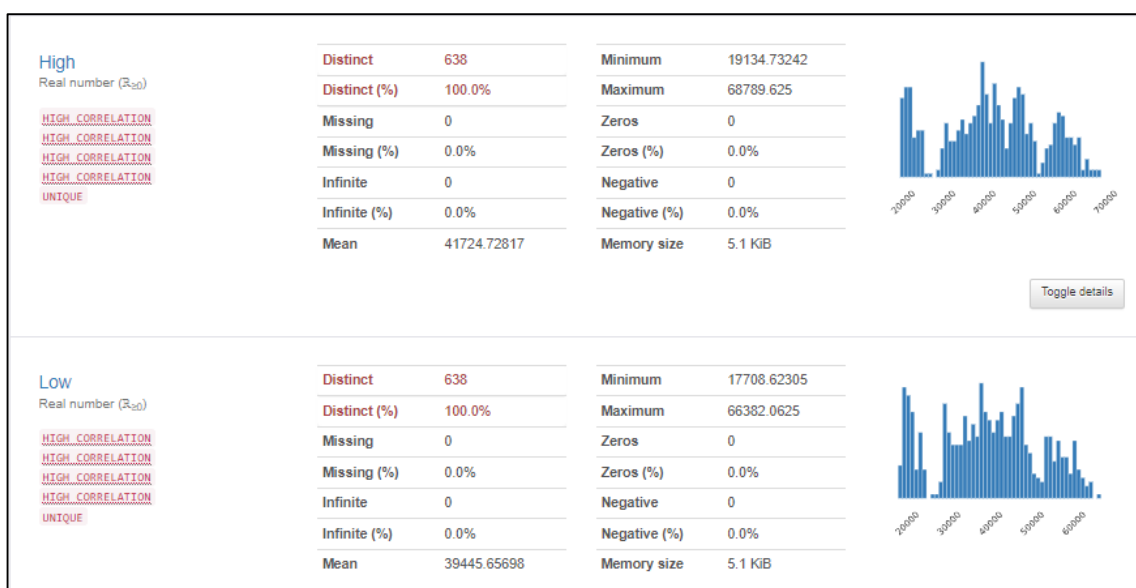


Рисунок 2.7 – Результат розвідувального аналізу даних курсу Bitcoin за ознаками High, Low

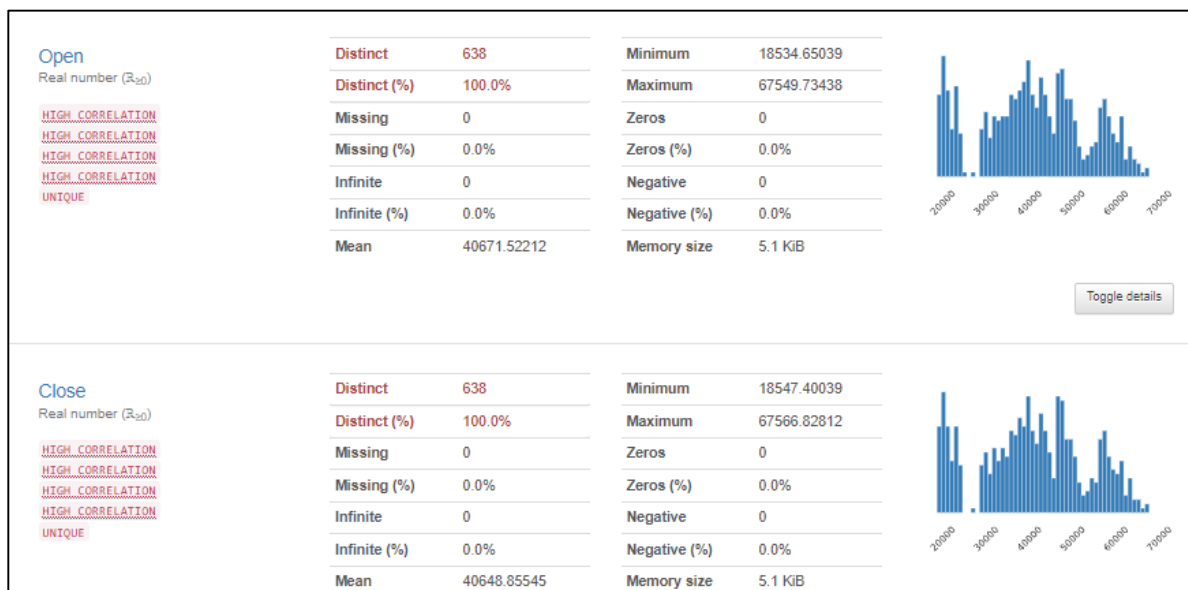


Рисунок 2.8 – Результат розвідувального аналізу даних курсу Bitcoin за ознаками Open, Close

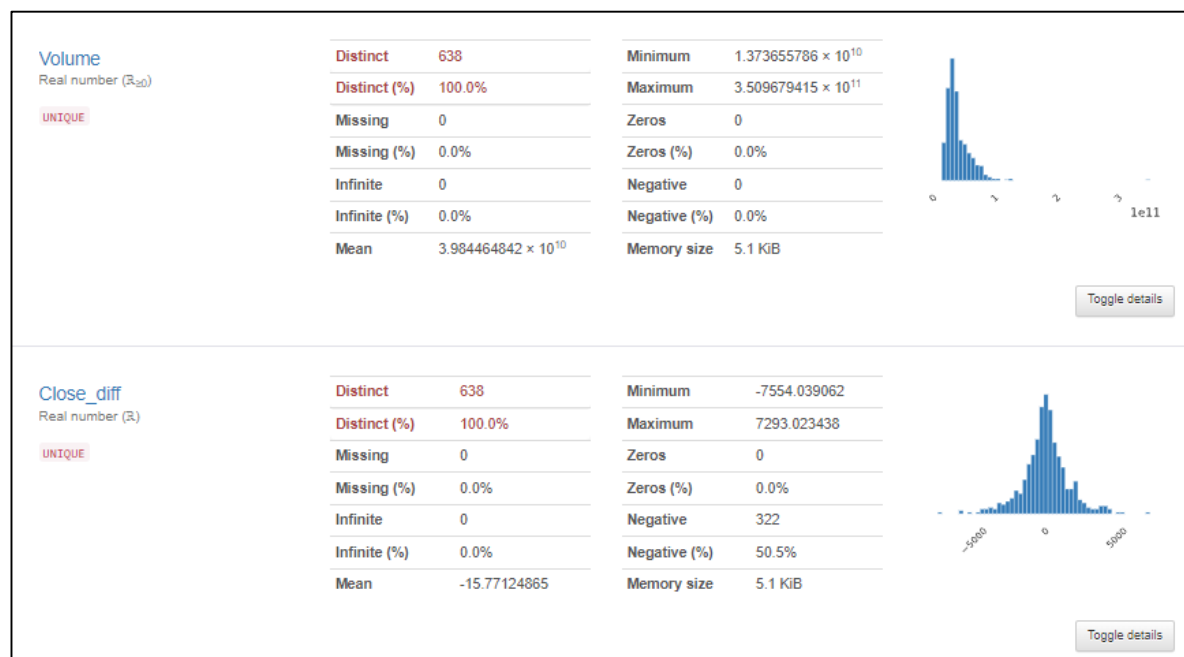


Рисунок 2.9 – Результат розвідувального аналізу даних курсу Bitcoin за ознаками Volume, Close_diff

Використовуючи ознаки Open, high, Low, Close яка має назву «OHLC-ознаки» будують OHLC-діаграму. Це діаграма метою якої є перегляд падіння свічок та підняття свічок. Червоний колір символізує падіння а зелений підйом.

На рисунку 2.10 зображена OHLC-діаграма криптовалюти Bitcoin за даними сайту Yahoo.

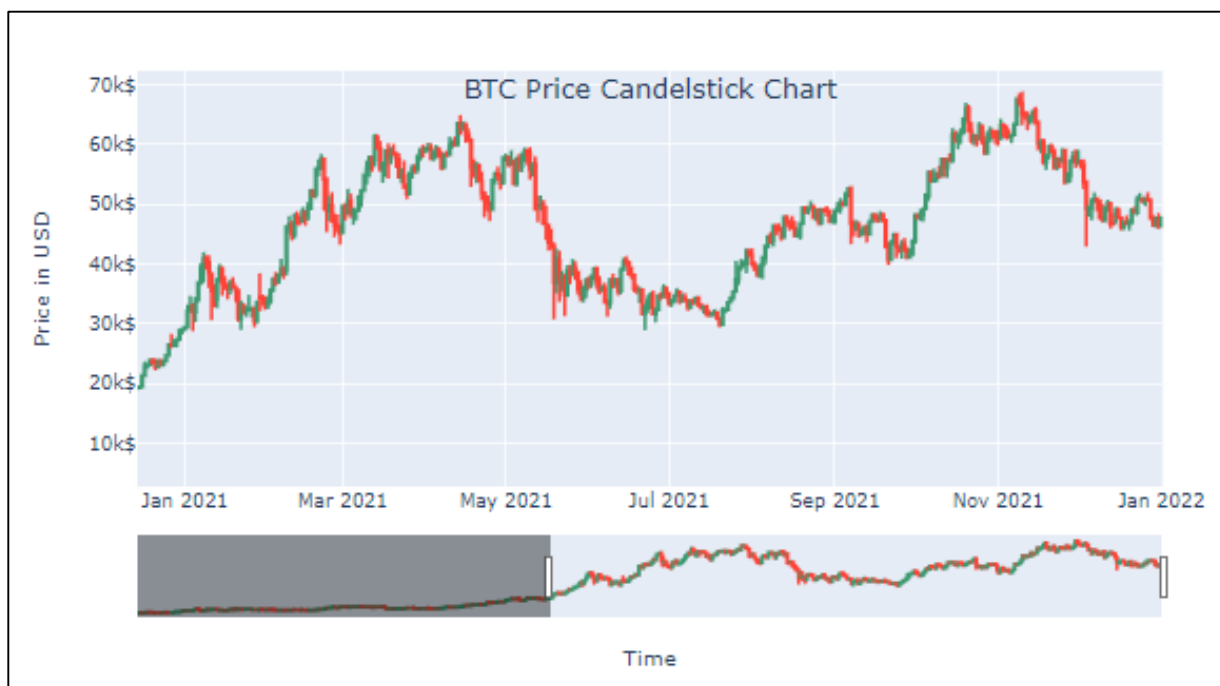


Рисунок 2.10 – OHLC-діаграма криптовалюти Bitcoin

Проаналізуємо ряд значень ознак для відповіді на подані питання:

- 1) Можливо стаціоналізувати ряд взяттям першої чи другої різниці?
- 2) Чи має ряд сезонність, якщо відповідь так, то скільки діб може знаходитись в її періоді?

Для перевірки чи є ряд стаціонарним використовується критерій Діка-Фуллера (Dickey-Fuller) [8]. Тест Дікі-Фуллера – є одним із тестів на одиничні корені (Unit root test). Статистика тесту (DF-статистика) – це проста t-статистика для перевірки значущості коефіцієнтів лінійної регресії.

На рисунку 2.11 зображено програмний код для перевірки на стаціонарність курсу криптовалюти Bitcoin збереженого як ряд даних series.

```

def check_stationarity(series):
    # Thanks to https://machinelearningmastery.com/time-series-data-stationary-python/

    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 2.11 – Програмний код для перевірки на стаціонарність курсу криптовалюти Bitcoin

Результат застосування цього критерію до ознаки «Close» наведено на рисунку 2.12.

```

ADF Statistic: -1.257597
p-value: 0.648342
Critical Values:
    1%: -3.441
    5%: -2.866
    10%: -2.569
Non-stationary

```

Рисунок 2.12 – Результат перевірки на стаціонарність курсу криптовалюти Bitcoin

Як видно, з рисунку 2.10, можна зробити висновок про нестаціонарність. Нульова гіпотеза є нестаціонарним, має значущість p-value 0,65, тобто набагато більше, ніж 0,05.

Отже, слід спробувати стаціонаризувати цей ряд взяттям першої (рис. 2.13) і другої (рис. 2.14) різниці.

```

ADF Statistic: -26.335628
p-value: 0.000000
Critical Values:
    1%: -3.441
    5%: -2.866
    10%: -2.569
Stationary

```

Рисунок 2.13 – Результат перевірки на стаціонарність першої різниці курсу криптовалюти Bitcoin

