

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

## МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

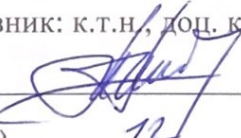
на тему:

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


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

 Жук В. Л.

Керівник: к.т.н., доц. каф. САІТ


 Жуков С. О.  
«02» 12 2022 р.

Опонент: к.т.н., доц. каф. ЗІ

 Куперштейн Л. М.  
«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.


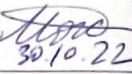


### ЗАВДАННЯ НА МАГІСТЕРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ

Жуку В'ячеславу Леонідовичу

1. Тема роботи: “Інформаційна технологія прогнозування курсу криптовалюти Gala у 2021-2022 роках”,  
керівник роботи: Жуков С. О., 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 Gala на світовому ринку за період з 01.04.2020 р. по 30.09.2022 р.
4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)
  - 1) Аналіз проблеми прогнозування курсу Gala.
  - 2) Ідентифікація та вибір оптимальної моделі для прогнозування курсу криптовалюти Gala.
  - 3) Розроблення інформаційної технології аналізу та прогнозування курсу криптовалюти Gala у 2021 – 2022 роках.
  - 4) Економічна частина
5. Перелік ілюстративного матеріалу:
  - Капіталізація ринку криптовалюти;
  - Графік курсу криптовалюти Gala та дати аномальних змін;
  - Таблиця похибки моделей;
  - OHLC-діаграма криптовалюти Gala за даними сайту Yahoo;
  - Блок-схема алгоритму інформаційної технології прогнозування курсу Gala у 2021-2022 роках;
  - Діаграма перестановок із важливістю ознак;

– Реальні тестові дані курсу криптовалюти Gala та їх прогноз 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	 01.10.22	 30.10.22
4	Буреннікова Н.В., д.е.н. проф. каф. ЕПВМ	 15.10.22	 26.10.22

7. Дата видачі завдання «16» 09 2022 року

#### КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів МКР	Строк виконання етапів роботи	Примі
1	Аналіз предметної області	09.2022	
2	Ідентифікація та вибір оптимальної моделі для прогнозування криптовалюти Gala	09.2022	
3	Розроблення інформаційної технології для аналізу та прогнозування курсу криптовалюти Gala	10.2022	
4	Економічна частина	10.2022	
5	Оформлення матеріалів до захисту МКР	11.2022	

Студент



Жук В. Л.

Керівник роботи



Жуков С. О.



## АНОТАЦІЯ

УДК 004.09

Жук В. Л. Інформаційна технологія прогнозування курсу криптовалюти Gala 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. 95 с.

На укр. мові. Бібліогр.: 20 назв; рис.: 60; табл.: 5.

В магістерській кваліфікаційній роботі проведено аналіз предметної області, обґрунтовано доцільність та актуальність даного дослідження. Проведено розвідувальний аналіз даних, проаналізовано фактори які впливають на курс криптовалюти Gala, виділено аномальні дати. Побудовано ряд моделей з використанням методів машинного навчання та зроблено прогноз курсу криптовалюти Gala. Виділено оптимальну модель та відібрано важливі ознаки, які найбільше вплинули на прогноз. На основі оптимальної моделі розроблено інформаційну технологію прогнозування курсу криптовалюти Gala 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 Gala.

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

## **ABSTRACT**

Zhuk V. L. Information technology for forecasting the course of the Gala cryptocurrency in 2021-2022. Master's qualification work in the specialty 126 – information systems and technologies, educational and professional program – information technologies of data and image analysis. Vinnytsia: VNTU, 2022. – 95 p.

In Ukrainian language. Bibliographer: 20 titles; fig.: 60; table: 5.

In the master's qualification work, attention was paid to the analysis of the subject area, the expediency and relevance of this research was substantiated. Factors affecting the cryptocurrency exchange rate were analyzed, anomalous dates were highlighted. A forecast of the 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.

The illustrative part consists of 8 posters with simulation results.

In the section of the economic part, the issue of the feasibility of developing and implementing an information system for monitoring passenger road transport is considered.

Key words: information system, digital currency, cryptocurrency, neural network, machine learning, Gala.

## ЗМІСТ

ВСТУП.....	4
1 АНАЛІЗ ПРОБЛЕМИ ПРОГНОЗУВАННЯ КУРСУ GALA.....	6
1.1 Аналіз проблем прогнозування курсу криптовалюти Gala.....	6
1.2 Формалізація постановки задачі .....	9
1.3 Аналіз інформаційних технологій для прогнозування та вибір оптимальної.....	12
1.4 Висновки.....	20
2 ІДЕНТИФІКАЦІЯ ТА ВИБІР ОПТИМАЛЬНОЇ МОДЕЛІ ДЛЯ ПРОГНОЗУВАННЯ КУРСУ КРИПТОВАЛЮТИ GALA.....	21
2.1 Аналіз можливостей мови програмування Python.....	21
2.2 Розвідувальний аналіз даних.....	23
2.3 Синтез нових ознак .....	42
2.4 Побудова моделей методами машинного навчання .....	46
2.5 Вибір оптимальної моделі .....	47
2.6 Висновки .....	48
3 РОЗРОБЛЕННЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ КУРСУ КРИПТОВАЛЮТИ GALA У 2021-2022 РОКАХ	50
3.1 Розроблення інформаційної технології прогнозування курсу Gala .....	50
3.2 Прогнозування тестових даних.....	54
3.3 Висновки .....	60
4 ЕКОНОМІЧНА ЧАСТИНА.....	61
4.1 Комерційний та технологічний аудит науково-технічної.....	61
4.2 Прогнозування витрат на виконання науково-дослідної (дослідно-конструкторської) роботи .....	64
4.3 Розрахунок економічної ефективності науково-технічної розробки за її можливої комерціалізації потенційним інвестором .....	69
4.4 Висновки .....	75

	3
ВИСНОВКИ.....	76
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	78
Додаток А (обов'язковий). Технічне завдання.....	80
Додаток Б (обов'язковий). Протокол перевірки кваліфікаційної роботи на наявність текстових запозичень.....	83
Додаток В (довідниковий). Лістинг програми .....	84
Додаток Г (обов'язковий). Ілюстративна частина .....	88

## ВСТУП

**Актуальність теми.** Сьогодні прогнозування стало частиною нашої реальності. Економіка складається з оцінок: політичних обставин, розвитку та видатності досягнень, збору в аграрній зоні, клімату, поточних обставин із видобутком корисних копалин для різних підприємств, а також показників вартості різних ресурсів і грошових стандартів [1]. Ця велика кількість видів поточних даних і прогнозів впливає на інтерес фінансового ринку.

Оскільки на купівлі та продажу криптовалюти можна заробити гроші, ринок криптовалют посів високе місце у світовій економіці. Однією із таких криптовалют є Gala. Щоб стати власником криптовалюти Gala цілком достатньо підключитися до сервісу її створення, тобто стати учасником єдиної мережі Digging. Прогнозування курсу криптовалюти Gala є актуальною та малодослідженою темою, оскільки на її ціну впливає безліч факторів. Для прогнозування криптовалюти можна використати моделі прогнозування часових рядів.

**Мета і задачі дослідження.** Метою дослідження є підвищення точності прогнозування курсу криптовалюти Gala за рахунок використання методів машинного навчання.

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

- проаналізувати існуючі проблеми в прогнозуванні курсу криптовалют;
- провести аналіз існуючих технологій та вибрати оптимальну модель машинного навчання;
- підвищити точність прогнозування курсу криптовалюти Gala на криптовалютному ринку.

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

**Предметом дослідження** є методи інформаційних технологій, які використовувались для аналізу та прогнозування курсу криптовалюти Gala.



**Методи дослідження.** У роботі використано методи розвідувального аналізу для дослідження даних, методи аналізу часових рядів для прогнозування даних та технології машинного навчання для оброблення інформації.

**Новизна одержаних результатів.** Дістала подальший розвиток інформаційна технологія прогнозування курсу криптовалюти Gala у 2021-2022 роках, яка дозволяє підвищити точність прогнозування цього курсу.

**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.) [2].

# 1 АНАЛІЗ ПРОБЛЕМИ ПРОГНОЗУВАННЯ КУРСУ GALA

## 1.1 Аналіз проблем прогнозування курсу криптовалюти Gala

Криптовалюта Gala була розроблена у якості ігрового токена, тому прогнозування курсу даної криптовалюти є складним завданням.

Із часу створення в 2019 році криптовалюта Gala налічує понад 13 мільйонів користувачів щомісяця а також реалізовано 26 000 nft найдорожчий екземпляр якого оцінюється в 3 мільйони доларів [3]. Місце Gala за капіталізацією серед 25 суміжних криптовалют зображений на рисунку 1.1

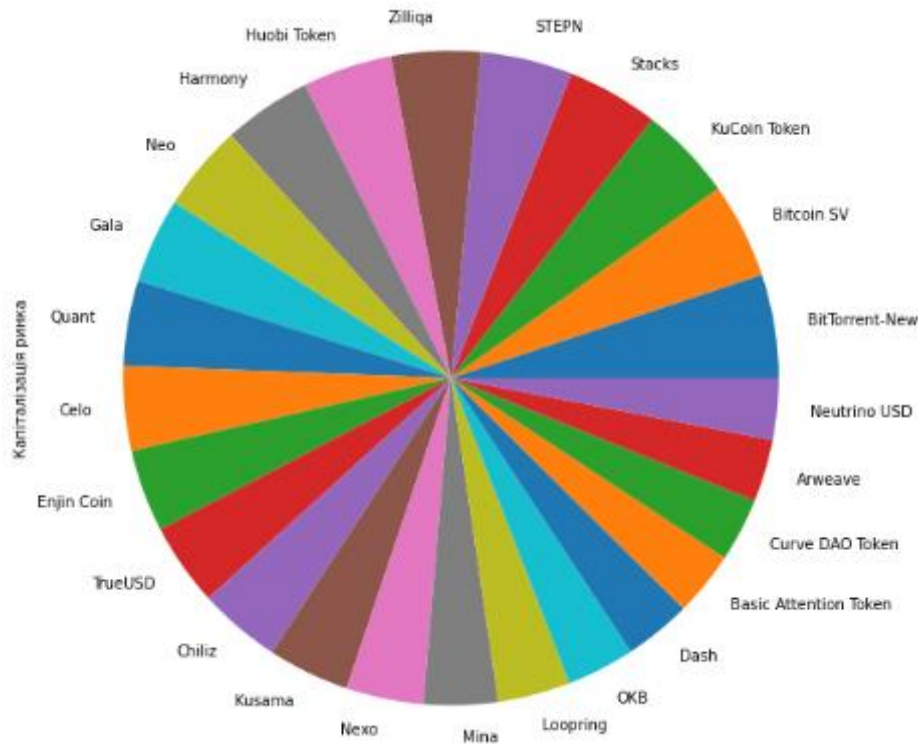


Рисунок 1.1 – Місце Gala за капіталізацією серед 25 суміжних криптовалют за даними сайту Yahoo станом на 12.04.2022 р.

Найбільше поширення ця криптовалюта отримала у таких країнах, як США, Канада, Мексика, Британія, Китай.

Gala games заснував Ерік Шірмейєр який також допоміг заснувати відому організацію соціальних і портативних ігор Zynga. За ініціативи Ширмейєра Zynga створила такі відомі хіти як Poker mafia, Wars і Farmville Schiermeyer який став відомим завдяки тому, що прагнув створювати нові популярні ігри.

У 2019 році втілюючи ідею створити ігрову організацію засновану на блокчейні і дати гравцям більше контролю над своїми іграми. Компанія із розробки ігор складається з 60 представників, напевно це не поодинокі гра а радше вона пропонує цілу виставку різних блокчейнів ігри, які ставлять розваги на перше місце, а блокчейн за лаштунками town star це відпочинок towns це просто програмна гра в реальному часі. Гравці мають справу з таким містом як simcity постійна взаємодія але навколо зірки вони дійсно володіють містом VOX це колекційний символ nft встановлений Gala games кожен VOX є винятковим і деякі VOX більш незвичайні ніж інші в першому рг протистояння 8888 vox було передано за типовою ціною 00888 eth 280 доларів США на годину написання [4].

Nft власники можуть заблокувати нагороду та придбати VOX coin граючи зі своїми символами VOX згодом власники VOX додатково отримують власний запис fbх який дозволяє їм оживити свої символи або навіть 3d-імпринтинг у тому що є в наявності є значною частиною у вдосконаленні гри функція критики локальної зони. Ігри постійно залучають і перевіряють припущення на своїй локальній території щодо конфлікту - це дозволяє гравцям формувати курс і план ігор. Клієнти можуть запускати центри прогнозування, центри підтримки, які допомагають мережі працювати, отже вони отримують винагороди наприклад: функція обмеженого випуску nft і різні функції відкритих дверей середовище центру складається з потрібної безпеки центри verification of-work row підтвердження виконаної роботи докази частки disconfirmation права власності на пропозицію та підтвердження потужності. Свідоцтво потужності row називаються піонерськими центрами оскільки вони є центрами 1-го рівня. З огляду на повне право власності 50 000 nft є першими користувачами організації які отримують nft від кожної наступної гри та Gala призначеного для їхнього центру permit pos - це платні центри, які працюють для конкретних ігор через

структуру оренди в розумній угоді доказом потужності є безкоштовні хаби які дозволять вам повністю мати ігри в біологічній системі хабів, позбавляючись від залежності від зосереджених механізмів сприяння наприклад amazon s3 [5].