```

ADF Statistic: -9.633372
p-value: 0.000000
Critical Values:
    1%: -3.441
    5%: -2.866
    10%: -2.569
Stationary

```

Рисунок 2.14 – Результат перевірки на стаціонарність другої різниці курсу криптовалюти Bitcoin

Як бачимо, і перша, і друга різниці ряду Bitcoin за 2021-2022 рр. є стаціонарними часовими рядами. Обов'язковим є якомога менший порядок, для цього і далі моделюємо та прогнозуємо саме першу різницю «Close_diff», яка навпроти «Close» вже є стаціонарною (рис. 2.15).

Date	High	Low	Open	Close	Volume	Close_diff
2021-01-02	7088.247559	6595.918457	6606.776367	6793.624512	47660646124	186.848145
2021-01-03	7003.220703	6673.335938	6797.396484	6733.387207	38976504903	-60.237305
2021-01-04	6878.953613	6696.484863	6738.382812	6867.527344	33185988584	134.140137
2021-01-05	6883.414062	6715.929199	6862.537598	6791.129395	29510409856	-76.397949
2021-01-06	7271.781250	6782.889648	6788.049805	7271.781250	46896904615	480.651855
...
2021-12-28	50679.859375	47414.210938	50679.859375	47588.855469	33430376883	-3051.562500
2021-12-29	48119.742188	46201.496094	47623.871094	46444.710938	30049226299	-1144.144531
2021-12-30	47879.964844	46060.312500	46490.605469	47178.125000	26686491018	733.414062
2021-12-31	48472.527344	45819.953125	47169.371094	46306.445312	36974172400	-871.679688
2022-01-01	47827.312500	46288.484375	46311.746094	47686.812500	24582667004	1380.367188

Рисунок 2.15 – Результат обчислення першої різниці курсу криптовалюти Bitcoin

Варто зазначити, що цільову ознаку для часових рядів доцільно використовувати саме «Close», а для класичних моделей машинного навчання на кшталт 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 т.ч. – на сезонність, з використанням ковзного середнього [8] (рис. 2.15). Перевірка на сезонність полягає у декомпозиції ряду на тренд «Trend», сезонну складову «Seasonal» та залишок «Resid» (рис. 2.14).

Взагалі модель ARIMA означає інтегроване ковзне середнє з авторегресією. Ця модель надає набір функцій, які є дуже потужними та гнучкими для виконання будь-яких завдань, пов'язаних із прогнозуванням часових рядів [9].

У машинному навчанні модель ARIMA зазвичай є класом статистичних моделей, які дають вихідні дані, які лінійно залежать від їхніх попередніх значень у комбінації стохастичних факторів [9].

При виборі відповідної моделі прогнозування часових рядів нам необхідно візуалізувати дані для аналізу тенденцій, сезонності та циклів. Коли сезонність є дуже сильною характеристикою часових рядів, необхідно розглянути таку модель, як сезонна ARIMA (SARIMA).

Модель ARIMA працює з використанням моделі розподіленого запізнення, в якій алгоритми використовуються для прогнозування майбутнього на основі значень, що запізнюються.

У машинному навчанні модель сезонного авторегресійного інтегрованого ковзного середнього (SARIMA) є кроком, відмінним від моделі ARIMA, заснованої на концепції сезонних трендів [10].

Сезонні варіації часових рядів можуть ураховувати періодичні моделі, що дозволяє робити точніші прогнози. Сезонний ARIMA (SARIMA) визначає як сезонний, так і несезонний компонент моделі ARIMA, що дозволяє фіксувати періодичні характеристики [10].

Вибираючи відповідну модель прогнозування, завжди візуалізуйте свої дані, щоб визначати тенденції, сезони та цикли. Якщо сезонність є сильною стороною ряду, розгляньте моделі із сезонним коригуванням, такі як модель SARIMA.

На рисунку 2.16 зображений результат декомпозиції часового ряду курсу криптовалюти на складові: на тренд «Trend», сезонну складову «Seasonal» та залишок «Resid».

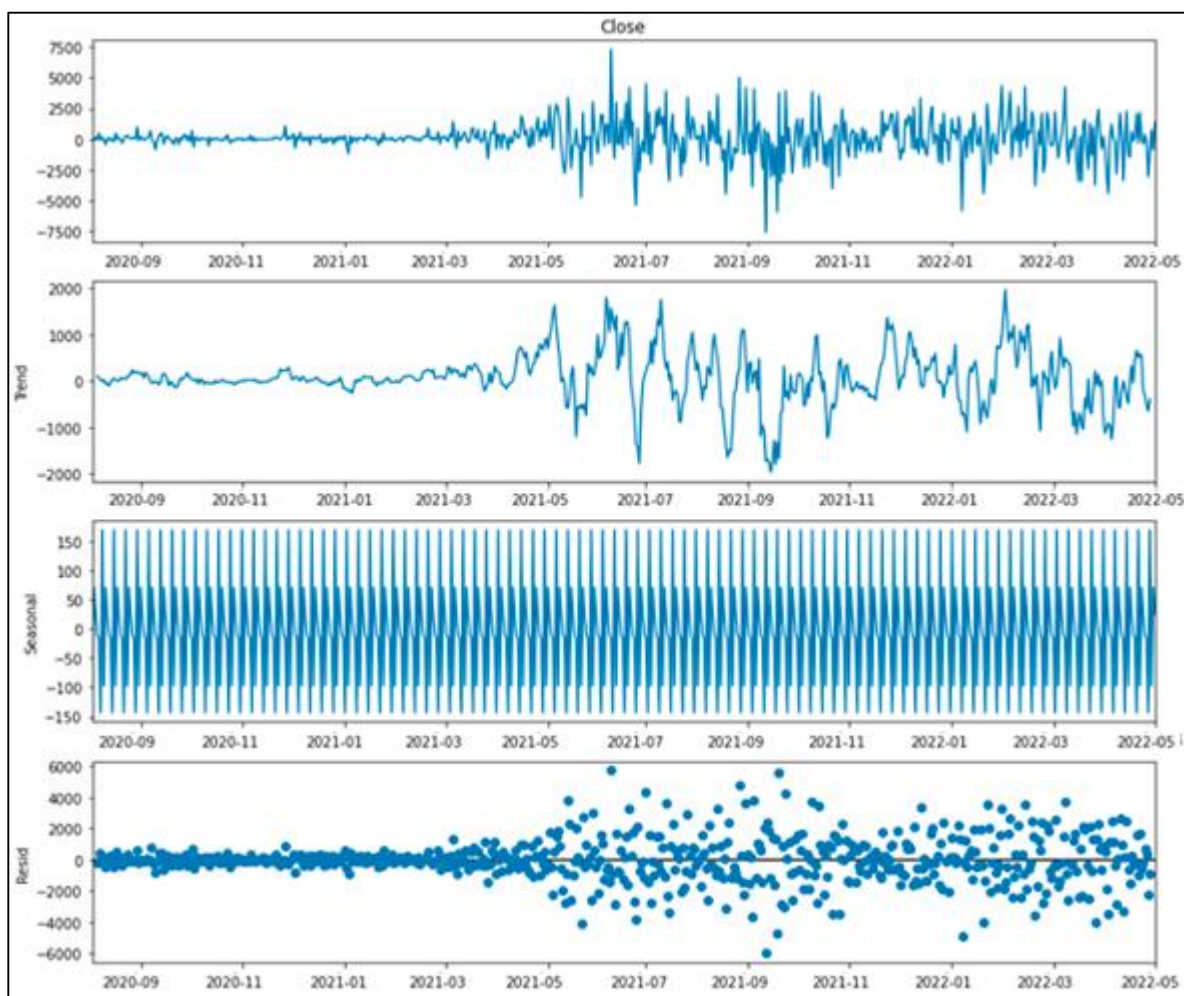


Рисунок 2.16 – Результат декомпозиції часового ряду курсу криптовалюти

Через велику кількість даних важко точно визначити період сезонності. Для вирішення цієї проблеми потрібно ще раз здійснити декомпозиції часового ряду за менший проміжок часу. Наприклад зробити декомпозицію та візуалізувати тільки дані за один місяць.

На рисунку 2.17 зображений результат декомпозиції часового ряду курсу криптовалюти Bitcoin за грудень 2021 року на складові, у т.ч. – на Trend та сезонну складову.

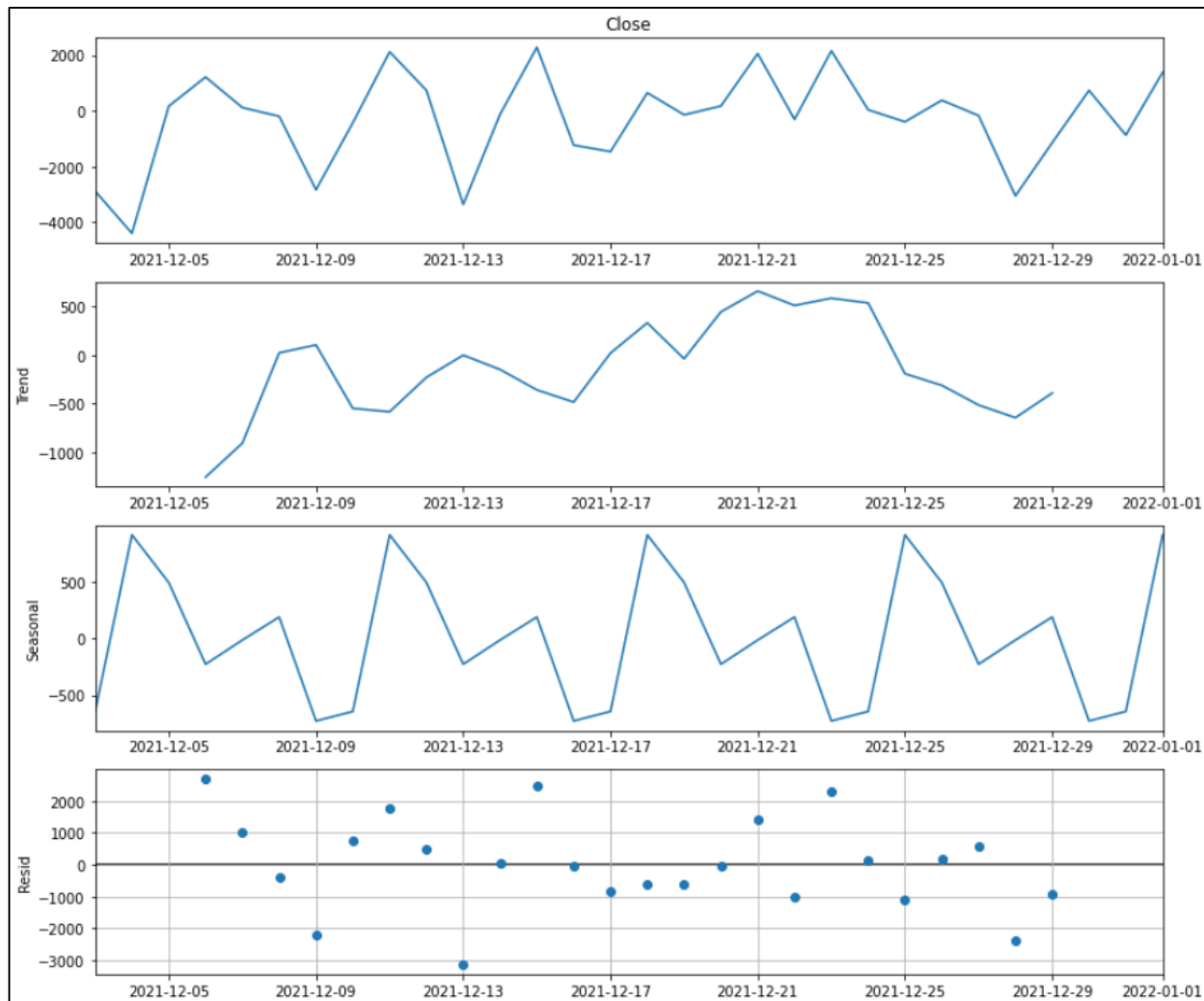


Рисунок 2.17 – Результат декомпозиції часового ряду курсу криптовалюти Bitcoin за останій місяць 2021 року на складові

На рисунку 2.18 зображений курс біткоіна за даними сайту Yahoo.

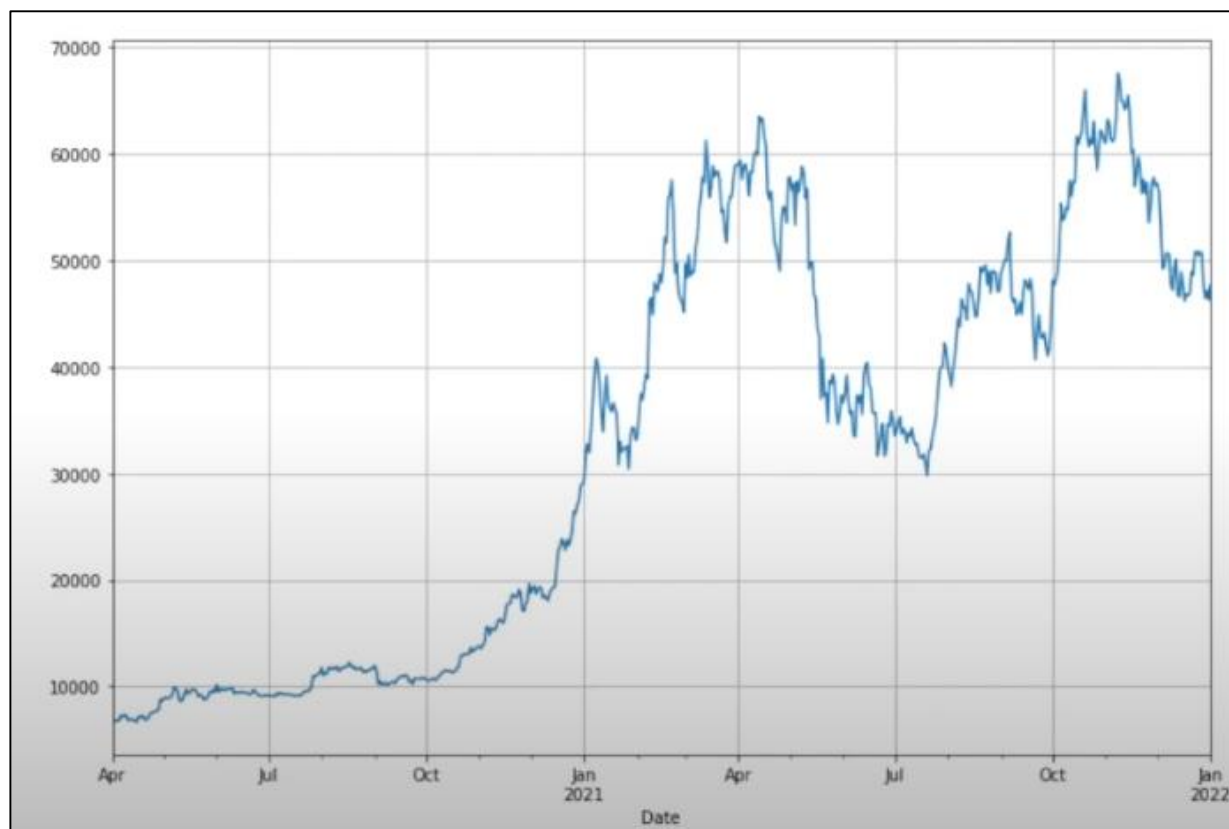


Рисунок 2.18 – Курс біткоїна за даними сайту Yahoo

Проаналізуємо який відсоток значень описує складова «Seasonal». Розглянемо весь ряд спостережень на рисунку 2.17. Як з нього видно, максимальна амплітуда значень цієї складової від -800 до 1000, тобто – десь 1800, у т.ч. час як основного ряду десь 70 тисяч. Отже, сезонна складова описує приблизно 2,5% даних, тобто нею можна знехтувати.

Проаналізуємо аномалії ринку. Аналіз періоду досліджень показав, що мав місце ряд аномалій (рис. 2.19-2.21).

Перша аномалія виявлена 8.01.2021 р. Відтоді індекс блакитних фішок Dow Jones Industrial Average впав більш ніж на 1000 пунктів, а Nasdaq впав на 5%. Це було найгірше денне падіння з 2020 року: о 19:00 біткойн впав на 5% до 32 860,91 долара. Ця аномалія призвела до великих втрат серед трейдерів та людей які зберігали свої кошти у криптовалюти [12].

На рисунку 2.20 відображена аномалія 8.01.2021 р. різкої зміни курсу криптовалюти.

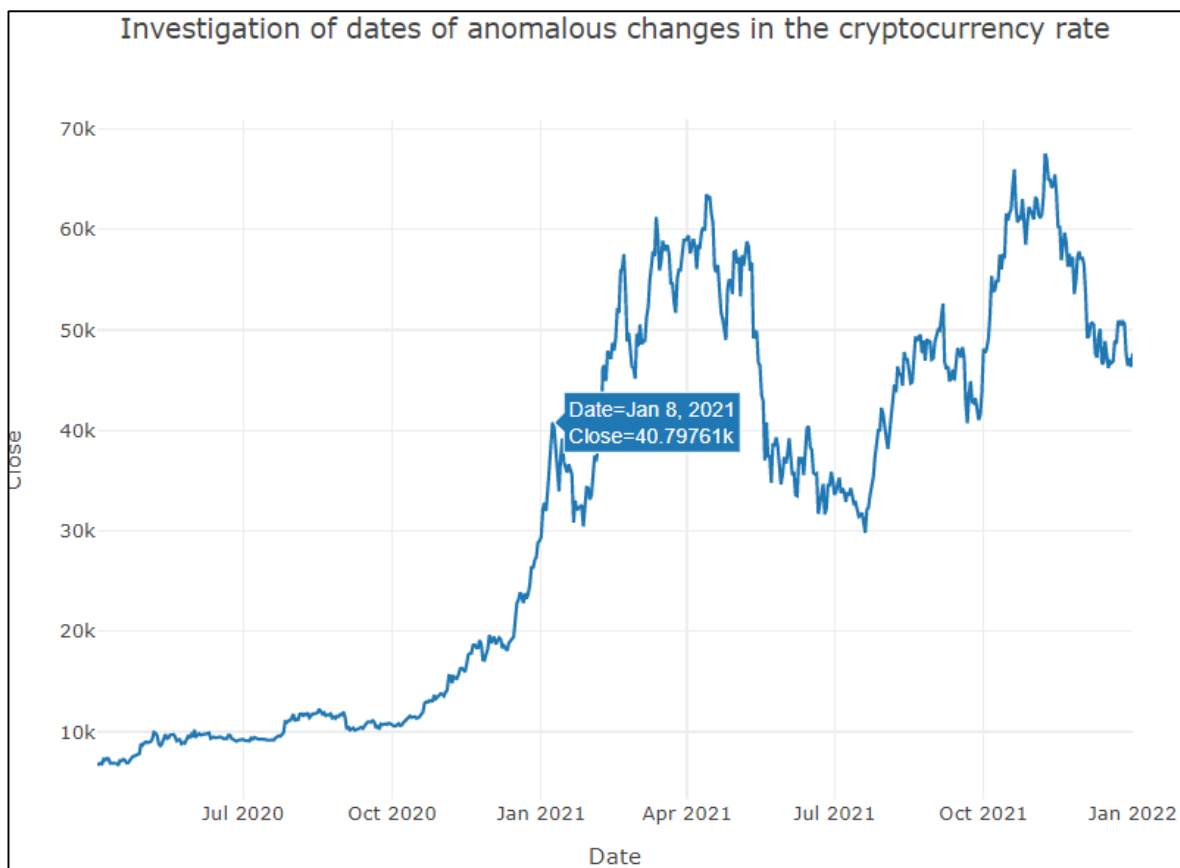


Рисунок 2.19 – Аномалія 8.01.2021 р. різкої зміни курсу криптовалюти

Друга аномалія виявлена 13.04.2021 р. Найпоширеніша криптовалюта Bitcoin 13 квітня досягла нового рекорду 62,732 долара, що пов'язано з виходом найбільшої криптовалютної біржі США Coinbase на IPO (первинне публічне розміщення акцій).

Coinbase – це криптовалютна біржа, яка була заснована у червні 2012 року. Її головний офіс знаходиться у Сан-Франциско, штат Каліфорнія. Платформа пропонує своїм користувачам електронний гаманець та майданчик для проведення транзакцій із різними криптовалютами [10].

Одна з особливостей Coinbase полягає в тому, що вона найбільш консервативна і наступна законодавчим нормам біржа. Біржа Coinbase працює лише згідно з правилами регуляторів. Вона вже отримала ліцензію на діяльність у багатьох штатах США. Coinbase після проведення IPO на NASDAQ 13 квітня 2021 через прямий лістинг стала першою криптобіржею у світі, котирується на фондовому ринку. Перед початком торгів компанія встановила довідкову ціну на

папери у розмірі \$250, проте перша сесія відкрилася на рівні \$381. Капіталізація на день IPO досягла \$85,8 млрд [10].

Coinbase має великий інтерфейс прикладного програмування (API), який налічує 6 ключових функцій: Wallet API, Coinbase Connect, Wallet Notifications, Merchant Checkouts, Coinbase Exchange API і Toshi.

На рисунку 2.20 відображена аномалія 13.04.2021 р. різкої зміни курсу криптовалюти.

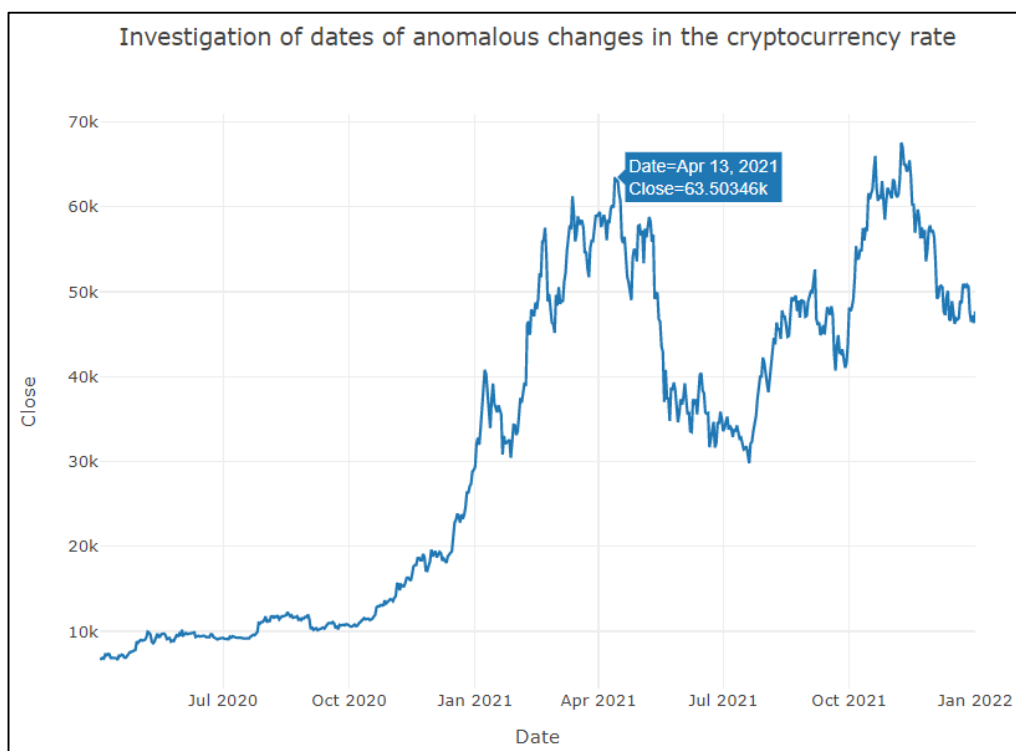


Рисунок 2.20 – Аномалія 13.04.2021 р. різкої зміни курсу

Третя аномалія виявлена 20.07.2021 р. Ціна біткоїну вперше з 22 червня опустилася нижче 30 тисяч доларів – сам біткоїн потяг за собою інші валюти. Валюта впала більш ніж на 5% до 29 750 доларів. З ранку понеділка (в Києві). Згідно з CoinDesk на 11:29 (За даними Києва) Ціна біткоїнів становить 29 847 доларів за штуку. Ethereum впав на 7,8% (\$1748), а жартівлива криптовалюта Dogecoin впала на 8,6% (\$0,16). Як зазначає CNBC, розпродаж біткойна стався після масштабного падіння світових фондових ринків. У понеділок індекс Dow Jones Industrial Average зафіксував найгірший день з жовтня минулого року. З

моменту досягнення максимуму в 65 000 доларів США в середині квітня ціна біткоїна впала більш ніж на 50%.

На рисунку 2.21 відображена аномалія 20.07.2021 р. різкої зміни курсу криптовалюти.

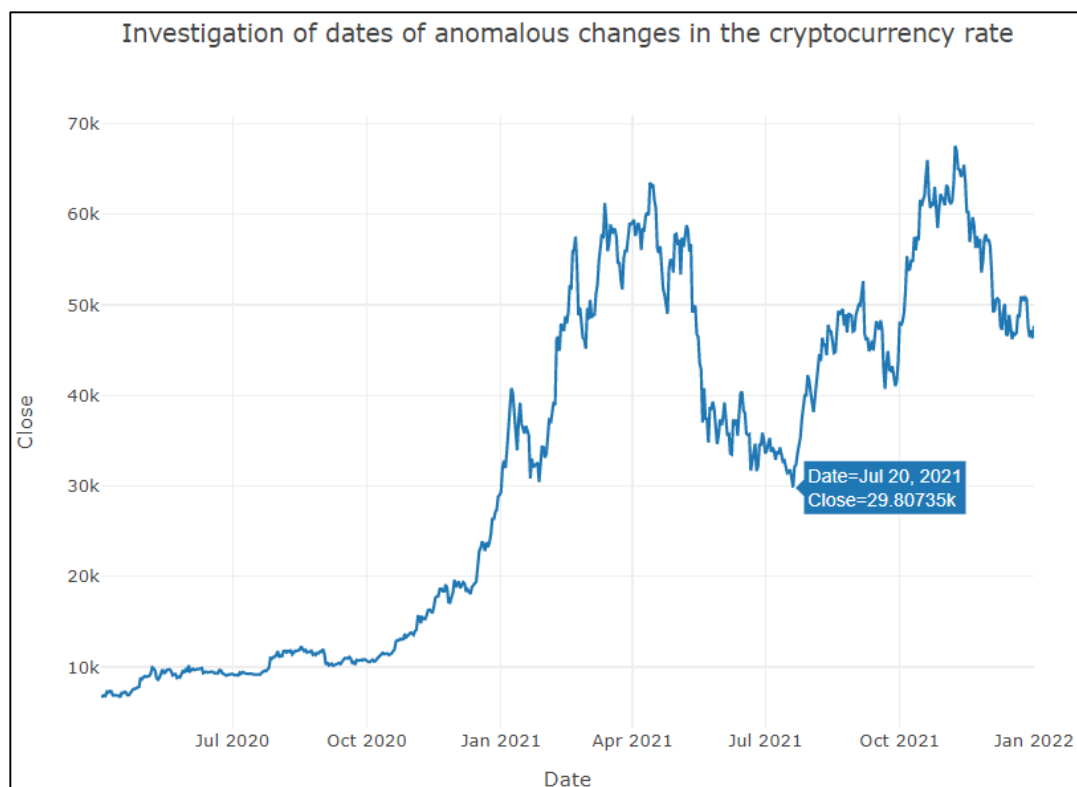


Рисунок 2.21 – Аномалія 20.07.2021 р. різкої зміни курсу

Четверта аномалія виявлена 29.09.2021 р. Як пише в середу газета Financial Times, причиною падіння вартості біткоїна стало повідомлення в акаунті Центробанку Китаю в WeChat. У ньому три асоціації банківської та інтернет-індустрії заявили, що фінансові та платіжні установи не повинні приймати криптовалюти як платежі і надавати пов'язані з ними послуги або продукти.

Зниження на ринку було викликане підвищенням відсоткових ставок центробанками світу, націленим на боротьбу з високою інфляцією, та хвилею банкрутств та боргових проблем, що прокотилася компаніями в секторі криптовалют.

Цього разу підйом на криптовалютному ринку, що почався в понеділок, розходиться з динамікою фондового ринку. Американські акції в понеділок продовжили падати, а Dow Jones Industrial Average завершив торги в зоні "ведмежого" тренду – впав на 20% і більше з останнього піку вперше від початку пандемії коронавірусу [10].

На рисунку 2.22 відображена аномалія 29.09.2021 р. різкої зміни курсу криптовалюти.



Рисунок 2.22 – Аномалія 29.09.2021 р. різкої зміни курсу

П'ята аномалія виявлена 09.11.2021 р. Bitcoin обрушився майже на 20%. Вперше з кінця вересня він опустився нижче за 19 тисяч доларів. Про це повідомляла РБК-Україна з посиланням на показники криптобіржі Binance [10].

Біткойн впав до 18 400 доларів на Binance, найбільшій біржі криптовалют за обсягом, опустившись нижче 19 000 доларів за монету вперше з кінця вересня. Біткойн падає в ціні протягом трьох торгових сесій поспіль. Сьогодні вранці ціна впала нижче 20 000 доларів за монету вперше з 25 жовтня. Біткойн падає в ціні протягом трьох торгових сесій поспіль. Вона впала нижче 20 000 доларів за монету вперше з 25 жовтня. З початку року курс біткоіна впав у 2,3 рази, відступивши від рівня 46,2 тисячі доларів. Історичного максимуму в 69 000 доларів США було досягнуто в листопаді минулого року, а поточна ціна не перевищувала 48 000 доларів з моменту створення. Тобто в межах відносної

похибки. Binance, найбільша у світі криптовалютна біржа, підписала лист про наміри придбати FTX.com, оператора свого головного конкурента, біржі FTX. Тоді весь криптовалютний ринок обвалився, коли біржа FTX виявилася банкрутом. Ціна біткойна захоплює весь ринок [10].

Ще однією причиною падіння курсу біткойна стало розслідування Комісії з цінних паперів і бірж США (SEC) щодо компанії Marathon Digital Holdings, яка заснувала дата-центр, фінансований за рахунок емісії акцій. Американська влада розцінила це як порушення закону. Давно відомо, що SEC викликала Марафона до суду. Комісія почала розслідування щодо компанії за порушення федерального законодавства про цінні папери. Розслідування стосується партнерства з дата-центром у Монтані, проведеного спільно з Beowulf Energy у жовтні 2020 року. Два роки тому Marathon і Beowulf Energy підписали свій перший контракт на розробку рішень для видобутку з низьким енергоспоживанням. В обмін на перенесення центру майнінгу біткойнів поруч із Beowulf Power Station Beowulf Power Station отримає акції Marathon, повідомляє портал Cryptonews.

На рисунку 2.23 відображена аномалія 09.11.2021 р. різкої зміни курсу криптовалюти.



Рисунок 2.23 – Аномалія 09.11.2021 р. різкої зміни курсу криптовалюти.

Після виявлення аномальних даних формуємо із них датасет На рисунку 2.24 виведені всі аномальні зміни у один датасет за весь 2021 рік які будуть використовуватись під час нашого прогнозування по біткоїну. У рядку ds відображаються саме ті дати які стали аномальними.

	ds	lower_window	upper_window	prior_scale	holiday
0	2021-01-08	0	0	10	anomalous_dates
1	2021-01-27	0	0	10	anomalous_dates
2	2021-04-13	0	0	10	anomalous_dates
3	2021-07-20	0	0	10	anomalous_dates
4	2021-09-06	0	0	10	anomalous_dates
5	2021-09-29	0	0	10	anomalous_dates
6	2021-11-08	0	0 <td 10	anomalous_dates	
7	2021-12-17	0	0	10	anomalous_dates

Рисунок 2.24 – Параметри аномальних змін та їх дати

На рисунку 2.25 відображений графік курсу біткоїну синьою лінією та аномальні дати з рисунку 2.24.що, відображаються червоними лініями

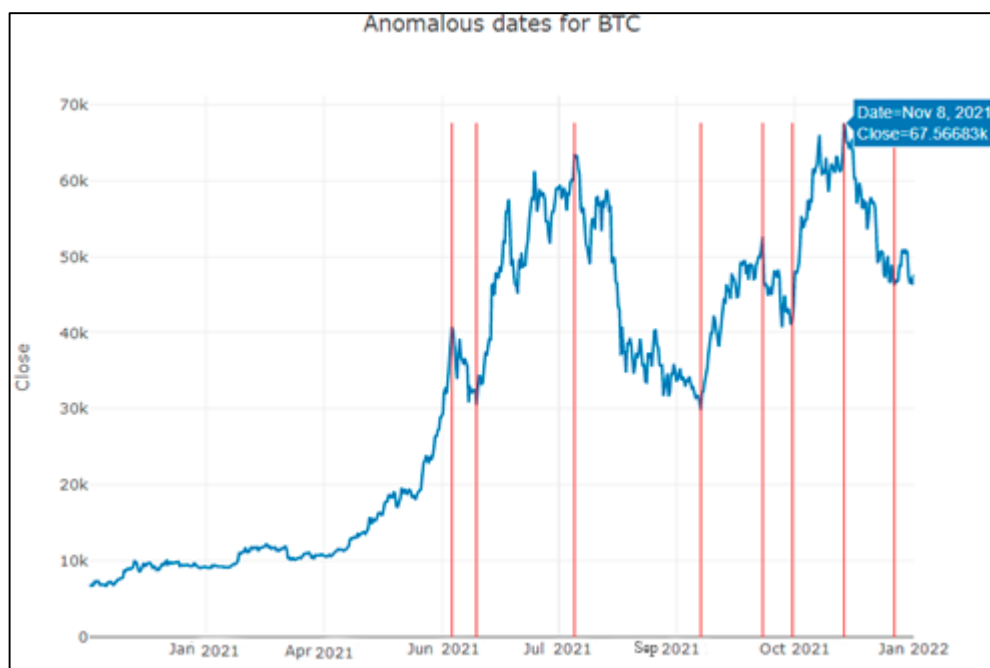


Рисунок 2.25 – Курс криптовалюти та їхні дати аномальних змін

2.4 Побудова моделей методами машинного навчання

Побудовані тренувальний, валідаційний та тестовий датасети з усіма ознаками курсу криптовалюти Bitcoin за даними 2021 року наведено на рисунку 2.25.

```
Origin dataset has 249 rows and 60 features
Get training dataset with 229 rows
Get validation dataset with 10 rows
Get test dataset with 10 rows
```

Рисунок 2.25 – Розмір побудованих тренувального, валідаційного та тестового датасетів з усіма ознаками курсу криптовалюти Bitcoin за даними 2021 року

Відобразимо ряд моделей з використанням методів машинного навчання. Використаємо ознаки. Здійснимо їх стандартизацію за допомоги метода «sklearn.preprocessing.StandardScaler».

Ідентифікуємо такі моделі (рис. 2.26):

	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 2.26 – Моделі машинного навчання та їх параметри

Лінійна регресія є привабливою моделлю, оскільки представлення настільки просте.

Представлення — це лінійне рівняння, яке поєднує певний набір вхідних значень (x), розв'язок якого є прогнозованим результатом для цього набору вхідних значень (y). Таким чином, як вхідні значення (x), так і вихідне значення є числовими [13].

Лінійне рівняння призначає один масштабний коефіцієнт для кожного вхідного значення або стовпця, який називається коефіцієнтом і позначається великою грецькою літерою Beta (B). Також додається один додатковий коефіцієнт, що надає лінії додатковий ступінь свободи (наприклад, переміщення вгору і вниз на двовимірному графіку), і його часто називають перехопленням або коефіцієнтом зміщення [13].

Регресія KNN (KNeighbors Regressor)— це непараметричний метод. Розмір околиці має бути встановлений аналітиком або може бути обраний за допомогою перехресної перевірки (ми побачимо це пізніше), щоб вибрати розмір, який мінімізує середньоквадратичну помилку.

Регресія вектора підтримки Epsilon. Вільними параметрами в моделі є C та $epsilon$ [14].

Реалізація заснована на `libsvm`. Складність часу підгонки більш ніж квадратична з кількістю вибірок, що ускладнює масштабування до наборів даних з більш ніж парою 10000 вибірок. Для великих наборів даних подумайте про використання `LinearSVR` або `SGDRegressor` замість цього, можливо, після трансформатора `Nystroem`.

Random Forest Regressor.

Випадковий ліс — це мета-оцінка, яка відповідає ряду класифікаційних дерев рішень на різних підвбірках набору даних і використовує усереднення для підвищення точності прогнозування та контролю над-підгонки. Розмір підвбірки контролюється за допомогою параметра `max_samples`, якщо `bootstrap=True` (за замовчуванням), інакше для побудови кожного дерева використовується весь набір даних [15].

Регресор мішування.

Regressor Bagging — це складний метаоцінювач, який пристосовує кожен базовий регресор до випадкових підмножин вихідного набору даних, а потім об'єднує їхні індивідуальні передбачення (чи шляхом голосування, чи шляхом усереднення), щоб сформувати остаточний прогноз. Така мета-оцінка зазвичай може бути використана як спосіб зменшити дисперсію оцінювача чорної скриньки (наприклад, дерева рішень) шляхом введення рандомізації в процедуру його побудови, а потім побудови ансамблю з нього [16].

Підвищення градієнта (XGB) відноситься до класу ансамблевих алгоритмів машинного навчання, які можна використовувати для задач класифікації або прогнозного моделювання регресії.

Ансамблі будуються з моделей дерева рішень. Дерева додаються по одному в ансамбль і підходять для виправлення помилок передбачення, допущених попередніми моделями. Це тип ансамблевої моделі машинного навчання, яку називають підвищенням [16].

Багатошарові перцептрони (MLP) — це штучна нейронна мережа, яка має 3 або більше шарів перцептронів. Ці шари — це один вхідний шар, 1 або більше прихованих шарів і один вихідний шар перцептронів. Дані надходять в одному напрямку, тобто вперед, з вхідних шарів → прихованих шарів (-ів) → вихідного рівня. Зворотне поширення — це техніка, при якій багатошаровий перцептрон отримує зворотний зв'язок щодо помилки в своїх результатах, а MLP відповідно коригує свої ваги, щоб робити більш точні прогнози в майбутньому. MLP використовується в багатьох техніках машинного навчання, таких як класифікація та регресія. Було показано, що вони дають дуже точні результати, зокрема, для задач класифікації [17].

За допомоги методу GridSearchCV були обрані найбільш оптимальні параметри моделей (рис. 2.28 – 2.29).