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

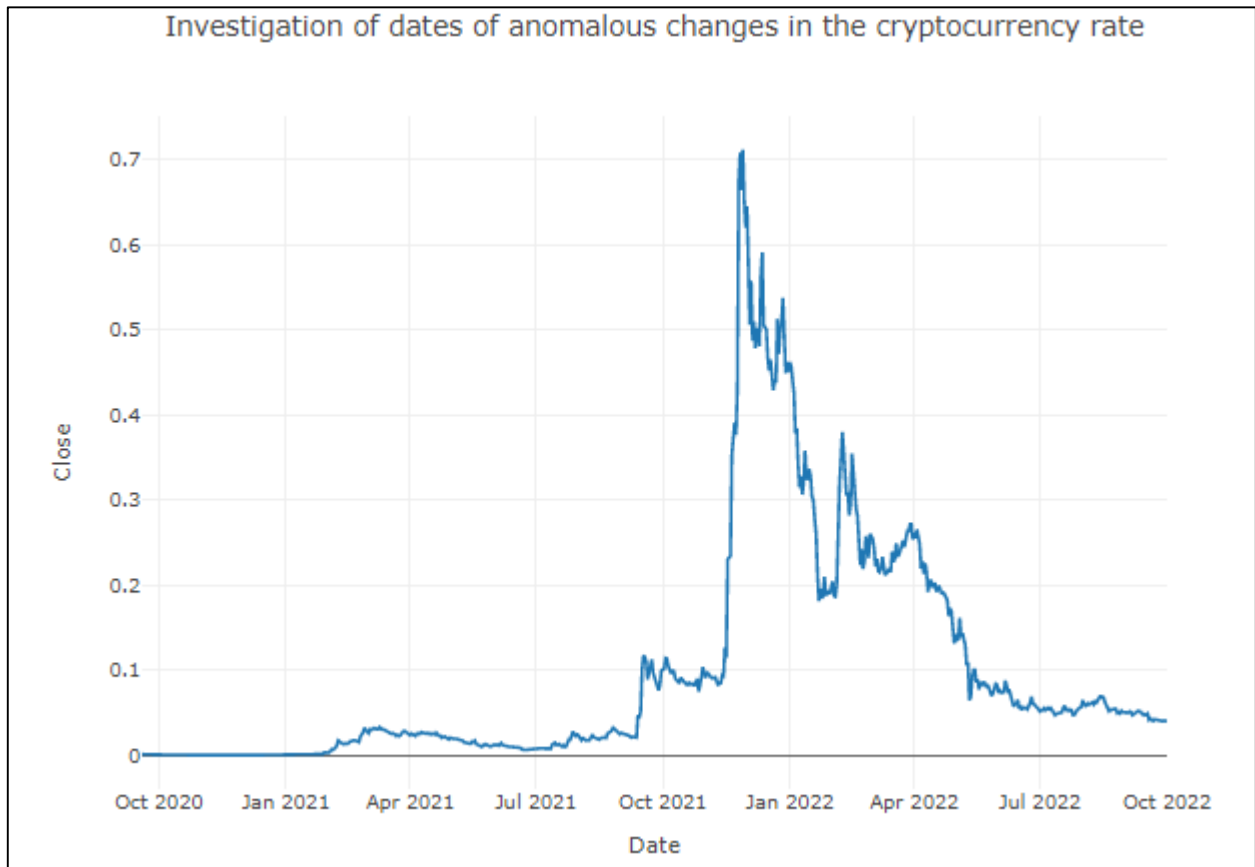


Рисунок 1.2 – Курс Gala за даними сайту Yahoo

Отже, прогнозування Gala є складною задачею, яка потребує чіткого аналізу курсу, а також факторів які впливають на курс. У кожній країні склалися специфічні ситуації стосовно криптовалют, відрізняється вона і в Україні, особливо поточна ситуація з криптовалютним регулюванням. Виникає необхідність вивчення впливу чинників на курс криптовалюти, наприклад Gala, адже сьогодні прогнозування руху курсу криптовалют та динаміки їх розвитку є складним завданням.

## 1.2 Формалізація постановки задачі

Оскільки датасети із криптовалютних ринків містять базові ознаки, тому вони подані на рисунку 1.3

	High	Low	Open	Close	Volume
Date					
2020-09-18	0.001534	0.001436	0.001437	0.001534	89173.0
2020-09-19	0.001810	0.001214	0.001535	0.001491	78421.0
2020-09-20	0.001784	0.001382	0.001492	0.001582	76880.0
2020-09-21	0.001584	0.001185	0.001582	0.001288	91765.0
2020-10-02	0.000932	0.000781	0.000901	0.000858	115883.0
...	...	...	...	...	...
2022-09-27	0.042246	0.039707	0.040853	0.040285	129931012.0
2022-09-28	0.040607	0.038653	0.040282	0.040172	106797618.0
2022-09-29	0.040925	0.039448	0.040176	0.040895	90196728.0
2022-09-30	0.041101	0.040065	0.040889	0.040543	111533658.0
2022-10-01	0.040816	0.040048	0.040539	0.040258	61000313.0

Рисунок 1.3 – Базові ознаки датасету по криптовалюті Gala за даними сайту Yahoo

Графіки барів, або графіки OHLC (Open-High-Low-Close), використовуються як трейдинговий інструмент для візуалізації та аналізу змін цін на облігації, валюти, акції, цінні папери тощо за певний період часу.

Графіки OHLC зручні для регулярного аналізу настроїв ринку та прогнозування цінових змін у майбутньому на основі виявлених патернів. На осі Y графіка OHLC розміщується шкала цін, але в осі X – шкала часу. У кожному конкретному періоді часу на графіку OHLC зображується графічний символ, який демонструє два діапазони показників: найвищу (максимум) і найнижчу (мінімум) ціну торгів, а також ціну відкриття та ціну закриття за аналізований період часу (наприклад, день). На графічному символі найвища та найнижча ціна представлені довжиною основної вертикальної лінії [6].

Ціна відкриття та ціна закриття представлені положенням міток на вертикальній лінії: мітка зліва вказує на ціну відкриття, а мітка праворуч – на ціну закриття. Кожному графічному символу на графіку OHLC може бути наданий колір. Це допоможе відрізнити "бичачий" ринок (ціна закриття вище, ніж ціна відкриття) від "ведмежого" (ціна закриття нижче, ніж ціна відкриття).

Надаємо дані OHLCV у деталях в інтервалі від 1 секунди до 1 дня.

- «Open» – Ціна відкриття інтервалу часу у валюті котирування (для Gala/USD ціна буде USD).
- «High» – Найвища ціна, досягнута протягом даного інтервалу часу, у валюті дослідження.
- «Low» – Найнижча ціна, досягнута протягом даного інтервалу часу, у валюті дослідження.
- «Close» – Ціна закриття даного інтервалу часу, у валюті дослідження.
- «Volume» – Кількість купленого або проданого активу, що відображається в базовій валюті.

API на Yahoo містить ще «Adj Close», зазвичай він рівний параметру «Close» тому ним нехтують.

Саме ознака «Close» і вважається ознакою, яка відображає поточний курс криптовалюти.

Важливо точно сформулювати набори даних для підготовки, затвердження (для вибору ідеальної моделі) і тестування (для остаточної оцінки очікуваної точності) [7].

Відповідно до зобов'язань, викладених в окремому дорученні, під час виконання завдання за допомогою методів штучного інтелекту показник «Target» повинен визначатися швидкістю до кінця кожного «Close» дня відповідно до формули:

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



де  $t$  – час, днів;  $d$  – кількість днів, на які слід навчити модель робити прогноз (зсув робиться саме назад, тоді з'являється можливість по значеннях ознак у  $t$  добу прогнозувати цільову ознаку у  $t+d$  добу).

Ще один компонент моделей часових рядів полягає в тому, що і тестовий, і реальний набори даних вибираються в кінці інформації про сприйняття. Основна причина для схвалення набору даних полягає в тому, щоб допомогти у виборі ідеальної моделі, яка найкраще передбачить майбутні якості, а не минулі якості.

Взагалі кажучи, тестова інформація – це останні  $d$  («forecasting\_days»). Усі минулі якості структурують підготовчий набір даних.[8]

Знову ж таки, після визначення ідеальної моделі, яка дає найкращі очікування щодо інформації про схвалення для певного (або кількох) показників, її слід перерахувати на основі розширеного набору даних підготовки, тобто наборів даних підготовки та схвалення, об'єднаних разом.

Прогнозування часового ряду зображено на рисунку 1.4



Рисунок 1.4 – Прогнозування часового ряду

### **1.3 Аналіз інформаційних технологій для прогнозування та вибір оптимальної**

Цифрова валюта є однією із найсучасніших, найменш вивчених і найпростіших способів вкладення ресурсів у фінанси. Постійно зростаюча кількість людей потребує вкладення ресурсів у бізнес, щоб отримати зарплату. Будь-хто, хто має доступ до ПК із доступом до Інтернету, може виконати в цифровій валюті. Цю поширеність також можна зрозуміти через популярність розмов на цю тему в Інтернеті. Останнім часом численні активи створюють враження, що пропонують оплату неактивною цифровою валютою без ваших відсотків. Незважаючи на великий ризик спекуляцій, постійно зростаюча кількість людей готова випробувати свої здібності, у світлі того факту, що потенційний чемпіон може досягти успіху в цьому тоталізаторі. Більше тонкощів можна знайти в [4]. На даний момент доступно понад 1000 значків різних криптографічних форм грошей. Більшість цифрових валют, у будь-якому випадку, не користуються попитом на тій підставі, що вони фактично такі ж, як їхні предки та намагайтеся не використовувати переконливі великі літери. Масштаб конвертації нових цифрових форм грошей здебільшого залежить від просування, що дозволяє вам передбачити масштаб їх обміну, досліджуючи зовнішні елементи. Більше тонкощів можна знайти в [9].

Блискучий інтерес до цифрових валют може принести величезні прибутки. Щоб отримати готівку, обмінюючи криптографічні форми грошей, ви повинні мати можливість точно передбачити майбутні зміни витрат.

Курс криптовалюти можна прогнозувати методами машинного навчання з використанням дерев рішень, нейронних мереж та інших моделей, які враховують вплив багатьох ознак, або – як періодичний ряд, навіщо використовувати авторегресію та включені рухомі нормальні моделі (такі як "ARIMA") та їхні аналоги або Facebook Prophet з різними параметрами сезонності та коефіцієнтів Фур'є для врахування ступеню їх нелінійності (рис. 1.5) [10].

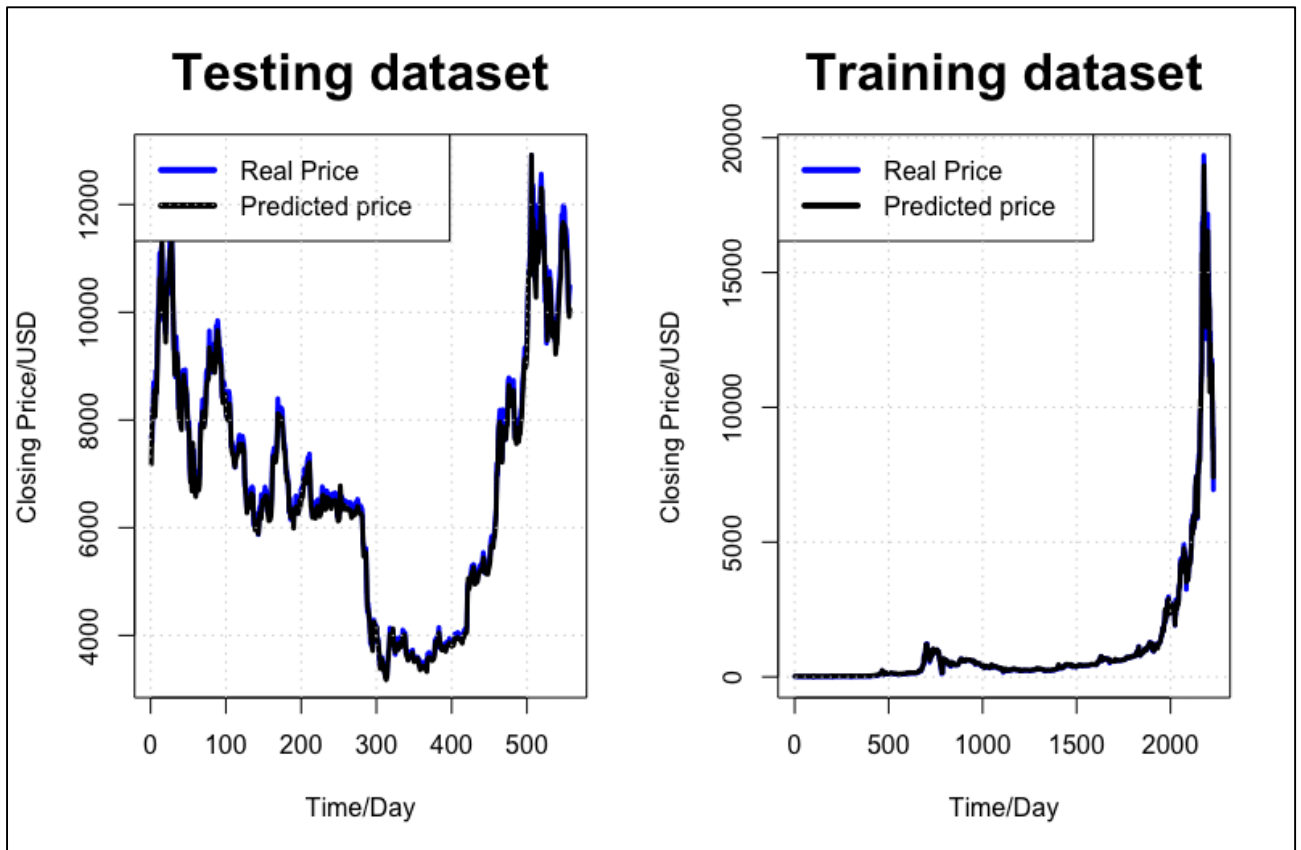


Рисунок 1.5 – Прогнозування часового ряду

Вирішення будь-якого питання вивчення інформації та прогнозування за допомогою штучного інтелекту зазвичай виконується на наступних етапах [7]:

- очищення інформації (розпізнавання та вилучення помилкової та дивної інформації);
- розвідувальний аналіз даних (англ.: EDA — Exploratory Data Analysis);
- видобування ознак (англ.: FE — Feature Engineering).