```

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

Tuning model 'KNeighbors Regressor'
Best parameters: {'leaf_size': 10, 'n_neighbors': 20}

Tuning model 'Support Vector Machines'
Best parameters: {'C': 1.0, 'kernel': 'linear', 'tol': 0.001}

Tuning model 'Linear SVR'
Best parameters: {'C': 2.0}

```

Рисунок 2.28 – Оптимальні параметри ідентифікованих моделей, побудованих для багатьох ознак для валідаційних даних курсу криптовалюти

```

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

Tuning model 'Bagging Regressor'
Best parameters: {'max_features': 0.05, 'n_estimators': 5, 'warm_start': False}

Tuning model 'XGB Regressor'
Best parameters: {'learning_rate': 0.2, 'max_depth': 3, 'n_estimators': 90}

Tuning model 'MLP Regressor'
Best parameters: {'hidden_layer_sizes': 2, 'learning_rate': 'adaptive', 'learning_rate_init': 0.01, 'max_iter': 1000, 'solver': 'lbfgs'}

CPU times: user 6min 4s, sys: 3.56 s, total: 6min 8s
Wall time: 4min 14s

```

Рисунок 2.29 – Оптимальні параметри ідентифікованих моделей, побудованих для багатьох ознак для валідаційних даних курсу криптовалюти (продовження)

Після обчислення усіх моделей на валідаційних даних здійснюється порівняння прогнозів з відомими цільовими значеннями «Close» у валідаційному датасеті за 3-ма метриками: r^2_score , RMSE, MSE.

Регресія KNN (2 K-nearest Neighbours Regression) — це непараметричний метод, який інтуїтивно наближає зв'язок між незалежними змінними та безперервним результатом шляхом усереднення спостережень у тому самому околиці. Розмір околиці має бути встановлений аналітиком або може бути обраний за допомогою перехресної перевірки (ми побачимо це пізніше), щоб вибрати розмір, який мінімізує середньоквадратичну помилку [18].

Хоча метод досить привабливий, він швидко стає непрактичним, коли розмірність збільшується, тобто коли є багато незалежних змінних.

Похибки усіх моделей, побудованих для багатьох ознак для валідаційних даних курсу криптовалюти наведено на рисунку 2.30.

name_model	type_data	r2_score	rmse	mape
MLP Regressor	valid	-0.198116	1351.685712	2.446971
Random Forest Regressor	valid	-1.346753	1891.733477	3.600522
XGB Regressor	valid	-1.681957	2022.330814	3.910245
ARIMA_manual	valid	-2.134046	2186.144947	4.155419
ARIMA_auto	valid	-2.134046	2186.144947	4.155419
Prophet_4_days_3_order	valid	-3.500235	2619.653422	4.378228
Prophet_7_days_3_order	valid	-3.484681	2615.122624	4.434402
Prophet_5_days_3_order	valid	-3.451903	2605.548101	4.459947
Prophet_14_days_12_order	valid	-3.980607	2755.924964	4.67282
KNeighbors Regressor	valid	-2.878011	2431.816508	4.742777
Prophet_7_days_12_order	valid	-4.348056	2855.776728	4.823952
Prophet_14_days_3_order	valid	-4.210608	2818.840293	4.854828
Prophet_5_days_12_order	valid	-4.572239	2915.017259	4.903258
Prophet_4_days_12_order	valid	-4.853504	2987.680986	4.904383
Linear SVR	valid	-3.263582	2549.843676	4.947899
Support Vector Machines	valid	-3.347282	2574.750635	5.000937
Linear Regression	valid	-7.766195	3656.215868	7.098678
Bagging Regressor	valid	-11.517699	4369.06509	7.56984

Рисунок 2.30 – Похибки усіх моделей, побудованих для багатьох ознак для валідаційних даних курсу криптовалюти

Як видно, похибка оптимальної моделі за усіма трьома метриками «MLP Regressor» має низьку відносну похибку, порівняно малу похибку RMSE і дуже погану точність за метрикою r^2_score , що означає, що самі значення прогнозуються достатньо точно, але напрямок зростання чи падіння курсу прогнозується дуже погано (майже у протифазі, оскільки коефіцієнт кореляції r^2_score є від'ємним). Отже, оптимальна модель є гарною тільки за метрикою MAPE, тобто за відотною похибкою.

2.5 Висновки

У другому розділі було проведено розвідувальний аналіз даних, синтез нових ознак та перевірка різних моделей для обрання оптимальної для задачі прогнозування курсу криптовалюти.

Розвідувальний аналіз даних показав графіки аномальних дат та зміни курсу, відносно яких було проаналізовано причини, що викликали такі різкі зміни.

Внаслідок синтезу нових ознак було відібрано 57 ознак, з яких 5 основних та 50 згенерованих бібліотекою TSFRESH.

Було побудовано та проаналізовано моделі машинного навчання для обраних ознак.

На основі валідаційних даних курсу криптовалюти було проведено аналіз похибки різних моделей аналізу та обрано оптимальну модель: MLP Regressor. Як видно, похибка оптимальної моделі за усіма трьома метриками «MLP Regressor» є гарною тільки за метрикою MAPE, тобто за відносною похибкою.

У разі прогнозування даних на розширеному датасеті була отримана похибка у розмірі 2.44691, що свідчить про те, що модель добре навчена і дає досить точну оцінку майбутнім змінам курсу.

3 РОЗРОБЛЕННЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ТА ПРОГНОЗУВАННЯ ЗА НЕЮ КУРСУ BITCOIN

3.1 Розроблення інформаційної технології прогнозування курсу Bitcoin

Для проведення прогнозування було пророблено такі попередні етапи:

1) Формалізація постановки задачі.

Для досягнення поставленої мети необхідно розв'язати такі задачі:

- Провести аналіз закономірностей зміни курсу криптовалюти Bitcoin;
- Здійснити вибір оптимальних інформаційних технологій та моделей машинного навчання;
- Побудувати моделі зміни курсу та вибрати з них оптимальну;
- Здійснити прогнозування курсу криптовалюти на тестових даних та оцінити ефективність роботи оптимальної моделі.

2) Вибір оптимальних інформаційних технологій.

3) Аналіз аномальних змін курсу криптовалюти та причин їх виникнення.

Було виявлено п'ять аномальних змін та описано їх розділі 2.3.

4) Створення Pandas Profiling Report для кращого розуміння даних курсу та їх змін з часом.

Pandas profiling – потужний інструмент data science, що дозволяє значно прискорити попередній аналіз даних. Його використання особливо ефективно при роботі з великими датасетами, коли досліджувати кожну ознаку вручну неможливо [21].

Для розв'язання поставленої задачі для тестових даних пропонується такий алгоритм:

1. Проаналізувати дані.

За ознаки OHLC будуємо так звану OHLC-діаграму яка допомагає нам побачити падіння та підйом курсу. Червоним кольором відображається падіння курсу а зеленим період підйому

На рисунку 3.1 зображена OHLC-діаграма криптовалюти Bitcoin за даними сайту Yahoo.

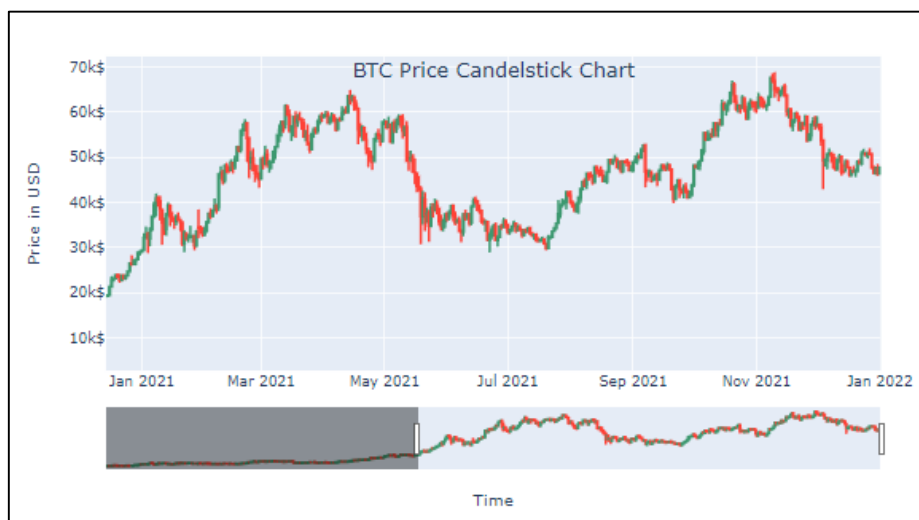


Рисунок 3.1 – OHLC-діаграма криптовалюти Bitcoin за даними сайту Yahoo

2. Синтез ознак. Для тестових даних синтезу на рисунку 3.2 – 3.3 такі ознаки.