Для того щоб визначити найбільш впливові чинники на формування ринкового курсу криптовалют необхідно визначити як саме досліджувати важливість чинників для певної криптовалюти. Найбільш важливі предиктори визначають за допомогою кореляції та коефіцієнтів моделі. Тому подальшому виникає потреба дослідити які існують методи кореляції, які моделі можна використовувати для прогнозування часових рядів [11].

З теорії часових рядів виокремимо для себе розділи для підготовки часового ряду та певних моделей прогнозу. Хоча статистичні моделі які зазвичай

використовуються в аналізі та прогнозування часових рядів вимагаю стаціонарності ряду, а в загальному випадку більші справжніх часових рядів стаціонарними назвати важко, та доволі точно працюють на такого вигляду ряду. Для даної роботи було обрано методи машинного навчання для прогнозування та виділення важливості кожного з елементів інформаційного впливу. Оскільки в загальному вони дають кращі результати ніж статистичні. Проте без використання підготовки даних для прогнозування не можна обійтися від теорії часових рядів, тому що за необхідне є якісне перетворення та підготовка даних. А за підготовку даних можна взяти підготовку даних як звичайних декількох статистичних моделей і в подальшому робити різні перетворення та досліджувати як вони впливають на різні моделі. Тому одним з важливих етапів визначення впливу інформаційних чинників є сам вибір інформаційних чинників, які можуть впливати на формування ринку. Обрання відповідного лагу для ряду, певних ковзних статистик тощо. Та обов'язково тренування моделей, щоб вони були адекватними.

Для проведення розрахунків рекомендується використано публічний ноутбук [9], адаптувавши його до постановки задачі в індивідуальному завданні. Також існують і інші методи прогнозування курсу криптовалюти.

Walletinvestor був розроблений у 2017 році, щоб задовольнити потреби ринку в чудових і безкоштовних спеціалізованих оцінках ставок цифрових грошей. Оскільки фізичне створення цифр є неймовірно виснажливим і потребує важливої інформації, ми вирішили створити штучну систему оцінки, засновану на здібностях мозку. Після плідних випробувань і позитивної критики вони почали потроху розширювати сферу адміністрування, щоб відповідати ринковим моделям і вимогам клієнтів. Більше тонкощів можна знайти в [12]. Починаючи з початку 2018 року, сайт надає цифри понад 50 000 пропозицій, кілька тисяч матчів Forex, понад 10 000 активів і основних товарів.

До середини 2019 року Walletinvestor проведено спеціалізоване дослідження та оцінки понад 50 000 грошових статей, що супроводжуються

основними елементами преміум-класу, такими як спеціальні припущення (клієнти преміум-класу можуть надавати цифри з урахуванням попередньо встановлених проміжків часу) або підтвердження дизайну діаграми. Більше тонкощів можна знайти в [13]. На межі 2019/2021 років вони додали ринки житла до свого портфолію, регулярно оновлюючи оцінки вартості землі приблизно в 28 000 міських районах США. Більше тонкощів можна знайти в [14]. Щомісяця сайт отримує 2 000 000 незнайомих гостей і 6 000 000 онлайн-відвідувань. Кількість зареєстрованих клієнтів постійно зростає, на сьогоднішній день їх понад 70 000. Наявність сайту відображено на рисунках 1.6 – 1.9.

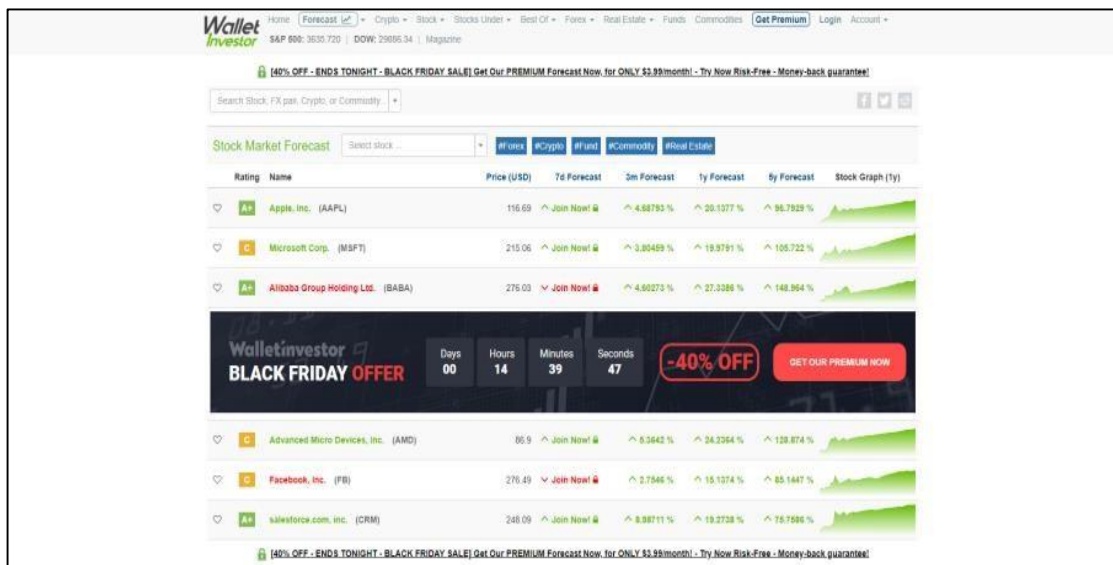


Рисунок 1.6 – Основний вигляд сайту Walletinvestor

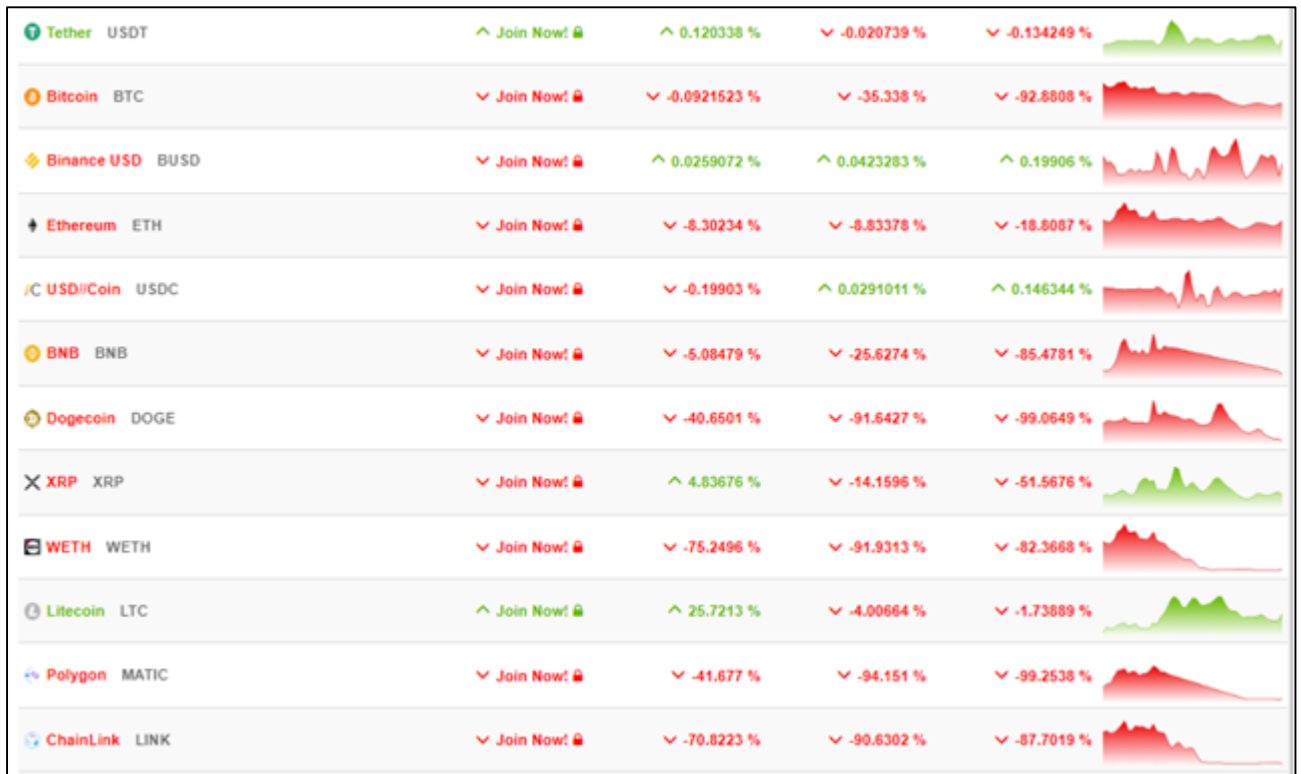


Рисунок 1.7 – Вигляд сайту Walletinvestor з прогнозами

Belinvestor – це сайт, на якому ви можете відслідковувати найновіші новини FOREX, огляди, припущення про грошові координати та фондові файли. Довідка також пропонує онлайн-ілюстрації та твердження.

Організація лише пропонує миттєву оцінку цифрових валют. На жаль, стратегії, які використовуються для очікування, також не розкриваються і є бізнесом. Більше тонкощів можна знайти в [15].

Trader – це структура яка дозволяє припустити розвиток стандарту конвертації готівки в торговій точці. Дані про минулі результати обміну (часові ряди) використовуються як початковий етап – найбільша, найменша вартість, вартість закриття та обсяг обміну щодня. Більше тонкощів можна знайти в [14]. Фреймворк передбачає супутні розрахунки для дослідження інформації: рухомий нормаль 3-х видів: прямий, чудовий, зі здачами; - MACD - гістограми, відомі маркери, такі як RSI, OBV, Williams R%, Candles, Point and Figure та багато інших. Клієнт може скласти власні рівняння для дослідження інформації. Переваги також включають можливість застосувати маркер до звичайного показника, що є фундаментальним, наприклад, під час створення гістограми



MACD, у якій рухома нормаль визначається за контрастом між двома рухомими нормальми. Більше тонкощів можна знайти в [16].

Elliott Wave Conjecture — одна з найбільших спеціалізованих експертних фірм на планеті. Вони щоденно охоплюють усі важливі пакети ресурсів, включаючи Forex, продукти, криптографічні гроші, світові рекорди, акції США. Допомога надає зручні та вичерпні відповіді щодо визначення та навчання для роздрібних та інституційних клієнтів у всьому світі, зокрема:

- Ідеальна передача останнього оновлення припущень з огляду на встановлений план;

- Ідеальні, прості у виконанні та майстерно виглядають контури. Діаграми також включають додатковий візуальний апарат, який допомагає учасникам ефективно розрізняти праву половину ринку (підйом, низхідний тренд або бічний) у світлі стратегії розташування. Більше тонкощів можна знайти в [17];

- Допомога доступна 24 години щодня;

- Більш вірне використання гіпотези хвилі Елліотта.

У довгостроковій перспективі вони розвивали додаткові методи для роботи над надійністю вимірювального приладу, включаючи ринкові зв'язки, цикли, групування коливань і структуру циркуляції.

Нейронні мережі використовуються для передбачення випадків. Ainet - це програма не дуже цінна для дослідження часових рядів, але вона дає чудові результати для деяких доручень, які потребують втручання інформації. Несприятливі наслідки можна отримати шляхом екстраполяції інформації. Межа іспиту є фактором покарання, який програма покращує самостійно. Результатом є прямокутна структура з усіма налаштованими відомостями та мережею з аналогічною кількістю розділів, але деяка інформація відсутня. Програма намагається передбачити цінність цієї відсутньої інформації. Більше тонкощів можна знайти в [18]. Слабкість програми полягає в тому, що вона працює згідно з технікою «чорної скриньки» і не дозволяє клієнту, який цікавиться мозковими організаціями, з'ясувати розрахунок прогнозу.

NeuroShell – це унікальний нейропакет, призначений виключно для визначення секторів монетарного бізнесу. Це правило дизайнерів спонукало до легкого розуміння інтерфейсу та можливості роботи з мозковими мережами без програмної інформації. Перевагою NeuroShell Informal Investor є те, як він виконує принцип просування, використовуючи спадкові розрахунки. Це економить багато машинного часу, нещодавно витраченого на вибір найкращих меж конкретного вимірювання та аналіз конкретного внеску в організацію мозку.

Неформальний інвестор NeuroShell зосереджується на розбудові системи обміну. Сама структура обміну може включати показники, а також очікувані значення, отримані від мозкових мереж у своїх стандартах. Найпоширеніший спосіб побудови мозкової мережі дуже простий, однак жоден із життєво важливих періодів побудови мозкової мережі не залишається в тіні. Основною конструкцією NeuroShell Informal investor є багатошаровий перцептрон. Більше тонкощів можна знайти в [19].



Рисунок 1.8 – Основний вигляд сайту NeuroShell Trader

NeuroShell Trader - це програмування для створення фреймворків обміну. По суті, це все, що завгодно, але не структура обміну, це інструментальний відсік як звичайних методів, так і створених людиною

стратегій інтелекту (симуляції інтелекту), до яких ви можете приєднатися, щоб створити власні структури електронного обміну. Механізми обміну можуть включати стандартні спеціалізовані вказівки на розслідування та рішення, які торговці використовували протягом дуже тривалого часу, штучні інтелектуальні процедури, такі як нейронні мережі. Структури обміну, які ви створите, будуть, отже, перевіряти та продовжувати передавати повідомлення в майбутнє, коли надходить нова інформація.



Рисунок 1.9 – Використання індикаторів NeuroShell для зняття даних с датчиків

NeuroShell Trader має велику бібліотеку яка містить понад 800 технічних показників. На додаток до стандартних показників, таких як *moving average*, *rate-of-change* або *stochastics*, розробники NeuroShell Trader додали низку інших показників [20].

## **1.4 Висновки**

У першому розділі розглянуто проблеми прогнозування курсу криптовалюти Gala. Проаналізовано існуючі технології прогнозування, отримано базові ознаки датасету по криптовалюті Gala за даними сайту Yahoo, обрано для роботи методи машинного навчання для прогнозування та виділено фактори які впливають на курс. Проте без використання підготовлених даних для прогнозування не можна обійтися в теорії часових рядів, тому необхідним є якісне перетворення та підготовлення даних.

## 2 ІДЕНТИФІКАЦІЯ ТА ВИБІР ОПТИМАЛЬНОЇ МОДЕЛІ ДЛЯ ПРОГНОЗУВАННЯ КУРСУ КРИПТОВАЛЮТИ GALA

### 2.1 Аналіз можливостей мови програмування Python

З розвитком галузі інновацій даних, фреймворків даних Python ефективно використовується в області машинного навчання та прогнозування. Його легкість допомагає виконувати складні обчислення машинного навчання.

Python — це унікальна дешифрована об'єктно-розташована мова попереднього аранжування з серйозним потужним компонуванням. Його створив у 1990 році голландський інженер-програміст Гвідо ван Россум [7].

Python використовується для різних цілей: для створення ігор і веб-додатків, для підтримки внутрішніх інструментів для різних завдань. Мова також зазвичай використовується в логічній сфері для вивчення та вирішення прикладних питань [7].

Діалекти програмування через деякий час змінюються - інженери додають в них нові родзинки, а також виправляють ляпи. Це засіб, за допомогою якого виявляються різні адаптації мови. Наприклад, код, написаний на Python 2 загалом, не працюватиме в такому розумінні без додаткових змін [7].

Однією з потенційних незручностей Python є швидкість виконання коду. Python, звичайно, не є впорядкованою мовою. Код Python спочатку впорядковується у внутрішній байт-код, який потім виконується посередником Python. У більшості випадків використання Python призводить до більш повільних проектів на відміну від таких діалектів, як C.

У будь-якому випадку, сучасні комп'ютери з легкістю підтримують більшість додатків, тому для більшості додатків швидкість просування є вищим пріоритетом, ніж швидкість виконання, а програми на Python зазвичай створюються набагато швидше.

Крім того, Python легко досягається завдяки модулям, написаним на C або C++. Такі модулі можна використовувати для виконання частин програми, які створюють серйозне навантаження на процесор.

Біологічна система Python розвивається і може стати переважаючим етапом прикладного машинного навчання.

Основна причина використання Python для визначення часових рядів полягає в тому, що універсально корисна мова програмування може бути використана як для повороту подій, так і для створення.

Python чудово підходить для штучного інтелекту в світлі того факту, що самі обчислення штучного інтелекту важко зрозуміти. Працюючи з Python, дизайнеру не потрібно приділяти багато уваги безпосередньо створенню коду: він може зосередитися на вирішенні більш складних завдань, пов'язаних зі штучним інтелектом.

Проста лінгвістична структура мови Python допомагає інженеру перевіряти складні обчислення з незначним часом виконання.

Ще одна перевага Python — надзвичайна допомога та якісна документація. У Python є багато цінних активів, де інженер програмного забезпечення може знайти підтримку та керівництво, перебуваючи на будь-якому етапі трансформації [8].

Ще одна перевага Python у штучному інтелекті — це його адаптивність: наприклад, інженер має вибір між підходом до елемента та сценаріями. Python допомагає об'єднувати різну інформацію. Крім того, Python є особливо вигідним для тих дизайнерів, які створюють більшість коду, використовуючи IDE.

Як уже зазначалося, Python набув поширення завдяки своїй простій і чіткій структурі пунктуації. З цієї причини є багато інженерів Python, які готові відмовитися від проектів, пов'язаних зі машинним навчанням [9].

Записані вище змінні пояснюють, чому Python так ефективно використовується у сфері штучного інтелекту. Його легкість допомагає виконувати складні обчислення машинного навчання.



Основним прихильником, який спробував застосувати Python у реальності в 1990 році, був Гвідо Ван Россум. Нещодавно цей чоловік, відриваючись від інженера мови ABC голландської організації CWI, зрозумів, що житло не приносить нічого нового. Це стало початком для складання ще одного медіатора [12].

Захоплюючий момент полягає в тому, що основна робоча модель Python була створена на домашньому Mac Гвідо, і в двох-трьох кінцях тижня. Ті гроші розпорошуються, добиваються за допомогою Мережі [13].

У 1996 році, коли це підприємство набирало мінімальних обсягів, до повороту подій приєднався Стів Маєвські, який мав велику популярність в організаціях. Стів, як Гвідо, любив Mac, що могло стати передумовою їхніх скоординованих зусиль. Зауважте, що мова не названа «Python» за видом, як помилково очікують численні інженери. Під час удосконалення «Python» Гвідо отримав задоволення від перегляду сатиричного шоу «Повітряний карнавал Монті Пайтона», тому своїм завданням він назвав віддати належне Монті Пайтону [13].

З тих пір, як Python отримав фантастичний потенціал і був безперешкодно поширений в Інтернеті, коли він здавався центром прихильників - людей, які були зацікавлені у вдосконаленні Python як мови програмування. На початку своєї екскурсії ця назва мови створила невеликий перекладач із невеликими можливостями та повною відсутністю ООП, який не використовував його та спонукав до подальшого вдосконалення мови [12].

У 1991 році почав з'являтися основний метод розвитку ООП. Гвідо рекомендував зобов'язання в організаціях CNRI, яка розташована в Америці.

## **2.2 Розвідувальний аналіз даних**

Результат розвідувального аналізу даних курс Gala за даними сайту Yahoo за допомогою інструментарію Pandas Profiling Report зображений на рисунку 2.1

Dataset statistics		Variable types	
Number of variables	7	DateTime	1
Number of observations	713	Numeric	6
Missing cells	0		
Missing cells (%)	0.0%		
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	39.1 KiB		
Average record size in memory	56.2 B		

Рисунок 2.1 – Результат розвідувального аналізу даних курс Gala за даними сайту Yahoo за допомогою інструментарію Pandas Profiling Report

Результат розвідувального аналізу даних курс Gala за даними сайту Yahoo за допомогою інструментарію Pandas Profiling Report для характеристики Date зображено на рисунку 2.2

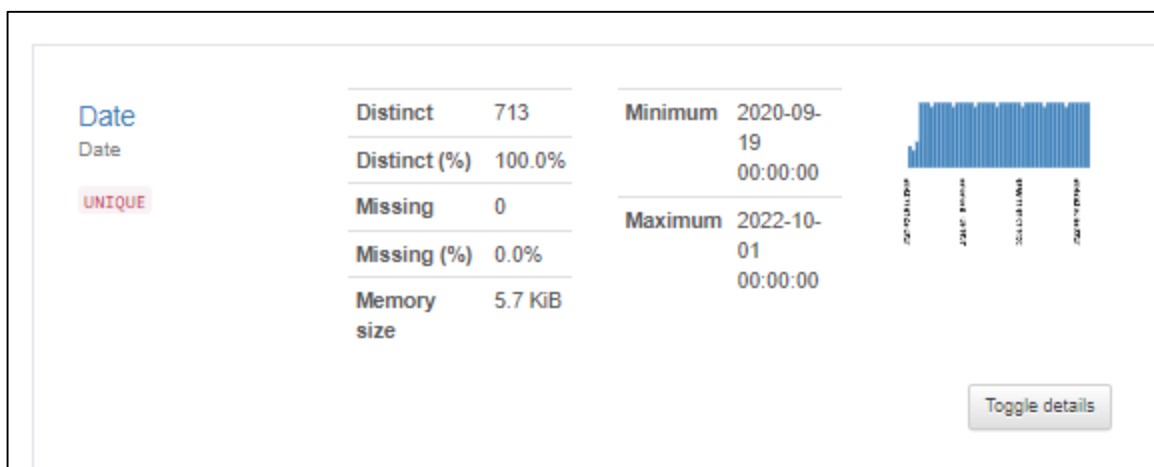


Рисунок 2.2 – Результат розвідувального аналізу даних курс Gala за даними сайту Yahoo за допомогою інструментарію Pandas Profiling Report (Date)

Результат розвідувального аналізу даних курс Gala за даними сайту Yahoo за допомогою інструментарію Pandas Profiling Report (для характеристик High, Low) зображено на рисунку 2.3

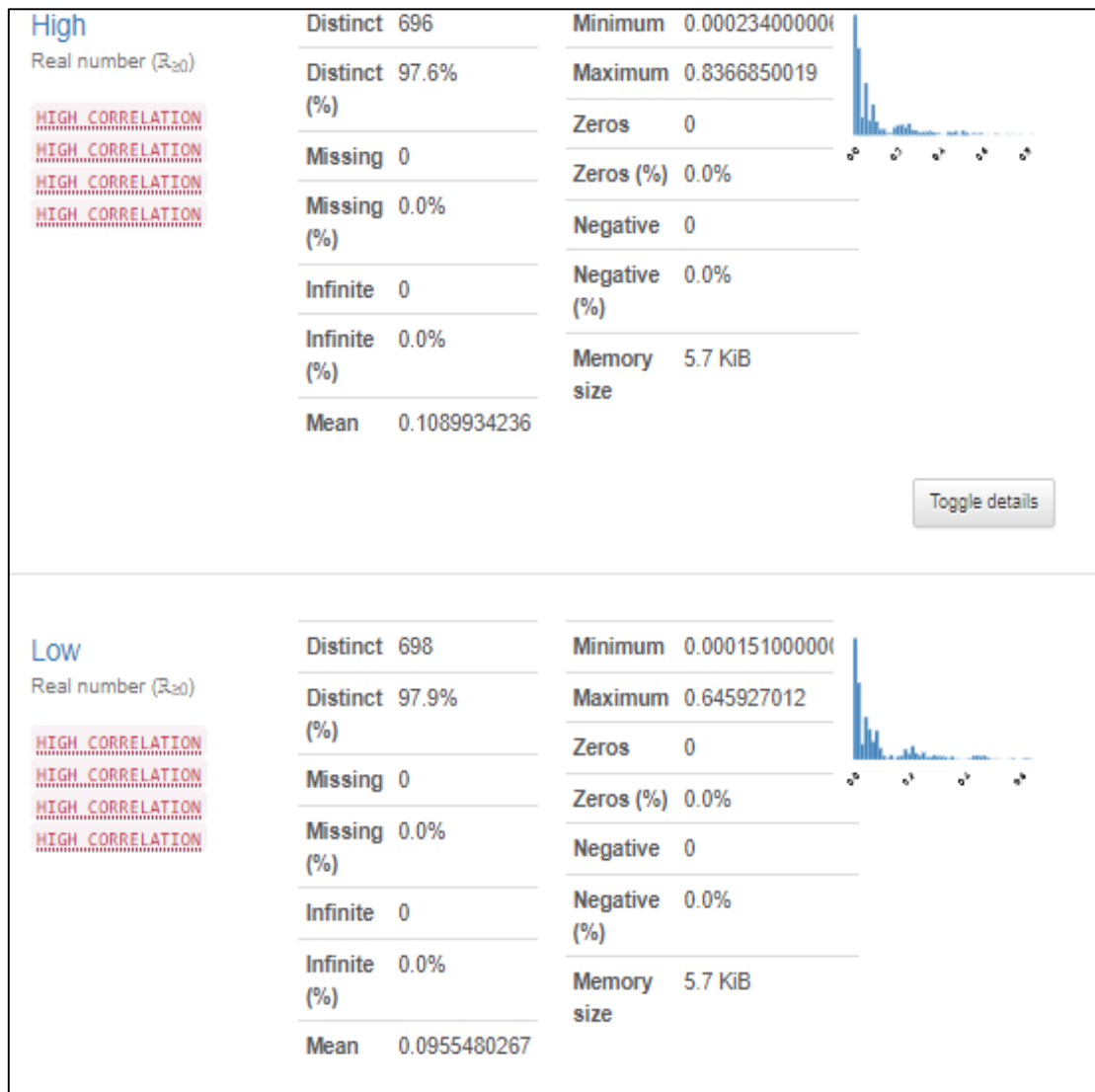


Рисунок 2.3 – Результат розвідувального аналізу даних курс Gala за даними сайту Yahoo за допомогою інструментарію Pandas Profiling Report (High, Low)