```
[ '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',
  'index_sum_values',
  'index_abs_energy',
  'index_median',
  'index_mean',
```

Рисунок 3.2 – Перелік синтезу ознак

```
'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',
'Date']
```

Рисунок 3.3 – Перелік синтезу ознак (продовження)

3. Використаємо ознаки. Здійснимо їх стандартизацію методом «sklearn.preprocessing.StandardScaler».

4. Формування формуванню тренувальних датасетів, валідаційних датасетів для відбору оптимальної моделі та тестових для надання оцінки точності прогнозу.

5. Вибір і тренування моделей на тренувальному датасеті (train). Прогнозування за кожною ідентифікованою моделлю валідаційних даних (valid).

6. Вибір оптимальної моделі серед всіх наведених.

7. Тренування найбільш оптимальної моделі за своїм датасетом (train+valid). Прогнозування за ідентифікованою моделлю тестових даних (test).

8. Надання оцінки точності прогнозу тестових даних.

3.2 Прогнозування тестових даних

Вибрана у розділі 2 оптимальна модель була заново ідентифікована на вже на розширеному датасеті «train+valid», після чого було зроблено прогнозування вже тестових даних та порівняння з реальними даними (рис. 3.2 – 3.3).

Багатошаровий перцептрон (MLP) – це алгоритм навчання з учителем, який вивчає функцію $f(*): R^m \rightarrow R^0$, навчанням на наборі даних, де m – кількість вимірювань для введення та o – кількість розмірів для виведення. Враховуючи набір функцій і ціль, він може вивчити апроксиматор нелінійної функції для класифікації чи регресії. Він відрізняється від логістичної регресії тим, що між вхідним та вихідним шарами може бути один або кілька нелінійних шарів, які називаються прихованими шарами [20].

Переваги багатошарового перцептронну:

– Можливість вивчати нелінійні моделі;

Можливість вивчення моделей у режимі реального часу (онлайн-навчання) із використанням `partial_fit`.

До недоліків багатошарового перцептронну (MLP) можна віднести:

– MLP із прихованими шарами мають невиконувати функцію втрат, коли існує більше одного локального мінімуму. Тому різні ініціалізації випадкових терезів можуть призвести до різної точності перевірки;

– MLP вимагає налаштування ряду гіперпараметрів, таких як кількість прихованих нейронів, шарів та ітерацій;

– MLP є чутливим до масштабування функцій.

У MLP процес обчислення виважених сум повторюється кілька разів. Відбувається обчислення виваженої суми входів для кожного елемента прихованого шару, до неї застосовується функція активації – зазвичай використовуються нелінійні функції випрямлений лінійний елемент або гіперболічний тангенс [20].

За промовчанням MLP використовує 100 прихованих вузлів.

Можлива налаштувати складність нейронної мережі за допомогою `l2` штрафу, щоб стиснути вагові коефіцієнти до близьких до нуля значень, аналогічно як це робиться в гребеневій регресії та лінійних класифікаторах. У `MLPClassifier` це відповідає параметр `alpha` (як і в моделях лінійної регресії). За промовчанням встановлено дуже низьке значення (невелика регулювання) [19].

Ваги в нейронній мережі задаються випадково перед початком навчання і випадкова ініціалізація впливає на процес навчання моделі.

Найбільш важливими параметрами є кількість шарів та кількість прихованих блоків в одному шарі. Доцільно розпочати з одного або двох прихованих шарів. Кількість вузлів на прихованому рівні часто аналогічна числу вхідних функцій.

Алгоритм, який використовується для обчислення ваги, задається за допомогою параметра `solver`. Існує два простих у використанні алгоритму. Алгоритм `adam`, виставлений за умовчанням, дає хорошу якість у більшості ситуацій, але дуже чутливий до масштабування даних (тому важливо від масштабувати дані так, щоб кожна характеристика мала середнє 0 і дисперсію 1) [19].

Клас `MLPRegressor` реалізує багатошаровий перцептрон (MLP), який навчається з використанням зворотного розповсюдження без функції активації у вихідному шарі, що також можна розглядати як використання функції ідентифікації як функції активації. Отже, він використовує квадратну помилку як функцію втрат, але в виході є набір безперервних значень [20].

`MLPRegressor` також підтримує регресію з кількома виходами, коли у вибірці може бути більше однієї мети [20].

На рисунку 3.5 зображені оптимальні параметри моделі `MLP Regressor`, оптимальної за метрикою `MAPE`

	name_model	r2_score	rmse	mape	params
17	MLP Regressor	-0.198116	1351.685712	2.446971	{'hidden_layer_sizes': 2, 'learning_rate': 'ad...