Результат розвідувального аналізу даних курс Gala за даними сайту Yahoo за допомогою інструментарію Pandas Profiling Report для характеристик Open, Close зображений на рисунку 2.4

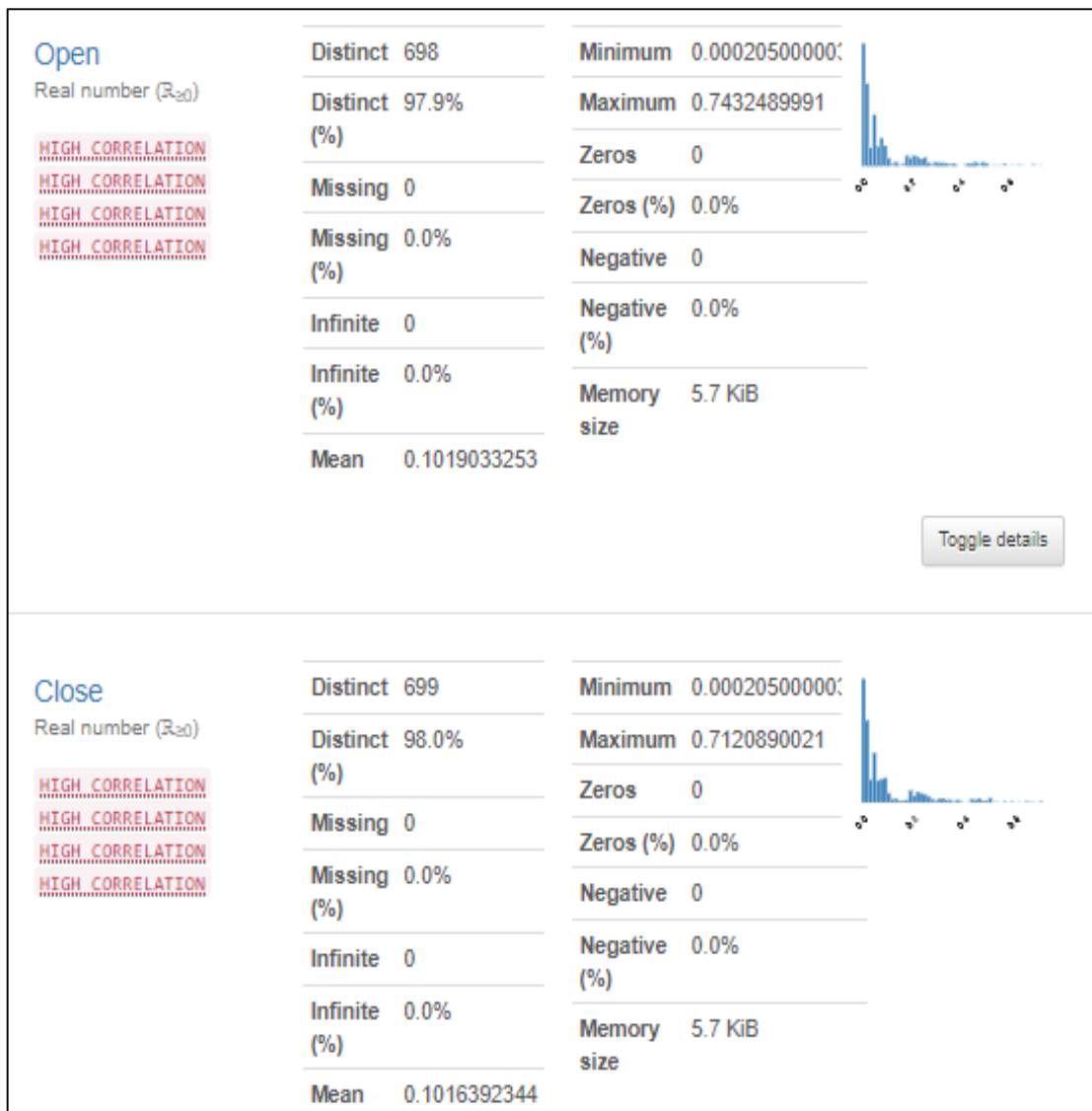


Рисунок 2.4 – Результат розвідувального аналізу даних курс Gala за даними сайту Yahoo за допомогою інструментарію Pandas Profiling Report (Open, Close)

Результат розвідувального аналізу даних курс Gala за даними сайту Yahoo за допомогою інструментарію Pandas Profiling Report для характеристик Volume, Close\_diff зображений на рисунку 2.5

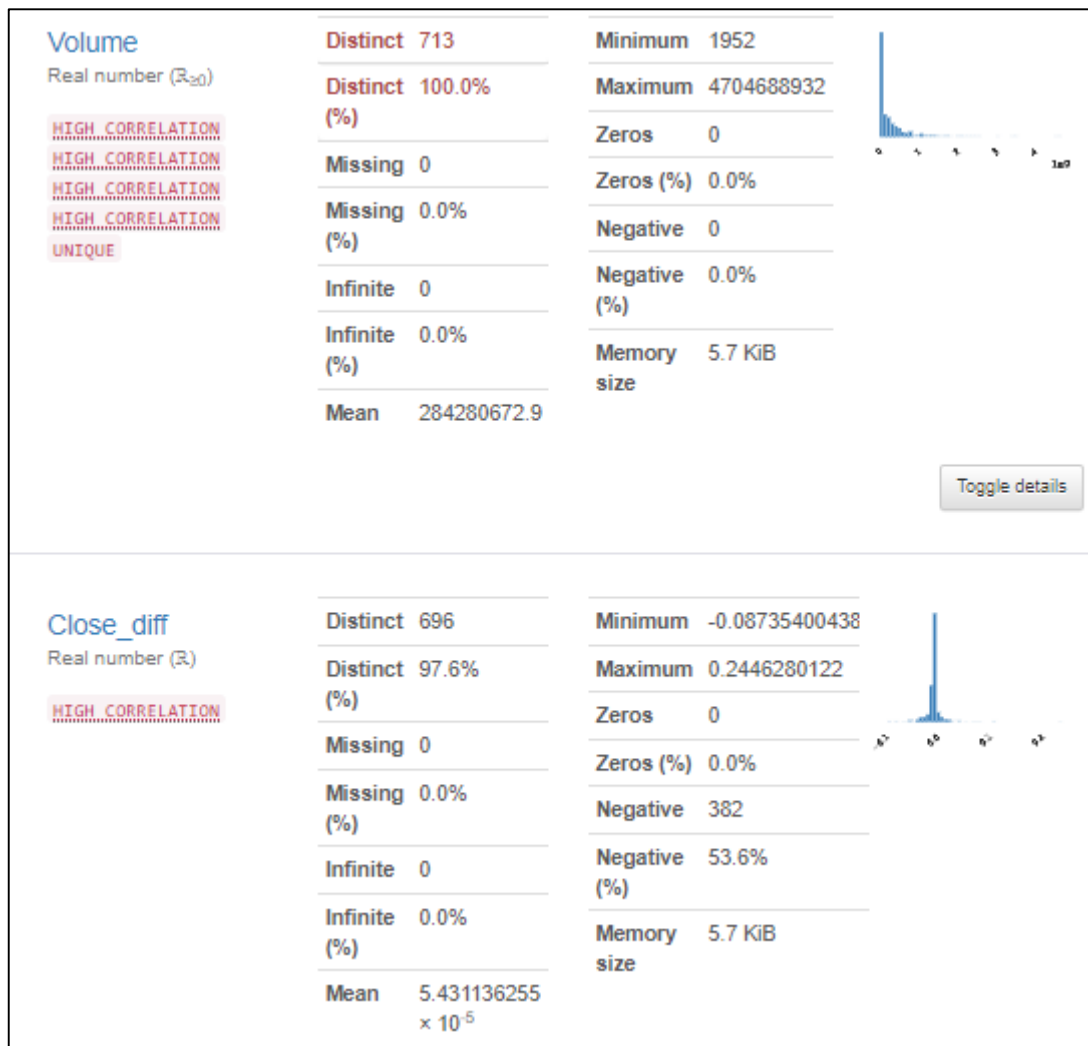


Рисунок 2.5 – Результат розвідувального аналізу даних курс Gala за даними сайту Yahoo за допомогою інструментарію Pandas Profiling Report (Volume, Close\_diff)

OHLC-ознаки такі як: «Open», «High», «Low», «Close» є дуже важливими, тому що за ними будують так звану OHLC-діаграму у вигляді вертикальних ліній червоного кольору (падіння курсу) та зеленого кольору (зростання) (рис. 2.6)

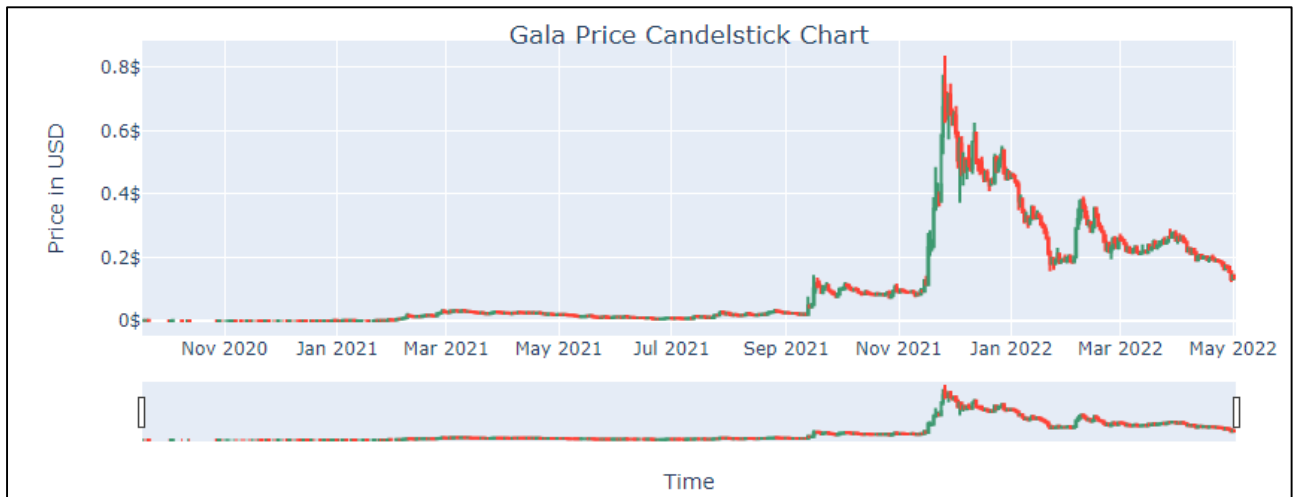


Рисунок 2.6 – OHLC-діаграма криптовалюти Gala за даними сайту Yahoo

Проаналізуємо ряд значень ознаки «Close» для того щоб отримати відповіді на декілька запитань:

1) Чи є ряд нестационарним і чи можна його стационаризувати взяттям першої чи другої різниці?

2) Чи має ряд якусь сезонність, а якщо – має, то скільки діб в її періоді?

Щоб перевірити, чи є ряд стационарним, використовується тест Дікі-Фуллера (рис. 2.7) [11]. Тест Дікі-Фуллера використовується для перевірки недейсних припущень про наявність одиничного кореня в моделі авторегресії. Вибіркова спекуляція змінюється залежно від адаптації тесту, але зазвичай передбачає наявність стационарності або стационарності шаблону. Названий на честь аналітиків Девіда Дікі та Уейна Фуллера, які створили тест у 1979 році.

Частина коду для перевірки на стационарність ряду криптовалюти Gala зображена на рисунку 2.7

```
# Stationarity check
check_stationarity(df['Close'])

ADF Statistic: -1.680958
p-value: 0.440949
Critical Values:
  1%: -3.442
  5%: -2.867
 10%: -2.570
Non-stationary
```

Рисунок 2.7 – Частина коду для перевірки на стаціонарність ряду криптовалюти  
Gala

Результат ознаки «Close» зображено на рисунку 2.8

```
# Stationarity check of the first difference of time series
check_stationarity(df['Close'].diff().dropna())

ADF Statistic: -6.142548
p-value: 0.000000
Critical Values:
    1%: -3.442
    5%: -2.867
   10%: -2.570
Stationary
```

Рисунок 2.8 – Результат перевірки на стаціонарність курсу криптовалюти Gala

Як бачимо, друга різниці ряду Gala за 2021-2022 рр. є стаціонарними часовими рядами. Оскільки важливим є якомога менший порядок, беремо і далі будемо моделювати та прогнозувати саме першу різницю «Adj Close», яка на відміну від оригінального «Close» вже є стаціонарною (рис. 2.9).