```

Optimal model is 'MLP Regressor' with parameters {'hidden_layer_sizes': 2, 'learning_rate':
'adaptive', 'learning_rate_init': 0.01, 'max_iter': 1000, 'solver': 'lbfgs'}
Tuning model 'MLP Regressor'

```

Рисунок 3.5 – Похибки і оптимальні параметри моделі `MLP Regressor`, оптимальної за метрикою `MAPE`

Після знаходження оптимальних параметрів моделі відображаєм реальні тестові дані курсу Bitcoin за останні 10 днів 2021 року та їхній прогноз за моделлю MLP Regressor (рис. 3.6).

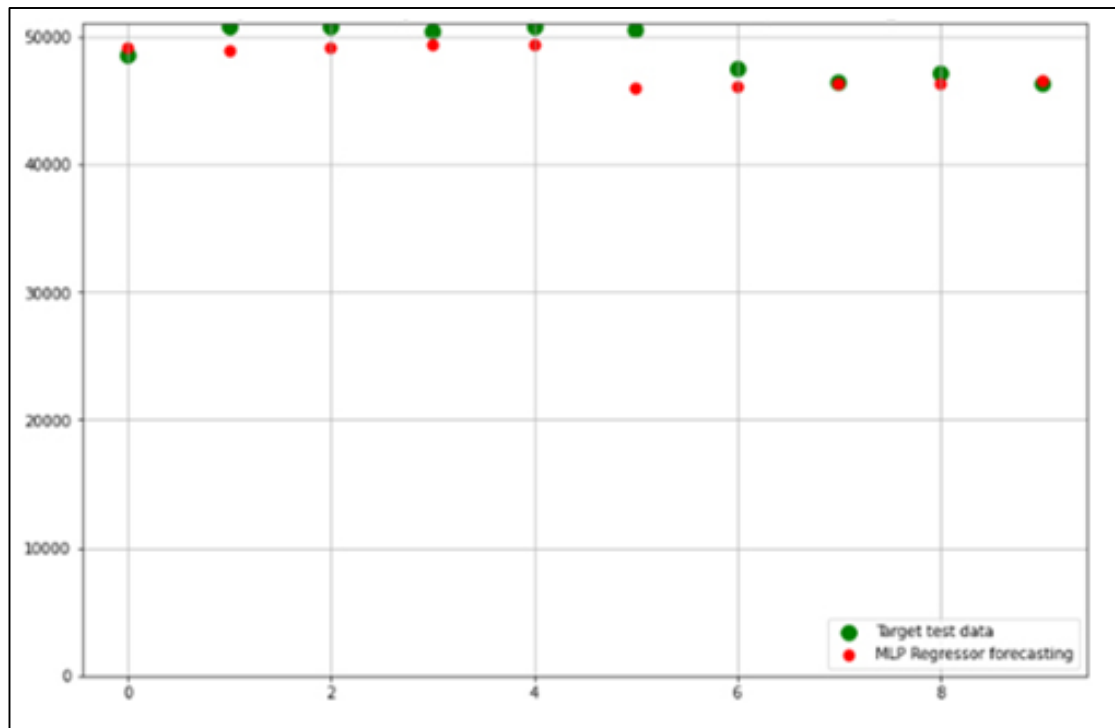


Рисунок 3.6 – Реальні тестові дані курсу криптовалюти та їх прогноз за моделлю MLP Regressor, оптимальної за метрикою MAPE

Як видно з графіків Реальні тестові дані курсу криптовалюти та їх прогноз за моделлю MLP Regressor, оптимальної за метрикою MAPE видають досить точний результат.

3.3 Висновки

У третьому розділі було розроблено інформаційну технологію прогнозу курсу криптовалюти біткоїну на основі моделі MLP Regressor за метрикою MAPE.

Було відображаено реальні тестові дані курсу Bitcoin за останні 10 днів 2021 року та їхній прогноз за моделлю MLP Regressor Як видно з графіків Реальні

тестові дані курсу криптовалюти та їх прогноз за моделлю MLP Regressor, оптимальної за метрикою MAPE видають досить точний результат.

4 ЕКОНОМІЧНА ЧАСТИНА

4.1 Комерційний та технологічний аудит науково-технічної розробки

Метою даного розділу є проведення технологічного аудиту, в даному випадку нової інформаційної технології прогнозування курсу криптовалюти Bitcoin у 2021-2022 роках. Особливістю розробки є підвищення точності прогнозування курсу цифрових валют за рахунок використання методів машинного навчання.

За аналоги були взяті ІТ, в основі яких лежить аналіз ринку та курсу криптовалюти, побудова графіків та їхніх діаграм. Отже, аналогом може бути TradingView, Finviz орієнтовна вартість яких для комерційного сегменту складає близько 120 000 грн.

Для проведення комерційного та технологічного аудиту залучають не менше 3-х незалежних експертів. Оцінювання науково-технічного рівня розробки та її комерційного потенціалу рекомендується здійснювати із застосуванням п'ятибальної системи оцінювання за 12-ма критеріями, у відповідності із табл. 4.1. [23].

Таблиця 4.1 – Рекомендовані критерії оцінювання комерційного потенціалу розробки та їх можлива бальна оцінка

Критерії оцінювання та бали (за 5-ти бальною шкалою)					
Кри-терій	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 умовах

Продовження табл. 4.1

Ринкові переваги					
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
4	Технічні та споживчі властивості продукту значно гірші, ніж в аналогів	Технічні та споживчі властивості продукту трохи гірші, ніж в аналогів	Технічні та споживчі властивості продукту на рівні аналогів	Технічні та споживчі властивості продукту трохи кращі, ніж в аналогів	Технічні та споживчі властивості продукту значно кращі, ніж в аналогів
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 динаміки	Ринок малий, але має позитивну динаміку	Середній ринок з позитивною динамікою	Великий стабільний ринок	Великий ринок з позитивною динамікою
7	Активна конкуренція великих компаній на ринку	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 фахівців	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 виробництві

Продовження табл. 4.1

Практична здійсненність					
11	Термін реалізації ідеї більший за 10 років	Термін реалізації ідеї більший за 5 років. Термін окупності інвестицій більше 10-ти років	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 років	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 продукту	Необхідно отримання великої кількості дозвільних документів на виробництво та реалізацію продукту, що вимагає значних коштів та часу	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

Усі дані по кожному параметру занесено в таблиці 4.2

Таблиця 4.2 – Результати оцінювання комерційного потенціалу розробки

Критерії оцінювання	ПІБ експертів		
	Богачук А. В.	Жук В. Л.	Гусак С. В.
	Бали		
Технічна здійсненність концепції	3	4	4
Наявність аналогів на ринку	3	3	4
Цінова політика	4	4	4
Технічні та споживчі властивості виробу	4	3	4
Експлуатаційні витрати	4	4	3
Ринок збуту	4	3	4
Конкурентоспроможність	3	4	3
Фахівці з технічної і комерційної реалізації	4	3	4
Фінансування	4	4	3
Матеріально-технічна база	3	3	3
Термін реалізації ідеї	4	4	4
Супровідна документація	4	3	4
Сума	42	41	43
Середньоарифметична сума балів	$(42+41+43) / 3 = 42$		

За даними таблиці 4.2 можна зробити висновок щодо рівня комерційного потенціалу даної розробки. Для цього доцільно скористатись рекомендаціями, наведеними в таблиці 4.3.

Таблиця 4.3 – Рівні комерційного потенціалу розробки

Середньоарифметична сума балів, розрахована на основі висновків експертів	Рівень комерційного потенціалу розробки
0-10	Низький
11-20	Нижче середнього
21-30	Середній
31-40	Вище середнього
41-48	Високий

Як видно з таблиці, рівень комерційного потенціалу розроблюваного нового програмного продукту є високим, що досягається за рахунок того, що оцінка вартості будинку – послуга, без якої не обійтись у багатьох випадках. Фактично будь-які операції з нерухомим майном вимагають розрахунку його ринкової вартості. Високий рівень комерційного потенціалу розроблюваного нового програмного продукту досягається за рахунок того, що використовуються методи машинного навчання, які дозволяють підвищити точність такого передбачення у порівнянні з аналогами.

4.2 Прогнозування витрат на виконання науково-дослідної (дослідно-конструкторської) роботи

Основна заробітна плата розробників, яка розраховується за формулою:

$$Z_o = \frac{M}{T_p} \cdot t, \quad (4.1)$$

де M – місячний посадовий оклад конкретного розробника (дослідника), грн.;

T_p – число робочих днів за місяць, 21 днів;

t – число днів роботи розробника (дослідника).

Результати розрахунків зведемо до таблиці 4.4.

Таблиця 4.4 – Основна заробітна плата розробників

Найменування посади	Місячний посадовий оклад, грн.	Оплата за робочий день, грн.	Число днів роботи	Витрати на заробітну плату, грн.
Керівник проекту	48000	2285,71	40	91428,571
Програміст	45000	2142,00	40	85714,286
Всього				177142,86

Так як в даному випадку розробляється програмний продукт, то розробник виступає одночасно і основним робітником, і тестувальником розроблюваного програмного продукту.

Додаткова заробітна плата розробників, які брали участь в розробці програмного продукту.

Додаткову заробітну плату прийнято розраховувати як 10 % від основної заробітної плати розробників та робітників:

$$Z_d = Z_o \cdot 10 \% / 100 \% \quad (4.2)$$

$$Z_d = (177142,86 \cdot 10 \% / 100 \%) = 17714,286 \text{ (грн).}$$

Нарахування на заробітну плату розробників. Згідно діючого законодавства нарахування на заробітну плату складають 22 % від суми основної та додаткової заробітної плати.

$$H_z = (Z_o + Z_d) \cdot 22 \% / 100\% \quad (4.3)$$

$$H_z = (177142,86 + 17714,286) \cdot 22 \% / 100 \% = 43258,29 \text{ (грн).}$$

Оскільки для розроблювального пристрою не потрібно витратити матеріали та комплектуючі, то витрати на матеріали і комплектуючі дорівнюють нулю.

Амортизація обладнання, яке використовувалось для проведення розробки.

Амортизація обладнання, що використовувалось для розробки в спрощеному вигляді розраховується за формулою:

$$A = \frac{Ц}{T_B} \cdot \frac{t_{\text{вик}}}{12} [\text{грн}]. \quad (4.4)$$

де Ц – балансова вартість обладнання, грн.;

T – термін корисного використання обладнання згідно податкового законодавства, років;

$t_{\text{вик}}$ – термін використання під час розробки, місяців.

Розрахуємо, для прикладу, амортизаційні витрати на комп'ютер балансова вартість якого становить 20000 грн., термін його корисного використання згідно податкового законодавства – 2 роки, а термін його фактичного використання – 1,90 міс.

$$A_{\text{обл}} = \frac{22000}{2} \times \frac{1,9}{12} = 1746,032 \text{ (грн)}.$$

Аналогічно визначаємо амортизаційні витрати на інше обладнання та приміщення. Розрахунки заносимо до таблиці 4.5.

Так як вартість ліцензійної ОС та спеціалізованих ліцензійних нематеріальних активів є безкоштовною, то $V_{\text{нем.ак}} = 0$ грн.

Таблиця 4.5 – Амортизаційні відрахування на матеріальні та нематеріальні ресурси ДЛЯ розробників

Найменування обладнання	Балансова вартість, грн.	Строк корисного використання, років	Термін використання обладнання, місяців	Амортизаційні відрахування, грн.
Комп'ютер та комп'ютерна периферія (Acer Aspire 5)	22000	2	1,90	1746,032
Офісне обладнання (меблі)	20000	4	1,90	793,651
Приміщення	850000	20	1,90	6746,032
Всього				9285,71

Тарифи на електроенергію для побутових споживачів (промислових підприємств) відрізняються від тарифів на електроенергію для населення. При цьому тарифи на розподіл електроенергії у різних постачальників (енергорозподільних компаній), будуть різними. Крім того, розмір тарифу залежить від класу напруги (1-й або 2-й клас). Тарифи на розподіл електроенергії для всіх енергорозподільних компаній встановлює Національна комісія з регулювання енергетики і комунальних послуг (НКРЕКП). Витрати на силову електроенергію розраховуються за формулою:

$$V_e = V \cdot P \cdot \Phi \cdot K_{\Pi}, \quad (4.5)$$

де V – вартість 1 кВт-години електроенергії для 1 класу підприємства, $V = 6,2$ грн./кВт;

P – встановлена потужність обладнання, кВт. $P = 0,4$ кВт;

Φ – фактична кількість годин роботи обладнання, годин.

K_{Π} – коефіцієнт використання потужності, $K_{\Pi} = 0,9$.

$$V_e = 0,9 \cdot 0,4 \cdot 8 \cdot 40 \cdot 6,2 = 803,52 \text{ (грн).}$$

Інші витрати та загальнопромислові 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» розраховуються як 50...100% від суми основної заробітної плати дослідників:

$$I_B = (Z_o + Z_p) \cdot \frac{H_{iB}}{100\%}, \quad (4.6)$$

де H_{iB} – норма нарахування за статтею «Інші витрати».

$$I_B = 177142,86 * 60\% / 100\% = 106285,7 \text{ (грн)}.$$

До статті «Накладні (загальнопромислові) 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 (загальнопромислові) витрати» розраховуються як 100...150% від суми основної заробітної плати дослідників:

$$H_{HЗВ} = (Z_o + Z_p) \cdot \frac{H_{HЗВ}}{100\%}, \quad (4.7)$$

де $H_{HЗВ}$ – норма нарахування за статтею «Накладні (загальнопромислові) витрати».

$$H_{HЗВ} = 177142,86 * 140\% / 100\% = 248000 \text{ (грн)}.$$

Витрати на проведення науково-дослідної роботи. Сума всіх попередніх статей витрат дає загальні витрати на проведення науково-дослідної роботи:

$$B_{\text{заг}} = 177142,86 + 19485,71 + 43258,29 + 9285,71 + 803,52 + 106285,7 + \\ + 248000 = 604261,81 \text{ (грн)}.$$

Розрахунок загальних витрат на науково-дослідну (науково-технічну) роботу та оформлення її результатів.

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

$$ЗВ = \frac{B_{\text{заг}}}{\eta} \text{ (грн)}, \quad (4.8)$$

де η – коефіцієнт, який характеризує етап (стадію) виконання науково-дослідної роботи.

Так, якщо науково-технічна розробка знаходиться на стадії: науково-дослідних робіт, то $\eta=0,1$; технічного проектування, то $\eta=0,2$; розробки конструкторської документації, то $\eta=0,3$; розробки технологій, то $\eta=0,4$; розробки дослідного зразка, то $\eta=0,5$; розробки промислового зразка, то $\eta=0,7$; впровадження, то $\eta=0,9$. Оберемо $\eta = 0,5$, так як розробка, на даний момент, знаходиться на стадії дослідного зразка:

$$ЗВ = 604261,81 / 0,5 = 1208524 \text{ (грн)}.$$

4.3 Розрахунок економічної ефективності науково-технічної розробки за її можливої комерціалізації потенційним інвестором

В ринкових умовах узагальнювальним 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. Саме зростання чистого прибутку забезпечить потенційному інвестору надходження додаткових коштів, дозволить покращити фінансові результати його діяльності, підвищить конкурентоспроможність та

може позитивно вплинути на ухвалення рішення щодо комерціалізації цієї розробки.

Для того, щоб розрахувати можливе зростання чистого прибутку у потенційного інвестора від можливого впровадження науково-технічної розробки необхідно:

- вказати, з якого часу можуть бути впроваджені результати науково-технічної розробки;

- зазначити, протягом скількох років після впровадження цієї науково-технічної розробки очікуються основні позитивні результати для потенційного інвестора (наприклад, протягом 3-х років після її впровадження);

- кількісно оцінити величину існуючого та майбутнього попиту на цю або аналогічні чи подібні науково-технічні розробки та назвати основних суб'єктів (зацікавлених осіб) цього попиту;

- визначити ціну реалізації на ринку науково-технічних розробок з аналогічними чи подібними функціями.

При розрахунку економічної ефективності потрібно обов'язково враховувати зміну вартості грошей у часі, оскільки від вкладення інвестицій до отримання прибутку минає чимало часу. При оцінюванні ефективності інноваційних проектів передбачається розрахунок таких важливих показників:

- абсолютного економічного ефекту (чистого дисконтованого доходу);
- внутрішньої економічної дохідності (внутрішньої норми дохідності);
- терміну окупності (дисконтованого терміну окупності).

Аналізуючи напрямки проведення науково-технічних розробок, розрахунок економічної ефективності науково-технічної розробки за її можливої комерціалізації потенційним інвестором можна об'єднати, враховуючи визначені ситуації з відповідними умовами.

Розробка чи суттєве вдосконалення програмного засобу (програмного забезпечення, програмного продукту) для використання масовим споживачем.

В цьому випадку майбутній економічний ефект буде формуватися на основі таких даних:

$$\Delta\Pi_i = (\pm\Delta\Pi_0 \cdot N + \Pi_0 \cdot \Delta N)_i \cdot \lambda \cdot \rho \cdot \left(1 - \frac{\vartheta}{100}\right), \quad (4.9)$$

де $\pm\Delta\Pi_0$ – зміна вартості програмного продукту (зростання чи зниження) від впровадження результатів науково-технічної розробки в аналізовані періоди часу;

N – кількість споживачів які використовували аналогічний продукт у році до впровадження результатів нової науково-технічної розробки;

Π_0 – основний оціночний показник, який визначає діяльність підприємства у даному році після впровадження результатів наукової розробки, $\Pi_0 = \Pi_6 \pm \Delta\Pi_0$;

Π_6 – вартість програмного продукту у році до впровадження результатів розробки;

ΔN – збільшення кількості споживачів продукту, в аналізовані періоди часу, від покращення його певних характеристик;

λ – коефіцієнт, який враховує сплату податку на додану вартість. Ставка податку на додану вартість дорівнює 20%, а коефіцієнт $\lambda = 0,8333$;

ρ – коефіцієнт, який враховує рентабельність продукту;

ϑ – ставка податку на прибуток, у 2022 році $\vartheta = 18\%$.

Припустимо, що при прогнозованій ціні 20000 грн. за одиницю виробу, термін збільшення прибутку складе 3 роки. Після завершення розробки і її вдосконалення, можна буде підняти її ціну на 2000 грн. Кількість одиниць реалізованої продукції також збільшиться: протягом першого року – на 5000 шт., протягом другого року – на 1000 шт., протягом третього року на 300 шт. До моменту впровадження результатів наукової розробки реалізації продукту не було:

$$\Delta\Pi_1 = (0 \cdot 2000 + (20000 + 2000) \cdot 5000) \cdot 0,8333 \cdot 0,22 \cdot (1 - 0,18) = 15033332,732 \text{ (грн).}$$

$$\Delta\Pi_2 = (0 \cdot 2000 + (20000 + 2000) \cdot (5000 + 1000)) \cdot 0,8333 \cdot 0,22 \cdot (1 - 0,18) = 19843999,206 \text{ (грн).}$$

$$\Delta\Pi_3 = (0*2000 + (20000 + 2000) * (5000+1000+300) * 0,8333 * 0,22) * (1 - 0,18) = 20836199,167 \text{ (грн)}.$$

Отже, комерційний ефект від реалізації результатів розробки за три роки складе 55713531,10 грн.

4.4 Розрахунок ефективності вкладених інвестицій і періоду їх окупності

Розраховуємо приведену вартість збільшення всіх чистих прибутків ПП, що їх може отримати потенційний інвестор від можливого впровадження та комерціалізації науково-технічної розробки:

$$ПП = \sum_1^T \frac{\Delta\Pi_i}{(1+\tau)^t}, \quad (4.10)$$

де $\Delta\Pi_i$ – збільшення чистого прибутку у кожному із років, протягом яких виявляються результати виконаної та впровадженої науково-дослідної (науково-технічної) роботи, грн;

T – період часу, протягом якою виявляються результати впровадженої науково-дослідної (науково-технічної) роботи, роки;

τ – ставка дисконтування, за яку можна взяти щорічний прогнозований рівень інфляції в країні, $\tau = 0,05 \dots 0,15$;

t – період часу (в роках).

Збільшення прибутку ми отримаємо, починаючи з першого року:

$$ПП = (15033332,732 / (1+0,1)^1) + (19843999,206 / (1+0,1)^2) + (20836199,167 / (1+0,1)^3) = 13666666,12 + 16399999,34 + 15654544,83 = 45721210,29 \text{ (грн)}.$$

Далі розраховують величину початкових інвестицій PV , які потенційний інвестор має вкласти для впровадження і комерціалізації науково-технічної розробки. Для цього можна використати формулу:

$$PV = k_{\text{інв}} * ЗВ, \quad (4.11)$$

де $k_{\text{інв}}$ – коефіцієнт, що враховує витрати інвестора на впровадження науково-технічної розробки та її комерціалізацію. Це можуть бути витрати на підготовку приміщень, розробку технологій, навчання персоналу, маркетингові заходи тощо; зазвичай $k_{\text{інв}}=2\dots5$, але може бути і більшим;

$ЗВ$ – загальні витрати на проведення науково-технічної розробки та оформлення її результатів, грн.

$$PV = 2 * 1208524 = 2417047,22 \text{ (грн)}.$$

Тоді абсолютний економічний ефект $E_{\text{абс}}$ або чистий приведений дохід (NPV , *Net Present Value*) для потенційного інвестора від можливого впровадження та комерціалізації науково-технічної розробки становитиме:

$$E_{\text{абс}} = ПП - PV, \quad (4.12)$$

$$E_{\text{абс}} = 45721210,29 - 2417047,22 = 43304163,07 \text{ (грн)}.$$

Оскільки $E_{\text{абс}} > 0$ то вкладання коштів на виконання та впровадження результатів даної науково-дослідної (науково-технічної) роботи може бути доцільним.

Для остаточного прийняття рішення з цього питання необхідно розрахувати внутрішню економічну дохідність або показник внутрішньої норми дохідності (IRR , *Internal Rate of Return*) вкладених інвестицій та порівняти її з так званою бар'єрною ставкою дисконтування, яка визначає ту мінімальну

внутрішню економічну дохідність, нижче якої інвестиції в будь-яку науково-технічну розробку вкладати буде економічно недоцільно.

Розрахуємо відносну (щорічну) ефективність вкладених в наукову розробку інвестицій E_6 . Для цього використаємо формулу:

$$E_6 = \sqrt[T_{ж}]{\left(1 + \frac{E_{abc}}{PV}\right)} - 1, \quad (4.13)$$

де $T_{ж}$ – життєвий цикл наукової розробки, роки.

$$E_6 = \sqrt[3]{(1 + 47914000,41/2574756,02)} - 1 = 1,664$$

Визначимо мінімальну ставку дисконтування, яка у загальному вигляді визначається за формулою:

$$\tau = d + f, \quad (4.14)$$

де d – середньозважена ставка за депозитними операціями в комерційних банках; в 2022 році в Україні $d = (0,09 \dots 0,14)$;

f – показник, що характеризує ризикованість вкладень; зазвичай, величина $f = (0,05 \dots 0,5)$.

$$\tau_{\min} = 0,14 + 0,05 = 0,19.$$

Так як $E_6 > \tau_{\min}$, то інвестор може бути зацікавлений у фінансуванні даної наукової розробки.

Розрахуємо термін окупності вкладених у реалізацію наукового проекту інвестицій за формулою:

$$T_{ок} = \frac{1}{E_6}, \quad (4.15)$$

$$T_{ок} = 1 / 1,697 = 0,60 \text{ (року)}.$$

Оскільки $T_{ок} < 3$ -х років, а саме термін окупності рівний 0,60 року, то фінансування даної наукової розробки є доцільним.

4.5 Висновки

Висновки до розділу: економічна частина даної роботи містить розрахунок витрат на розробку нового програмного продукту, сума яких складає 1208524 гривень. Було спрогнозовано орієнтовану величину витрат по кожній з статей витрат. Також розраховано чистий прибуток, який може отримати виробник від реалізації нового технічного рішення, розраховано період окупності витрат для інвестора та економічний ефект при використанні даної розробки. В результаті аналізу розрахунків можна зробити висновок, що розроблений програмний продукт за ціною дешевший за аналог і є високо конкуренто спроможним. Період окупності складе близько 0,60 роки.

ВИСНОВКИ

Магістерська кваліфікаційна робота на тему: «Інформаційна технологія прогнозування курсу криптовалюти Bitcoin у 2021-2022 роках» присвячена застосуванню методів машинного навчання для прогнозування курсу криптовалюти Bitcoin у 2021-2022 роках.

Розглянуто загальну характеристику об'єкту досліджень, проблеми прогнозування курсу криптовалюти Bitcoin, здійснено формалізацію постановки задачі та вибрано оптимальні інформаційні технології.

Проведено розвідувальний аналіз даних, синтез нових ознак та перевірка різних моделей для обрання оптимальної для задачі прогнозування курсу криптовалюти.

Розвідувальний аналіз даних показав графіки аномальних дат та зміни курсу, відносно яких проаналізовано причини, що викликали такі різкі зміни.

Внаслідок синтезу нових ознак відібрано 57 ознак, з яких 5 основних та 50 згенерованих бібліотекою TSFRESH. Побудовано та проаналізовано моделі машинного навчання для обраних ознак.

У результаті прогнозування даних з'ясовано, що похибка оптимальної моделі за метрикою MAPE становить 2.4%, що свідчить про те, що модель добре навчена і дає досить точну оцінку майбутнім змінам курсу.

Розроблено інформаційну технологію прогнозу курсу криптовалюти Bitcoin на основі моделі MLP Regressor за метрикою MAPE. Сформовано алгоритм роботи технології та отримані реальні тестові дані курсу та їх прогноз.

Економічна частина містить розрахунок витрат на розробку нового програмного продукту, сума яких складає 713158 гривень. Було спрогнозовано орієнтовану величину витрат по кожній з статей витрат. Також розраховано чистий прибуток, який може отримати виробник від реалізації нового технічного рішення, розраховано період окупності витрат для інвестора та економічний ефект при використанні даної розробки. В результаті аналізу розрахунків можна зробити висновок, що розроблений програмний продукт за ціною дешевший за

аналог і є високо конкурентоспроможним. Період окупності складе близько 0,60 роки.

В результаті виконання магістерської кваліфікаційної роботи з використанням методів машинного навчання розроблено інформаційну технологію прогнозування курсу криптовалюти Bitcoin, яка дозволила підвищити точність прогнозування курсу криптовалюти Bitcoin у 2021-2022 роках. Отже, завдання магістерської кваліфікаційної роботи було виконано в повному обсязі.

За результатами даної роботи написані тези доповіді, які були апробовані на Всеукраїнській науково-практичній інтернет-конференції «Молодь в науці: дослідження, проблеми, перспективи» (Вінниця, 2022-2023 рр.) та опубліковані у матеріалах цієї конференції.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Бурденюк А. В. ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПРОГНОЗУВАННЯ КУРСУ КРИПТОВАЛЮТИ BITCOIN У 2021-2022 РОКАХ. *Всеукраїнська науково-практична інтернет-конференція студентів аспірантів та молодих науковців «Молодь в науці: дослідження, проблеми, перспективи (МН-2023)»*. Вінниця, 2022. URL: <https://conferences.vntu.edu.ua/index.php/mn/mn2023/paper/viewFile/16824/14051>
2. Майбутнє цифрових валют – у фокусі уваги світових центральних банків URL: <https://bank.gov.ua/ua/news/all/maybutnye-tsifrovih-valyut--u-fokusi-uvagi-svitovih-tsentralnih-bankiv>
3. Цифрова валюта. URL: http://wikipedia.ua.nina.az/wiki/Цифрова_валюта
4. Криптовалюта: від появи до наших днів URL: <https://interfax.com.ua/news/press-release/600407.html>
5. How to Use Kaggle. URL: <https://www.kaggle.com/docs/notebooks>
6. Путівник мовою програмування Python Криптовалюта URL: <https://pythonguide.rozh2sch.org.ua/>
7. Береза Є.Д. Переваги використання мови Python для глибокого машинного навчання. *Наукові праці Харківського національного університету радіоелектроніки*. Харків, 2021. URL: https://openarchive.nure.ua/bitstream/document/18002/1/RiM_2021_KITS_167-168.pdf
8. Пономаренко В. С., Золотарьова І. О., Бутова Р. К., Плеханова Г. О. Інформаційні системи: навч. посіб. Харків : ХНЕУ, 2011. 176 с.
9. Жураковський Б. Ю., Зенів І. О. Технології прогнозування: навч. посіб. Київ : КПІ ім. Ігоря Сікорського, 2021. 271 с.
10. Predicting the price of Bitcoin using Machine Learning URL: <http://trap.ncirl.ie/2496/1/seanmcnally.pdf>.
11. Чому Python використовується для машинного навчання? URL: <https://medium.com/nuances-of-programming/чому-python-використовується-для-машинного-навчання-bc1a2d8f5963>

12. Машинне навчання. URL: uk.wikipedia.org/wiki/Машинне_навчання
13. Machine Learning, ML. URL: <https://www.it.ua/knowledge-base/technology-innovation/machine-learning>
14. Машинне навчання. URL: uk.wikipedia.org/wiki/Машинне_навчання
15. Random Forest Regression in Python. URL: <https://www.geeksforgeeks.org/random-forest-regression-in-python/>
16. Regression Example with XGBRegressor in Python. URL: www.datatechnotes.com/2019/06/regression-example-with-xgbregressor-in.html
17. LinearRegression. URL: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html
18. Класичне машинне навчання: завдання класифікації, узагальнення, кластеризації даних. URL: <https://evergreens.com.ua/ua/articles/classical-machine-learning.html>
19. Machine Learning, ML. URL: <https://www.it.ua/knowledge-base/technology-innovation/machine-learning>
20. Deep Neural Multilayer Perceptron (MLP) with Scikit-learn. URL: <https://towardsdatascience.com/deep-neural-multilayer-perceptron-mlp-with-scikit-learn-2698e77155e>
21. Random Forest Regression in Python. URL: <https://www.geeksforgeeks.org/random-forest-regression-in-python/>
22. Regression Example with XGBRegressor in Python. URL: www.datatechnotes.com/2019/06/regression-example-with-xgbregressor-in.html
23. Козловський В. О., Лесько О. Й., Кавецький В. В. Методичні вказівки до виконання економічної частини магістерських кваліфікаційних робіт : уклад. Вінниця : ВНТУ, 2021. 42 с.

Додаток А

Міністерство освіти і науки України
Вінницький національний технічний університет
Факультет інтелектуальних інформаційних технологій та автоматизації

ЗАТВЕРДЖУЮ

Завідувач кафедри САІТ

_____ д.т.н., проф. Мокін В. Б.

«19» _____ 09 _____ 2022 р.

ТЕХНІЧНЕ ЗАВДАННЯ

на магістерську кваліфікаційну роботу

«ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПРОГНОЗУВАННЯ КУРСУ
КРИПТОВАЛЮТИ BITCOIN У 2021-2022 РОКАХ»

08-53.МКР.002.02.000.ТЗ

Керівник: к.т.н., доц. каф. САІТ

_____ Козачко О. М.

«19» _____ 09 _____ 2022 р.

Розробив: студент гр. 2ІСТ-21м

_____ Бурденюк А. В.

«19» _____ 09 _____ 2022 р.

Вінниця 2022

1. Підстава для проведення робіт

Підставою для виконання роботи є наказ № 203 по ВНТУ від «14» 09 2022 р., та індивідуальне завдання на МКР, затверджене протоколом № 3 засідання кафедри САІТ від «14» 09 2022 р.

2. Джерела розробки:

- Мокін В. Б., Жуков С. О., Куперштейн Л. М., і Слободянюк О. В., «Інформаційна технологія прогнозування курсу криптовалют на основі комплексної інженерії ознак», *Вісник ВПІ*, вип. 2, с. 81–93, Квіт. 2022.
- Машинне навчання. URL: <https://www.it.ua/knowledge-base/technology-innovation/machine-learning>

3. Мета і призначення роботи:

Метою дослідження є підвищення точності прогнозування курсу криптовалюти Bitcoin за рахунок використання методів машинного навчання.

4. Вихідні дані для проведення робіт:

Набір даних, зібраний з сервісу Yahoo через API, який містить інформацію про ціну та інші ознаки про криптовалюту Bitcoin та датасет Центру Джона Хопкінса по ковіду.

5. Методи дослідження:

- використання моделей машинного навчання;
- інтерполювання даних.

6. Етапи роботи і терміни їх виконання:

1. Аналіз предметної області 20.09 – 30.09
2. Ідентифікація та вибір оптимальної моделі для прогнозування криптовалюти Bitcoin..... 01.10 – 20.10
3. Розроблення інформаційної технології для аналізу та прогнозування курсу криптовалюти Bitcoin 21.10 – 05.11
4. Економічна частина 06.11 – 19.11
5. Оформлення пояснювальної записки. 10.11 – 30.11

7. Очікувані результати та порядок реалізації:

Результати прогнозування курсу криптовалюти Bitcoin на певний період за оптимальною моделлю.

8. Вимоги до розробленої документації

Пояснювальна записка оформлена у відповідності до вимог «Методичних вказівок до виконання та оформлення магістерських кваліфікаційних робіт для студентів спеціальності 126 – «Інформаційні системи та технології» денної форми навчання».

9. Порядок приймання роботи

Публічний захист « 19 » 12 2022 р.
Початок розробки « 20 » 09 2022 р.
Граничні терміни виконання МКР « 30 » 11 2022 р.

Розробив студент групи 2ІСТ-21м _____ Бурденюк А. В.

Додаток Б

Протокол перевірки кваліфікаційної роботи на наявність текстових
запозичень

Назва роботи: «Інформаційна технологія прогнозування курсу криптовалюти Bitcoin у 2021-2022 роках»

Тип роботи: магістерська кваліфікаційна робота

Підрозділ: кафедра САІТ

Науковий керівник: Козачко О. М. к.т.н., доц. каф. САІТ

Показники звіту подібності Unicheck

Оригінальність	91,01 %
Схожість	8,99 %

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 самостійності її автора. 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.

Опис прийнятого рішення:

Робота допускається до захисту

Особа, відповідальна за перевірку



Жуков С. О.

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

Автор роботи



Бурденюк А. В.

Керівник роботи



Козачко О. М.

Додаток В

Лістинг програми

```
if is_other_ML:
    if (len(result) > 0) and (len(models) > 0):

        # Coefficients
        if opt_name_model2=='XGB Regressor':
            print('Feature importance diagram')
            # Coef. of the feature with nonzero importance
            xgb_coeff =
pd.DataFrame.from_dict(model_opt.get_booster().get_score(importance_type='weigh
t'), orient='index').reset_index(drop=False)
            coeff, _, is_score = add_fi_coeff(coeff, 'xgb_fi_coeff', None, xgb_coeff)

        # With the library xgboost
        fig = plt.figure(figsize = (15,15))
        axes = fig.add_subplot(111)
        xgb.plot_importance(model_opt,ax = axes,height = 0.5)
        plt.show()
        plt.close()

    else:
        # With the library sklearn
        try:
            coef_model = model_opt.coef_
            coeff, coeff_new, is_score = add_fi_coeff(coeff, 'lr_fi_score', coef_model)
        except:
            try:
                coef_model = feature_importances_
```

```

        coeff, coeff_new, is_score = add_fi_coeff(coeff, 'model_fi_score',
coef_model)
    except:
        print('The importance of the feature could not be obtained')
        is_score = False

if is_score:
    # Plot drawing
    coeff_non_zero = coeff_new[coeff_new['score']>0]
    plt.figure(figsize=(12, int(len(coeff_non_zero)*0.4)))
    coeff_non_zero = coeff_non_zero.sort_values(by='score', ascending=True)
    plt.barh(coeff_non_zero["feature"], coeff_non_zero["score"])
    plt.title("Feature importance diagram")
    plt.axvline(x=0, color=".5")
    plt.xlabel("Coefficient values")
    plt.subplots_adjust(left=0.3)

# Permutation feature importance diagram
if is_other_ML:
    if (len(result) > 0) and (len(models) > 0):
        try:
            perm_importance = permutation_importance(model_opt, test_mf, ytest_mf)

            # Save permutation feature importance values
            coef_model = perm_importance.importances_mean
            coeff, coeff_new, is_score = add_fi_coeff(coeff, 'perm_fi_score', coef_model)

            print('Permutation feature importance diagram:')
            coeff_non_zero = coeff_new[coeff_new['score'].abs()>1e-4]
            coeff_non_zero = coeff_non_zero.sort_values(by='score', ascending=True)

```



```

plt.figure(figsize=(12, int(len(coeff_non_zero)*0.4)))
plt.barh(coeff_non_zero["feature"], coeff_non_zero["score"])
plt.xlabel("Permutation Importance")
plt.show()

is_perm_importance = True

except: print('Permutation feature importance diagram for this optimal model is
not supported')

# Feature importance diagram with ELI5
if is_other_ML:
    if (len(result) > 0) and (len(models) > 0):
        try:
            print('Feature importance diagram with ELI5:')
            perm = PermutationImportance(model_opt).fit(test_mf,ytest_mf)

            # Save permutation feature importance values
            coef_model = perm.feature_importances_ # Feature importances,
                # computed as mean decrease
                # of the score when a feature
                # is permuted (i.e. becomes noise)
            coeff, _, is_score = add_fi_coeff(coeff, 'eli5_perm_fi_score', coef_model)

            # Display permutation feature importance values with ELI5
            display(eli5.show_weights(perm, feature_names = coeff.feature.tolist()))

            except: print('Feature importance diagram for this optimal model is not
supported in ELI5')

# Display and saving features importance values
if is_other_ML:

```

```
if coeff.isna().sum().sum()==0:
    print('Feature importance values:')
    fi_cols = coeff.columns.tolist()[1:]
    if len(fi_cols) > 0:
        coeff = coeff.sort_values(by=fi_cols, ascending=False)
        display(coeff)