	High	Low	Open	Close	Volume	Close_diff
Date						
2020-09-19	0.001810	0.001214	0.001535	0.001491	78421.0	-0.000043
2020-09-20	0.001784	0.001382	0.001492	0.001582	76880.0	0.000091
2020-09-21	0.001584	0.001185	0.001582	0.001288	91765.0	-0.000294
2020-10-02	0.000932	0.000781	0.000901	0.000858	115883.0	-0.000430
2020-10-03	0.001006	0.000813	0.000858	0.001006	79884.0	0.000148
...	...	...	...	...	...	...
2022-09-27	0.042246	0.039707	0.040853	0.040285	129931012.0	-0.000571
2022-09-28	0.040607	0.038653	0.040282	0.040172	106797618.0	-0.000113
2022-09-29	0.040925	0.039448	0.040176	0.040895	90196728.0	0.000723
2022-09-30	0.041101	0.040065	0.040889	0.040543	111533658.0	-0.000352
2022-10-01	0.040816	0.040048	0.040539	0.040258	61000313.0	-0.000285

Рисунок 2.9 – Результат обчислення першої різниці курсу криптовалюти Gala  
«Adj Close»

Насправді важливо, що можна використовувати «Close» як об'єктивний компонент для часових рядів, а «Close\_diff» — для зразкових моделей ШІ, наприклад, дерев вибору, організацій мозку тощо, які обробляють багато яскравих моментів.

Щоб перевірити, чи є ряд випадковим, використовується порядок `seasonal.seasonal_decompose` бібліотеки `Statsmodels.tsa` для розкладання ряду на частини, в тому числі - за нерегулярністю, використовуючи рухому нормаль [12]. Перевірка на нерегулярність полягає в розкладанні серії на візерунок «Візерунок», випадкову частину «Іноді» та решту. У випадку, якщо випадкова частина доступна і становить величезний рівень переваг серії, тоді припиняється її присутність і необхідність обчислювати елементи, які її враховують, і залучати моделі SARIMAX, а не ARIMA і інші для демонстрації часових рядів. Результат декомпозиції часового ряду курсу криптовалюти Gala на складові: на тренд «Trend», сезонну складову «Seasonal» та залишок «Resid» зображено на рисунку 2.10



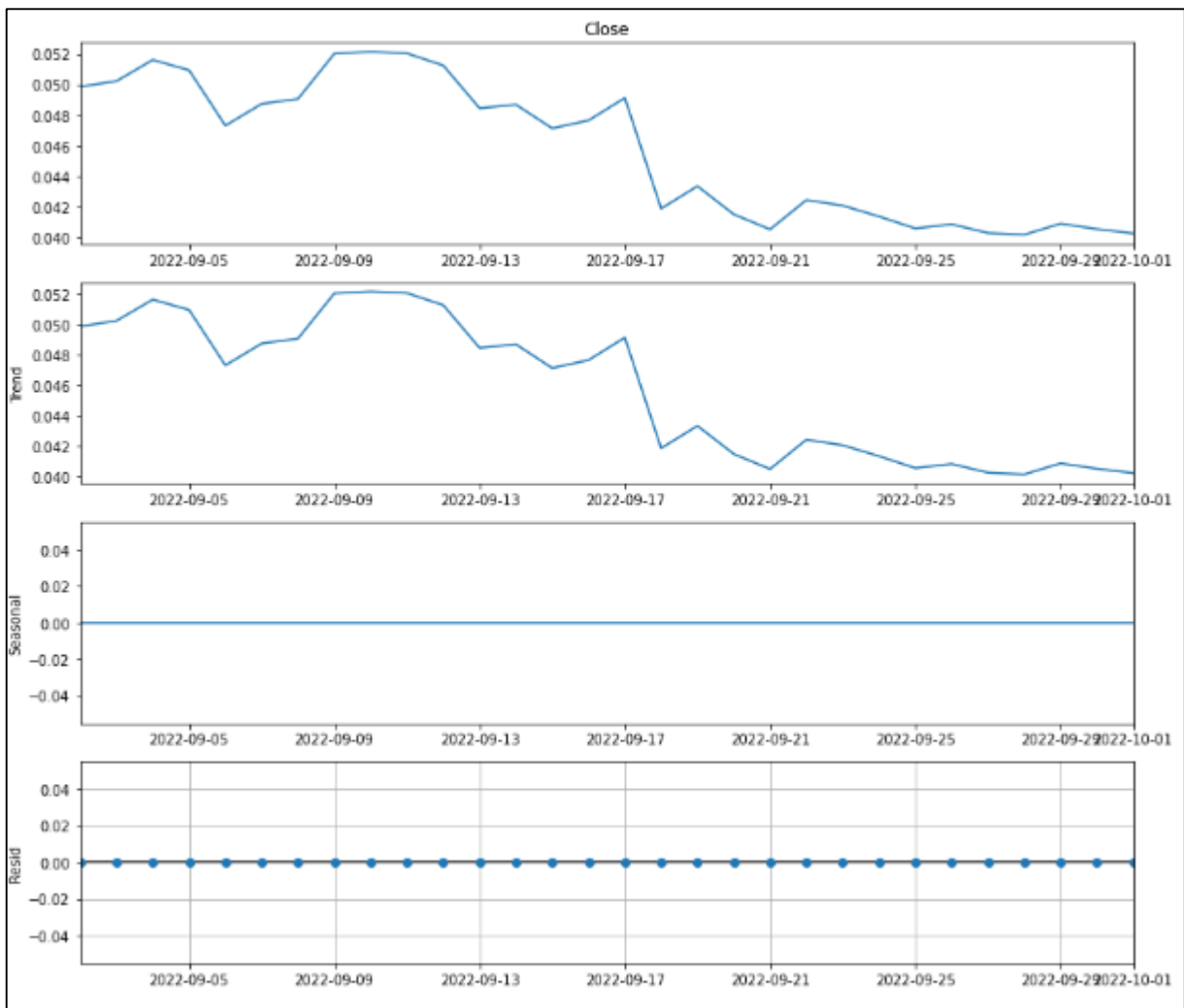


Рисунок 2.10 – Результат декомпозиції часового ряду курсу криптовалюти Gala на складові: на тренд «Trend», сезонну складову «Seasonal» та залишок «Resid»

На жаль, через величезну кількість інформації визначити конкретний час сезонності складно. Для цього ми розберемо і використаємо інформацію за останній місяць (рис. 2.11).

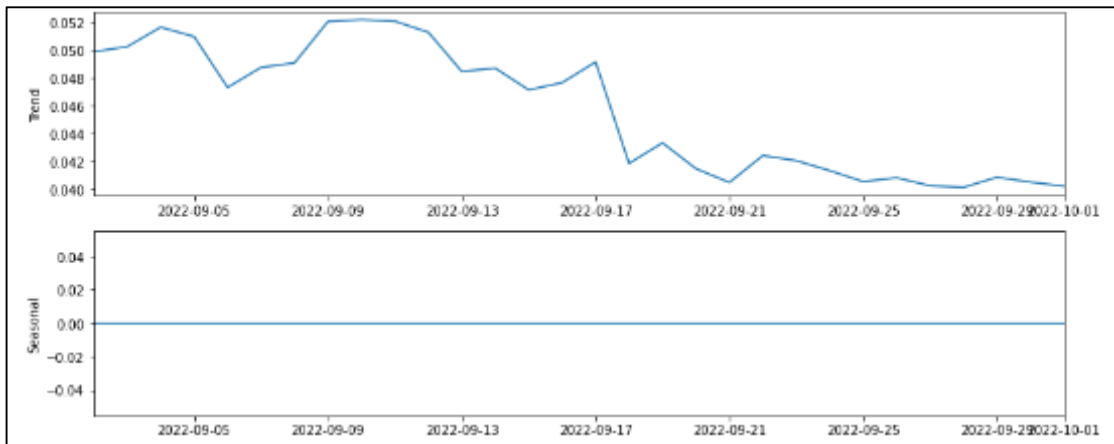


Рисунок 2.11 – Результат декомпозиції часового ряду курсу криптовалюти Gala за грудень 2022 року на складові тренд та сезонну складову

Із рисунку 2.11 стає зрозуміло, що будь-яка сезонність відсутня.

Проведений аналіз показав, що значення на рисунку 1.2 та основні закономірності суттєво відрізняються у 2021 і 2022 роках, тому для збільшення точності прогнозу пропонується будувати моделі тільки за даними 2022 року.

Проаналізуємо аномалії ринку. Аналіз періоду досліджень показав, що мав місце ряд аномалій (рис. 2.12).

	ds	lower_window	upper_window	prior_scale	holiday
0	2021-09-16	0	0	10	anomalous_dates
1	2021-11-29	0	0	10	anomalous_dates
2	2021-12-12	0	0	10	anomalous_dates
3	2021-12-19	0	0	10	anomalous_dates
4	2021-12-27	0	0	10	anomalous_dates
5	2022-01-22	0	0	10	anomalous_dates
6	2022-02-10	0	0	10	anomalous_dates
7	2022-02-14	0	0	10	anomalous_dates
8	2022-03-29	0	0	10	anomalous_dates
9	2022-05-04	0	0	10	anomalous_dates
10	2022-05-11	0	0	10	anomalous_dates
11	2022-08-14	0	0	10	anomalous_dates

Рисунок 2.12 – Перелік аномальних дати курсу криптовалюти

На рисунку 2.13 зображено аномалію 16.09.2021 р.

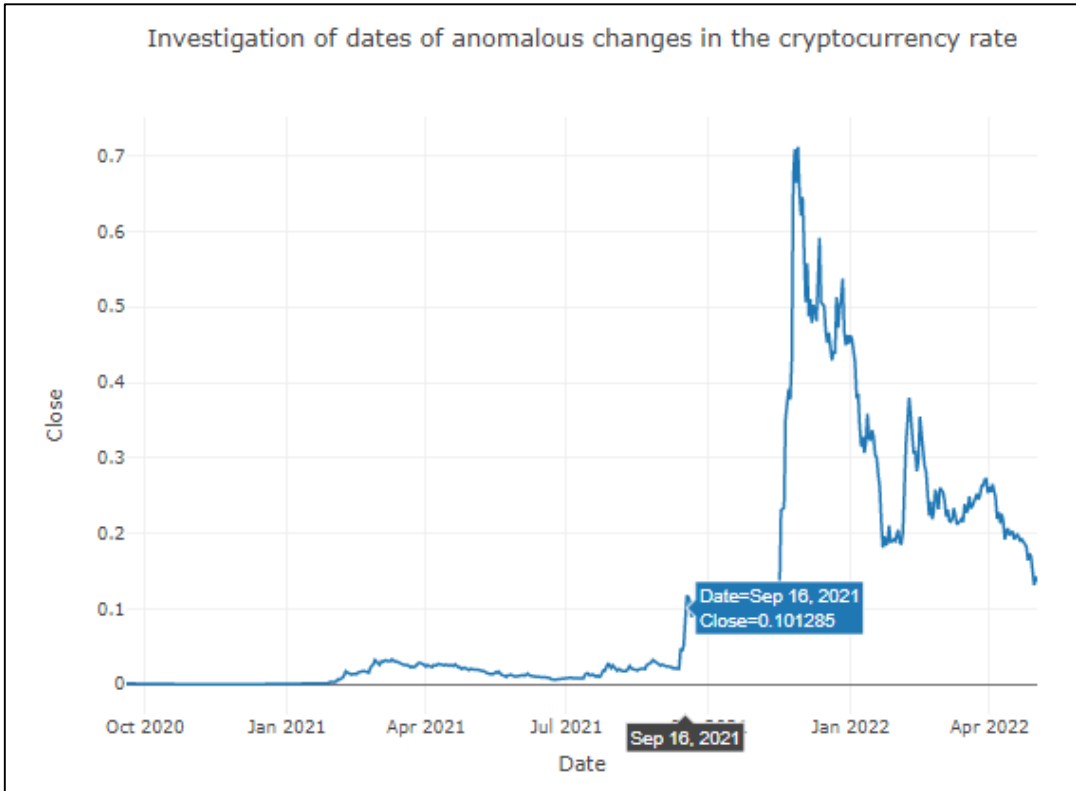


Рисунок 2.13 – Аномалія 16.09.2021 р. різкої зміни курсу криптовалюти

На рисунку 2.14 зображено аномалію 29.11.2021 р.

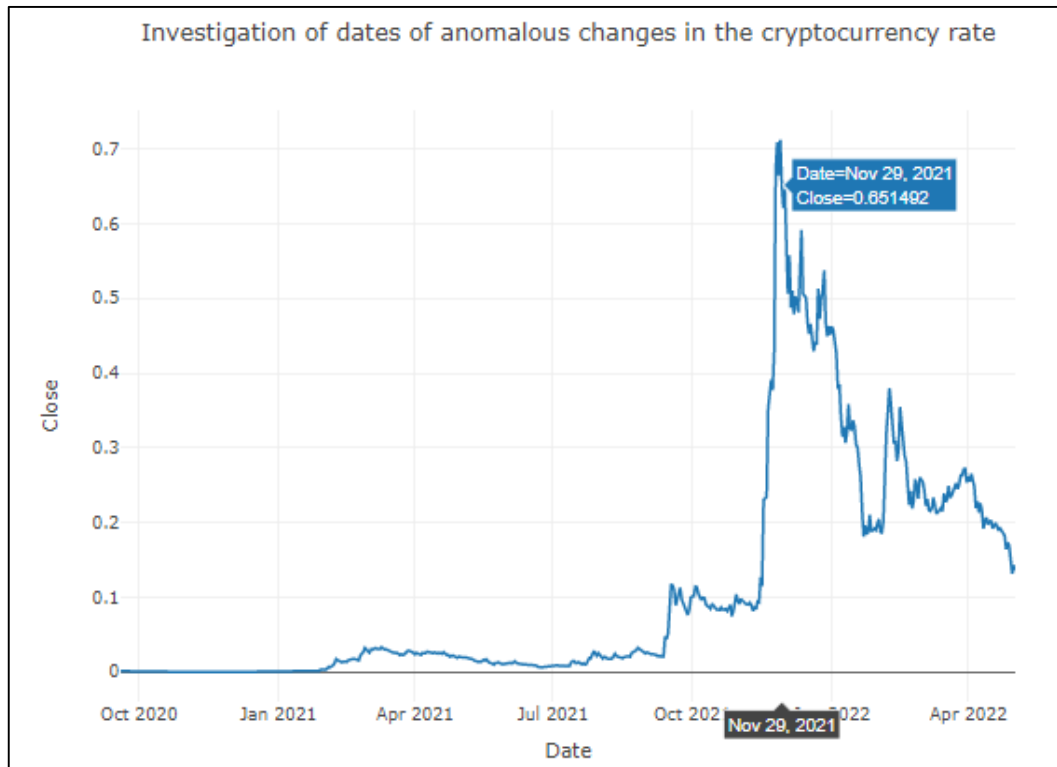


Рисунок 2.14 – Аномалія 29.11.2021 р. різкої зміни курсу криптовалюти

На рисунку 2.15 зображено аномалію 12.12.2021 р.

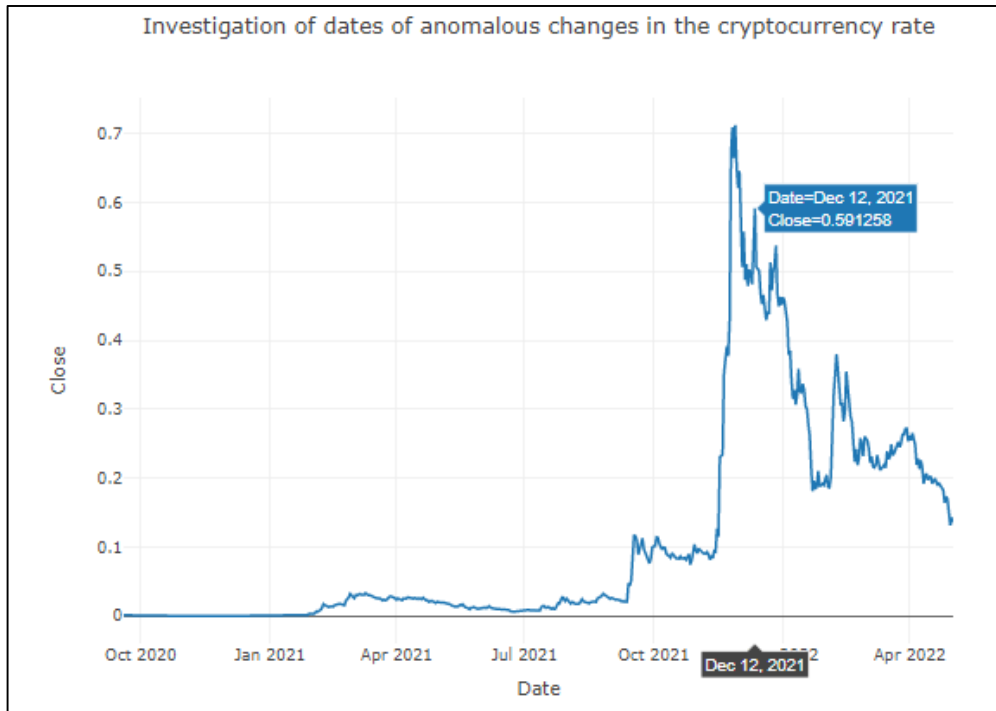


Рисунок 2.15 – Аномалія 12.12.2021 р. різкої зміни курсу криптовалюти

На рисунку 2.16 зображено аномалію 19.12.2021 р.

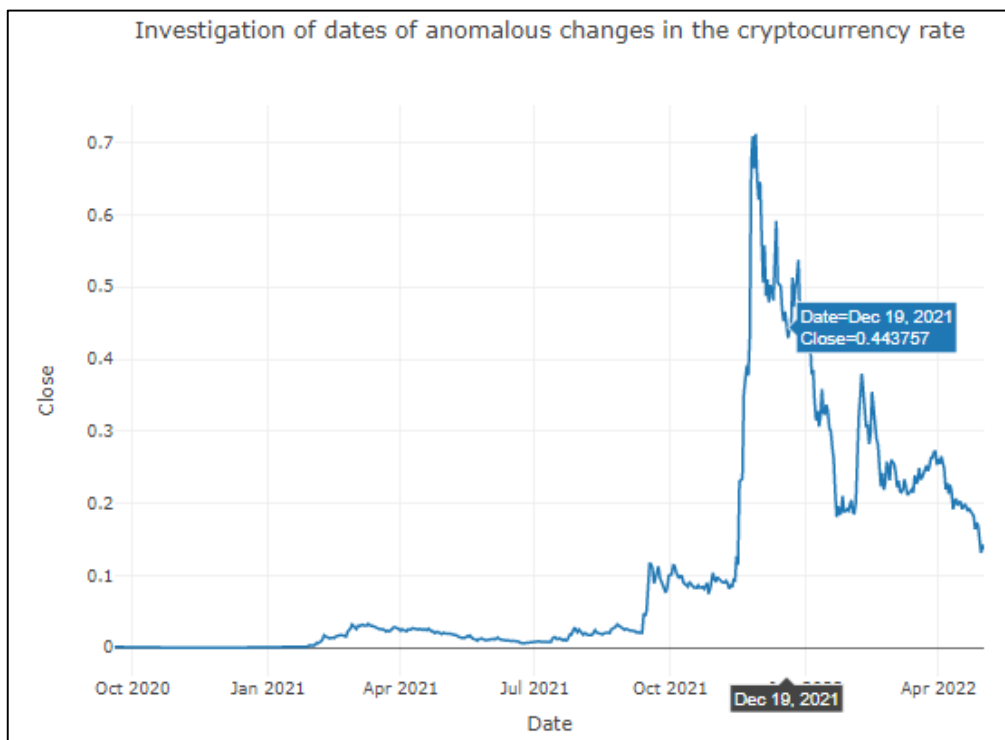


Рисунок 2.16 – Аномалія 19.12.2021 р. різкої зміни курсу криптовалюти

На рисунку 2.17 зображено аномалію 27.12.2021 р.

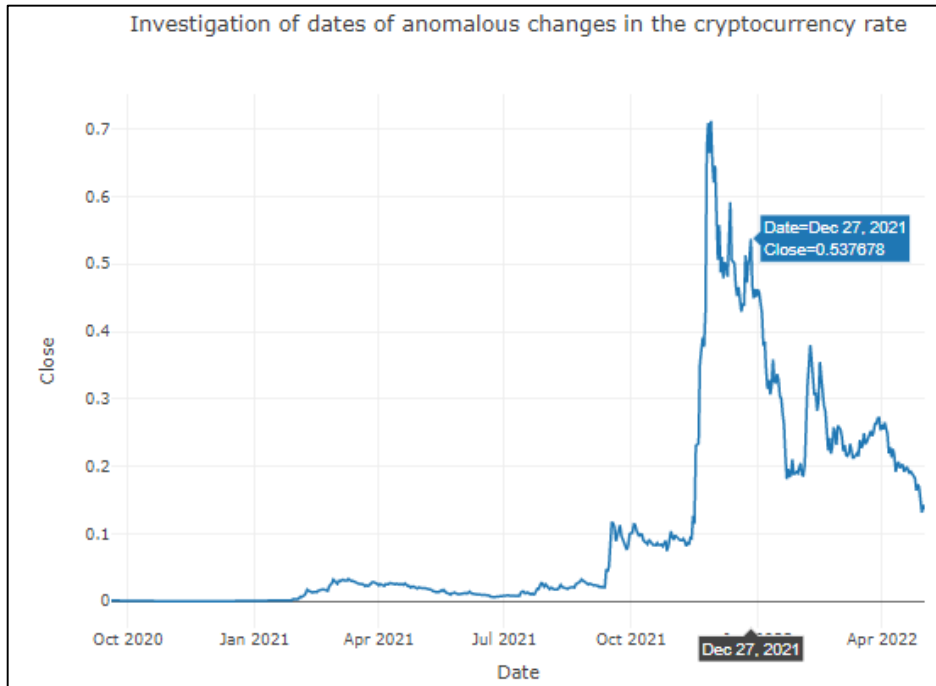


Рисунок 2.17 – Аномалія 27.12.2021 р. різкої зміни курсу криптовалюти

На рисунку 2.18 зображено аномалію 22.01.2022 р.

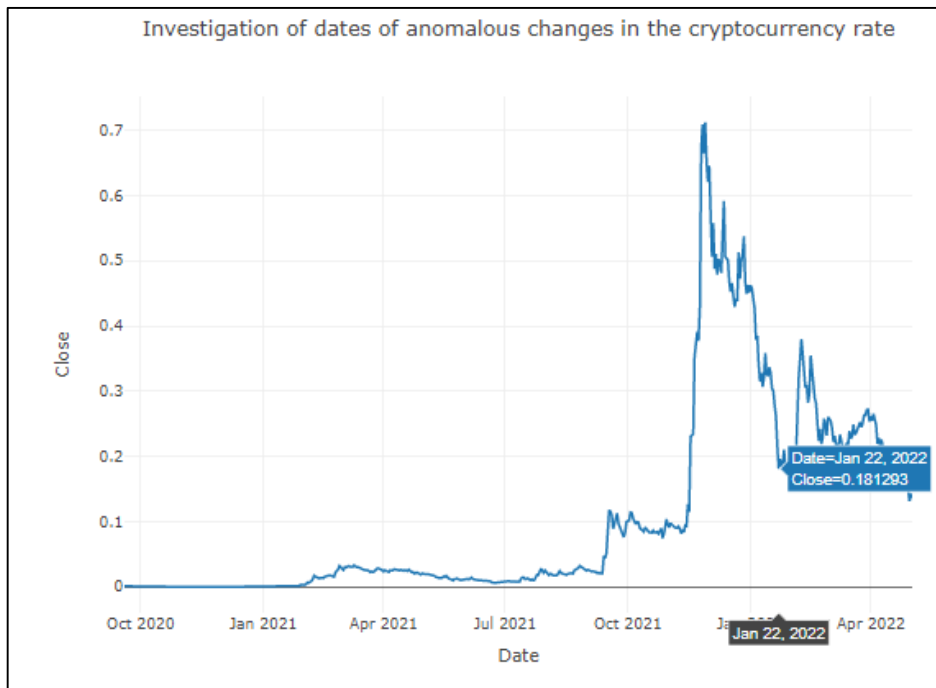


Рисунок 2.18 – Аномалія 22.01.2022 р. різкої зміни курсу криптовалюти

На рисунку 2.19 зображено аномалію 10.02.2022 р.

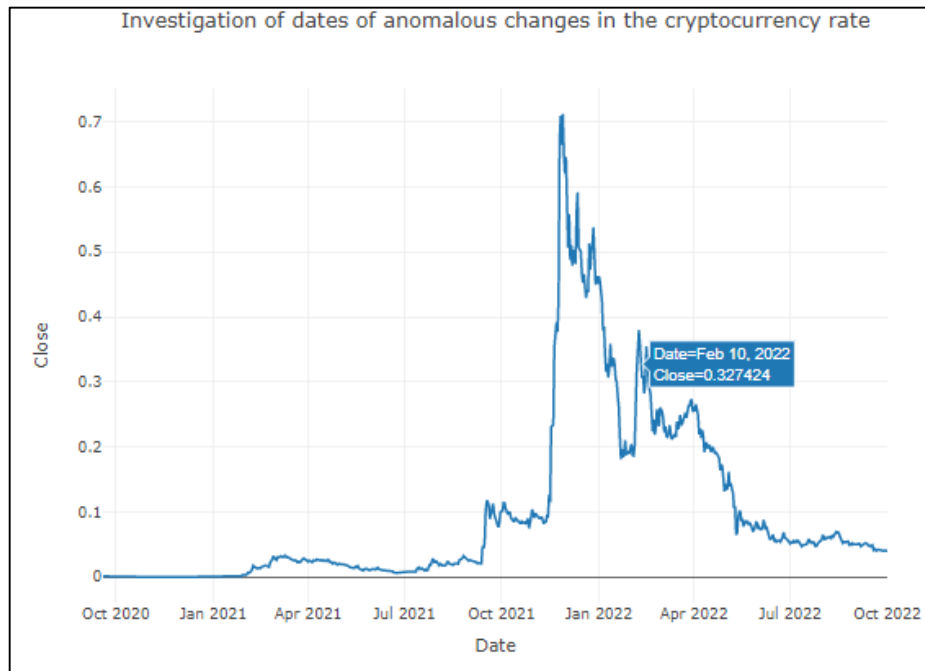


Рисунок 2.19 – Аномалія 10.02.2022 р. різкої зміни курсу криптовалюти

На рисунку 2.20 зображено аномалію 14.02.2022 р.