coeff.to_csv(f'feature_importance_for_optimal_model_{opt_name_model2}.csv',
            index=False)
```

ІЛЮСТРАТИВНА ЧАСТИНА**ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПРОГНОЗУВАННЯ КУРСУ
КРИПТОВАЛЮТИ БІТСОІН У 2021-2022 РОКАХ**

Виконав: студент гр. 2ІСТ-21м

_____ Бурденюк А. В.

«01» _____ 12 _____ 2022 р.

Керівник: к.т.н., доц. каф. САІТ

_____ Козачко О. М.

«02» _____ 12 _____ 2022 р.

Нормоконтроль: к.т.н., доцент

_____ Жуков С. О.

«02» _____ 12 _____ 2022 р.

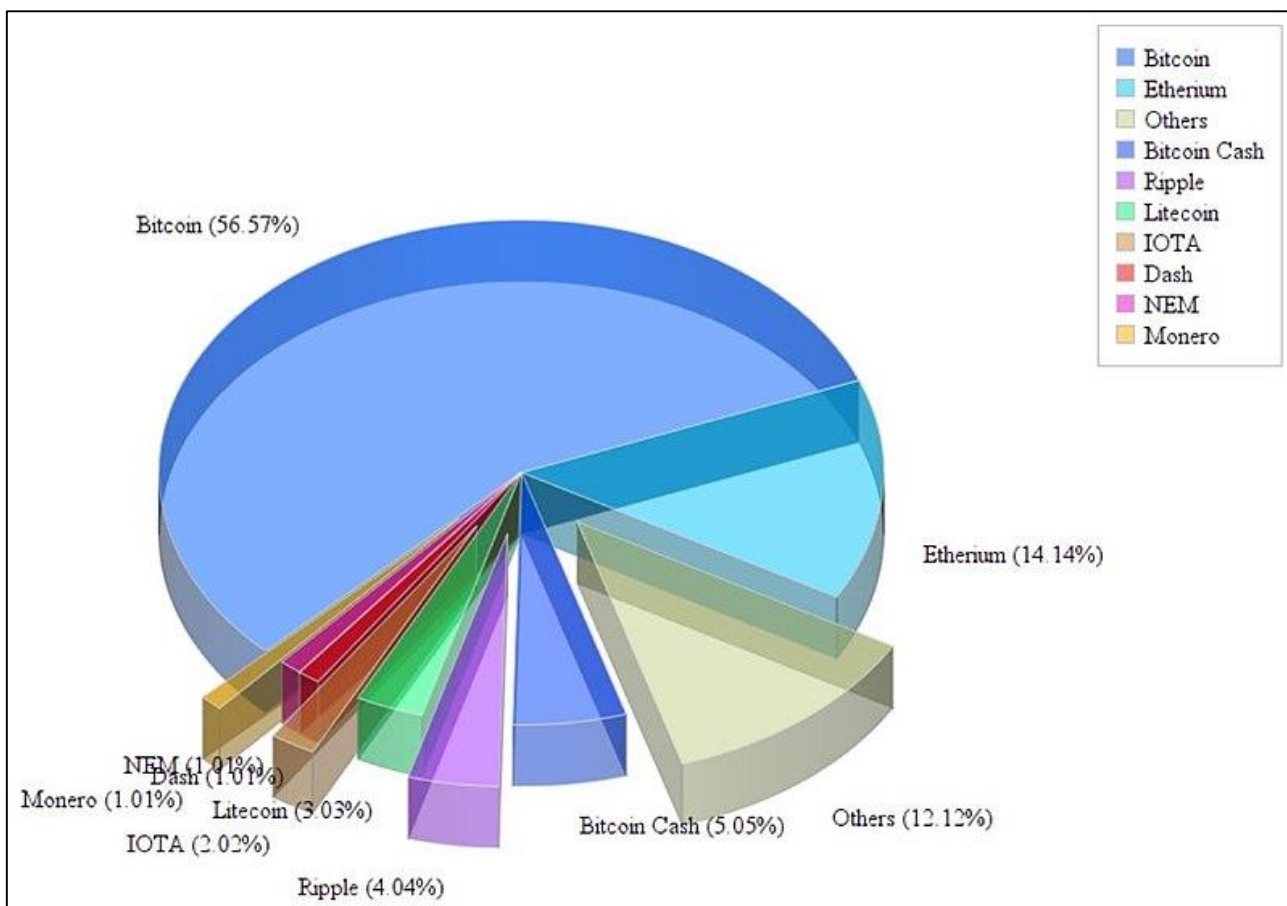


Рисунок Г.1 – Капіталізація ринку криптовалюти

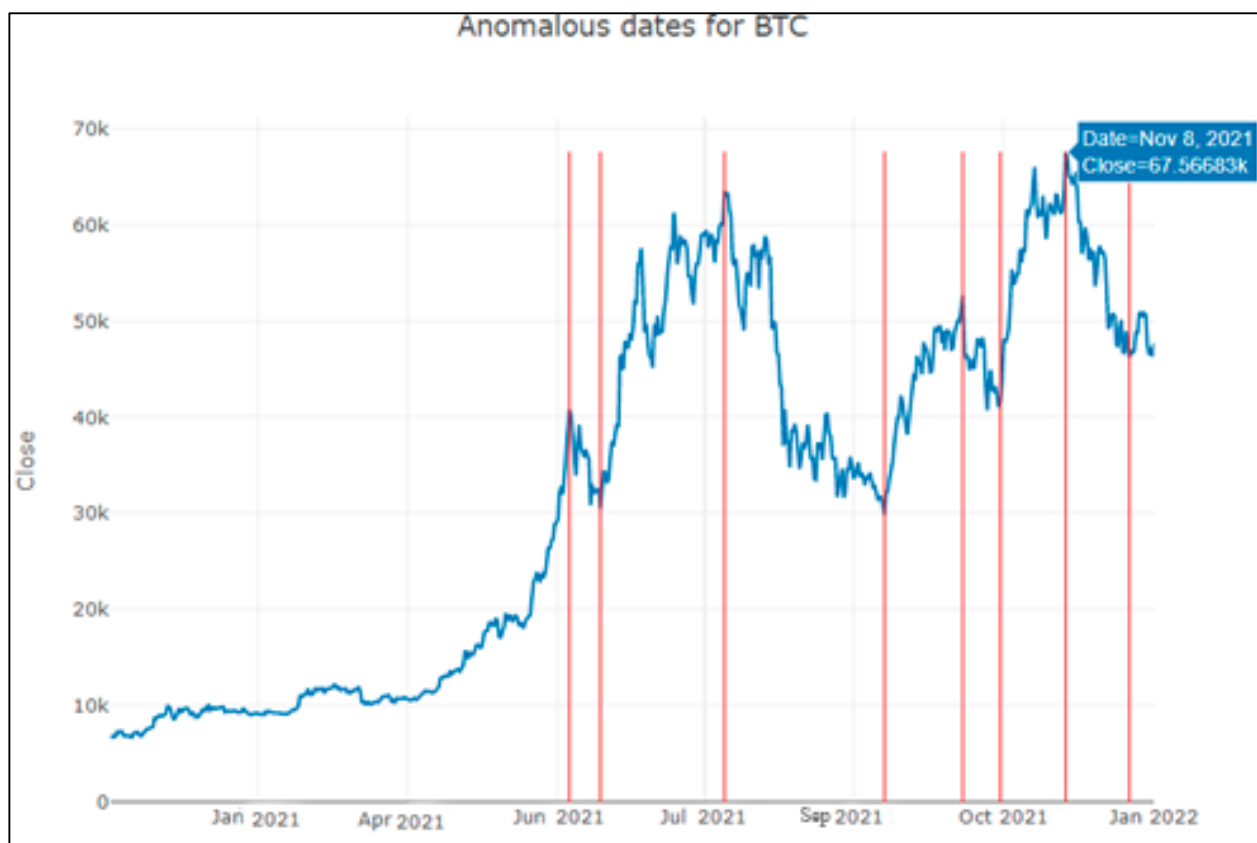


Рисунок Г.2 – Курс криптовалюти Bitcoin та дати аномальних змін

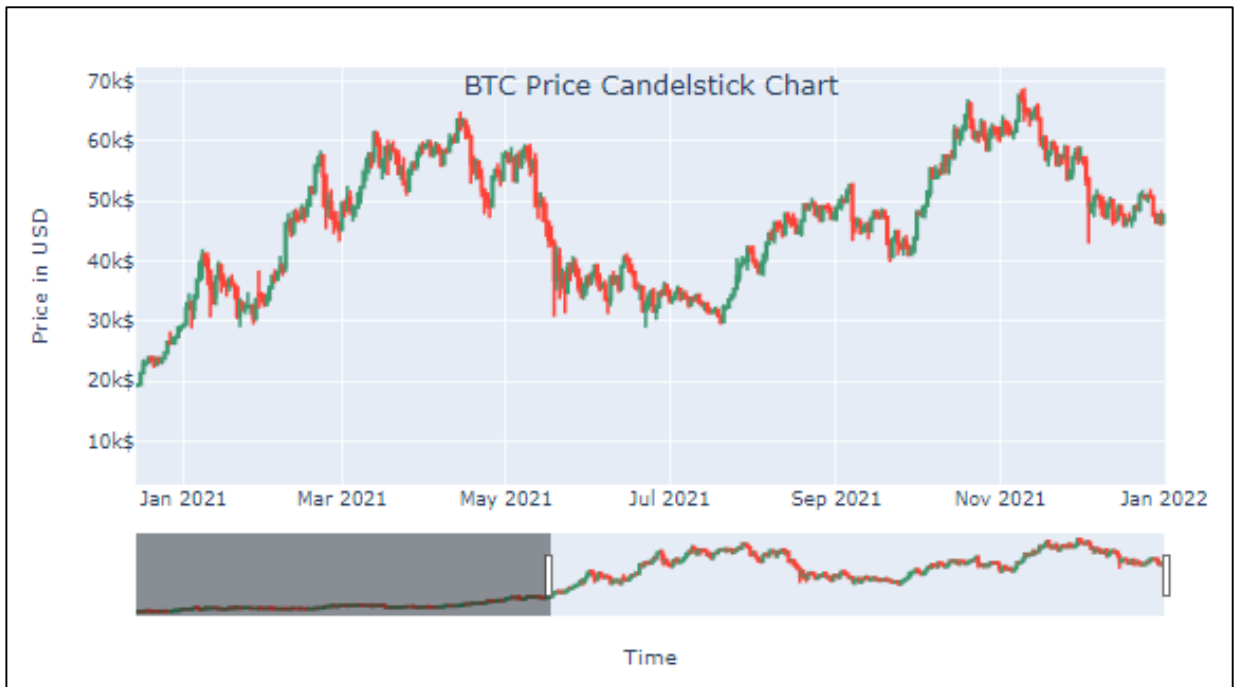


Рисунок Г.3 – OHLC-діаграма криптовалюти Bitcoin за даними сайту Yahoo

name_model	type_data	r2_score	rmse	mape
MLP Regressor	valid	-0.198116	1351.685712	2.446971
Random Forest Regressor	valid	-1.346753	1891.733477	3.600522
XGB Regressor	valid	-1.681957	2022.330814	3.910245
ARIMA_manual	valid	-2.134046	2186.144947	4.155419
ARIMA_auto	valid	-2.134046	2186.144947	4.155419
Prophet_4_days_3_order	valid	-3.500235	2619.653422	4.378228
Prophet_7_days_3_order	valid	-3.484681	2615.122624	4.434402
Prophet_5_days_3_order	valid	-3.451903	2605.548101	4.459947
Prophet_14_days_12_order	valid	-3.980607	2755.924964	4.67282
KNeighbors Regressor	valid	-2.878011	2431.816508	4.742777
Prophet_7_days_12_order	valid	-4.348056	2855.776728	4.823952
Prophet_14_days_3_order	valid	-4.210608	2818.840293	4.854828
Prophet_5_days_12_order	valid	-4.572239	2915.017259	4.903258
Prophet_4_days_12_order	valid	-4.853504	2987.680986	4.904383
Linear SVR	valid	-3.263582	2549.843676	4.947899
Support Vector Machines	valid	-3.347282	2574.750635	5.000937
Linear Regression	valid	-7.766195	3656.215868	7.098678
Bagging Regressor	valid	-11.517699	4369.06509	7.56984

Рисунок Г.4 – Таблица похибок моделей

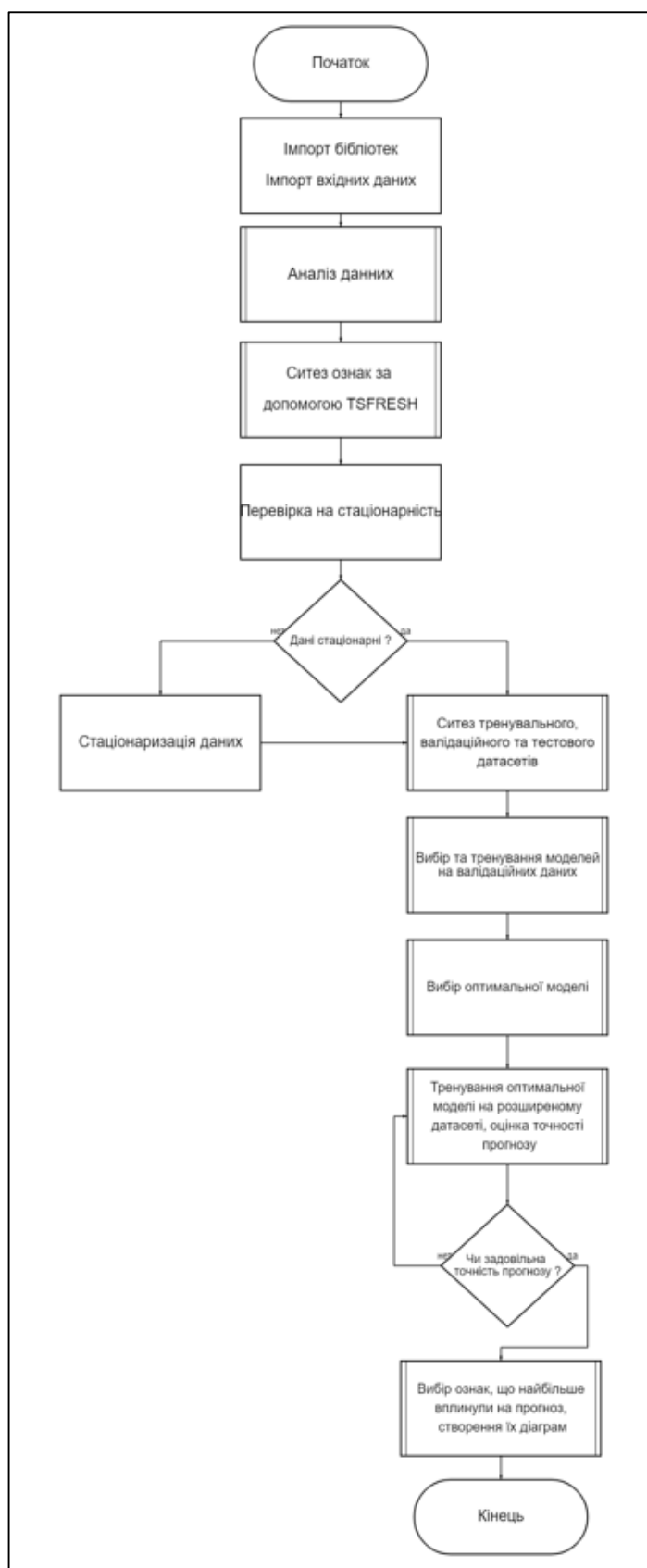


Рисунок Г.5 – Блок-схема алгоритму інформаційної технології прогнозування курсу Bitcoin у 2021-2022 роках

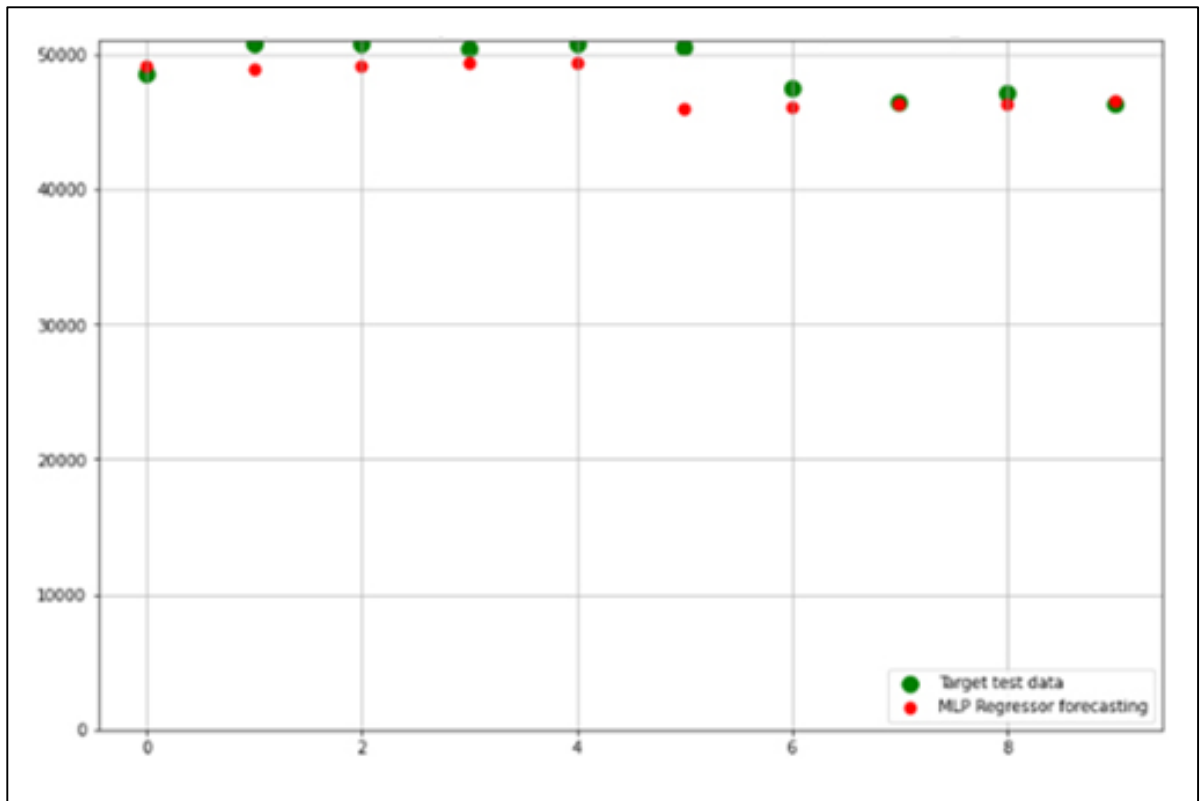


Рисунок Г.6 – Графік тестових даних курсу криптовалюти Bitcoin та їх прогноз за моделлю MLP Regressor