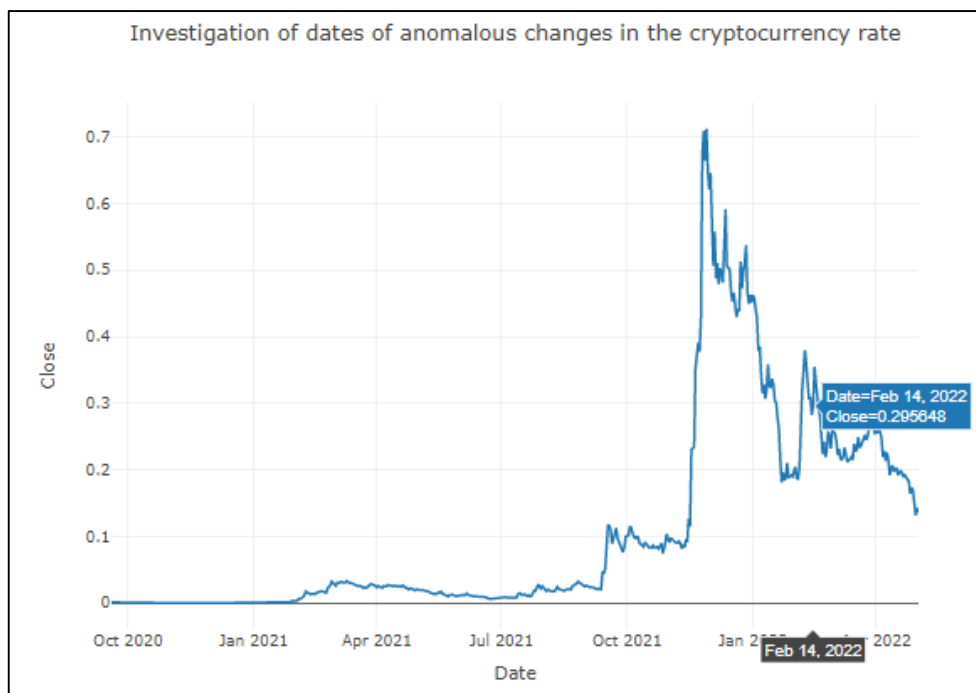


Рисунок 2.20 – Аномалія 14.02.2022 р. різкої зміни курсу криптовалюти

На рисунку 2.21 зображено аномалію 29.03.2022 р.

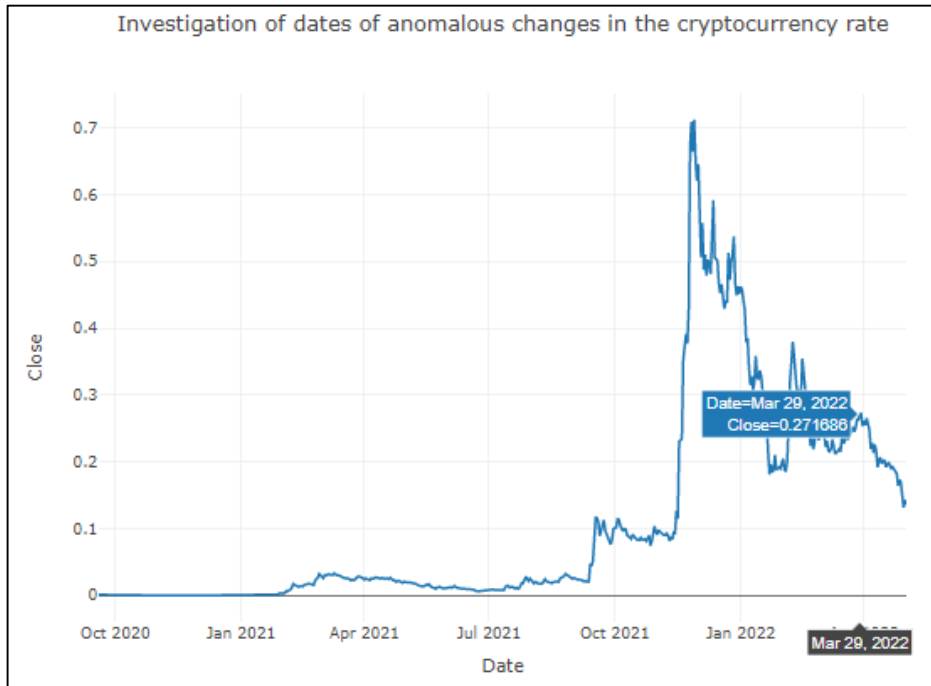


Рисунок 2.21 – Аномалія 29.03.2022 р. різкої зміни курсу криптовалюти

На рисунку 2.22 зображено аномалію 04.05.2022 р.

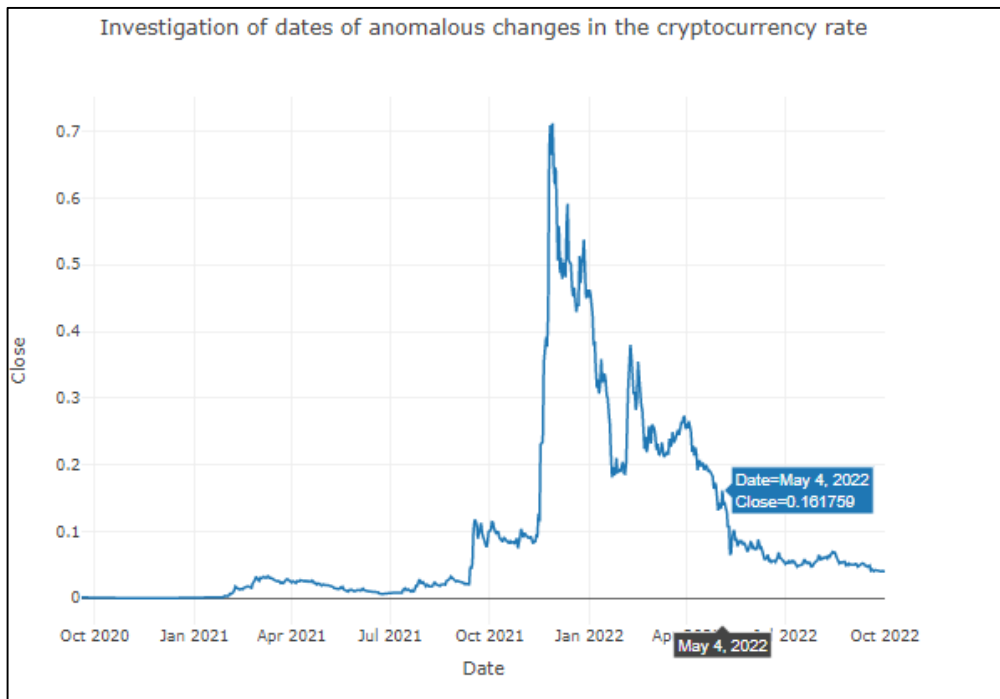


Рисунок 2.22 – Аномалія 04.05.2022 р. різкої зміни курсу криптовалюти

На рисунку 2.23 зображено аномалію 11.05.2022 р.



Рисунок 2.23 – Аномалія 11.05.2022 р. різкої зміни курсу криптовалюти

На рисунку 2.24 зображено аномалію 14.08.2022 р.

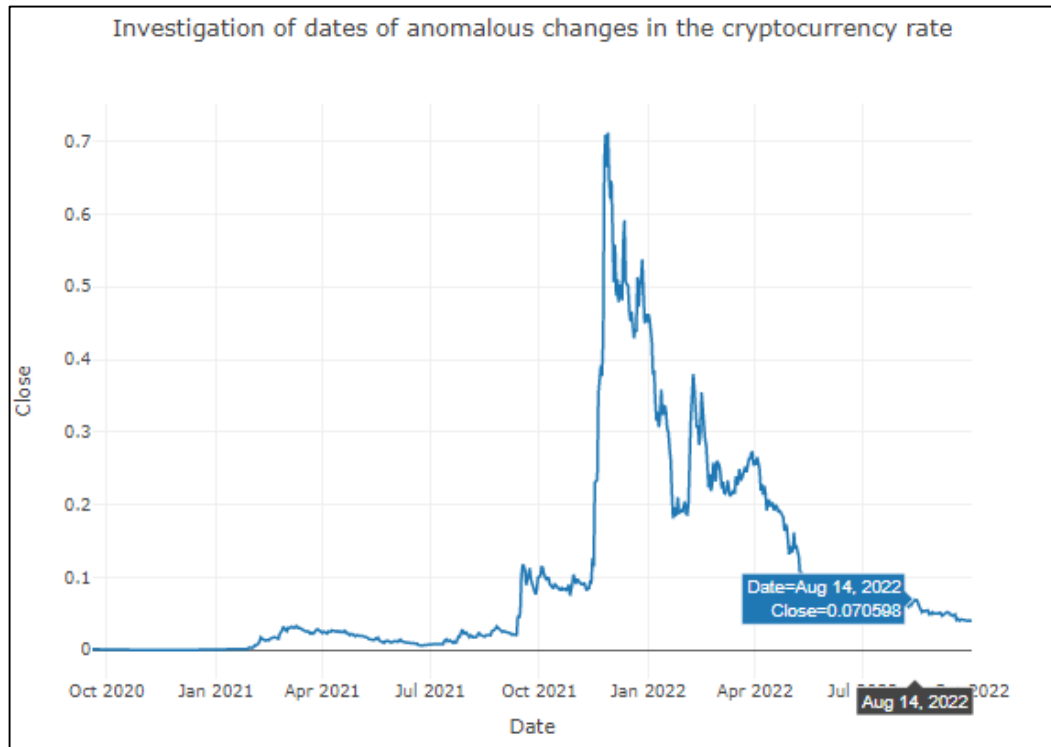


Рисунок 2.24 – Аномалія 14.08.2022 р. різкої зміни курсу криптовалюти



Можливі причини таких змін курсу –:

– 16.09.2021 індекс Dow Jones Industrial Average з блакитними фішками втратив понад 1000 пунктів, а Nasdaq впав на 5%, що стало найгіршим падінням за один день з 2021 року і це викликало падіння гали на 5% до 0,012 доларів США о 7:12 ранку за східним часом [2];

– 29.11.2021 найпопулярніша криптовалюта у світі Gala у понеділок, 29 листопада, збільшилася у вартості на майже 6% на тлі обнадійливих новин щодо нового коронавірусного штаму Omicron. Це призвело до приросту курсу до 0,83 доларів США [4];

– 12.12.2021 на тлі зміцнення індексів S&P 500 та Nasdaq інвестори починають знову повертатися до криптовалютних активів. Втім, капіталізація криптовалютного ринку за останній день знизилася на 11% до 1,21 трлн доларів, а індекс домінування біткоїну склав 45%. Це призвело до падіння Gala до 0,48 доларів США [7];

– 19.12.2021 Blockchain.com повідомив, що 90% існуючих біткойнів вже видобуто. У світі налічується близько 21 мільйона монет, з яких 18,9 перебувають у обігу. Це число було досягнуто лише через 12 років після покупки першої криптовалюти у січні 2009 року. Це призвело до падіння Gala до 0,46 доларів США [8];

– 27.12.2021 Tesla продала близько 10% активів BTC. Ілон Маск пояснив це бажанням продемонструвати ліквідність монети. Це призвело до приросту курсу Gala до 0,55 доларів США [8];

– 22.01.2022 криптовалюти продовжують втрачати ціну, головна криптовалюта світу Gala вже впала на 9%, а ціна Ethereum знизилася на 14%також знизився курс гали до 0,22 доларів США [9];

– 10.02.2022 NFT знову у тренді: обсяг торгів на OpenSea наблизився до рекордних \$5 млрд в свою чергу це призвело до приросту курсу гали до 0,31 доларів США [10];

– 14.02 2022 Litecoin здешевшав на 2,30 доларів США це і призвело до падіння курсу гали до 0,29 доларів США [10];

- 29.03.2022 компанія Luna Foundation інвестувала у Gala понад \$1 млрд. Це і призвело до приросту курсу гали до 0,28 доларів США [11];
- 04.05.2022 курс Біткоїна виріс до \$20116,50 так як ці криптовалюти пов'язані то це призвело до збільшення курсу гали до 0,26 доларів США [3];
- 11.05.2022 Китай запровадив заборону на всі операції з криптовалютою, визнавши незаконними усі транзакції, пов'язані з криптоактивами. Це спричинило суттєве зниження курсів низки криптовалют, а також Gala до 0,45 доларів США [9];
- 14.08.2021 разом із Gala падають й інші криптовалюти, причому деякі падають навіть сильніше. Наприклад, курс Ethereum за добу впав на 12,5%, Gala подешевшала на 15%, Terra – майже на 21,5%. Найменше втрачають Dogecoin (близько 7%), XRP (близько 9%), Cardano (10,5%), також впала Gala до 0,55 доларів США [5].

На рисунку 2.25 зображений створений датасет із дат аномальних змін, який потрібно використати протягом прогнозування курсу валюти.

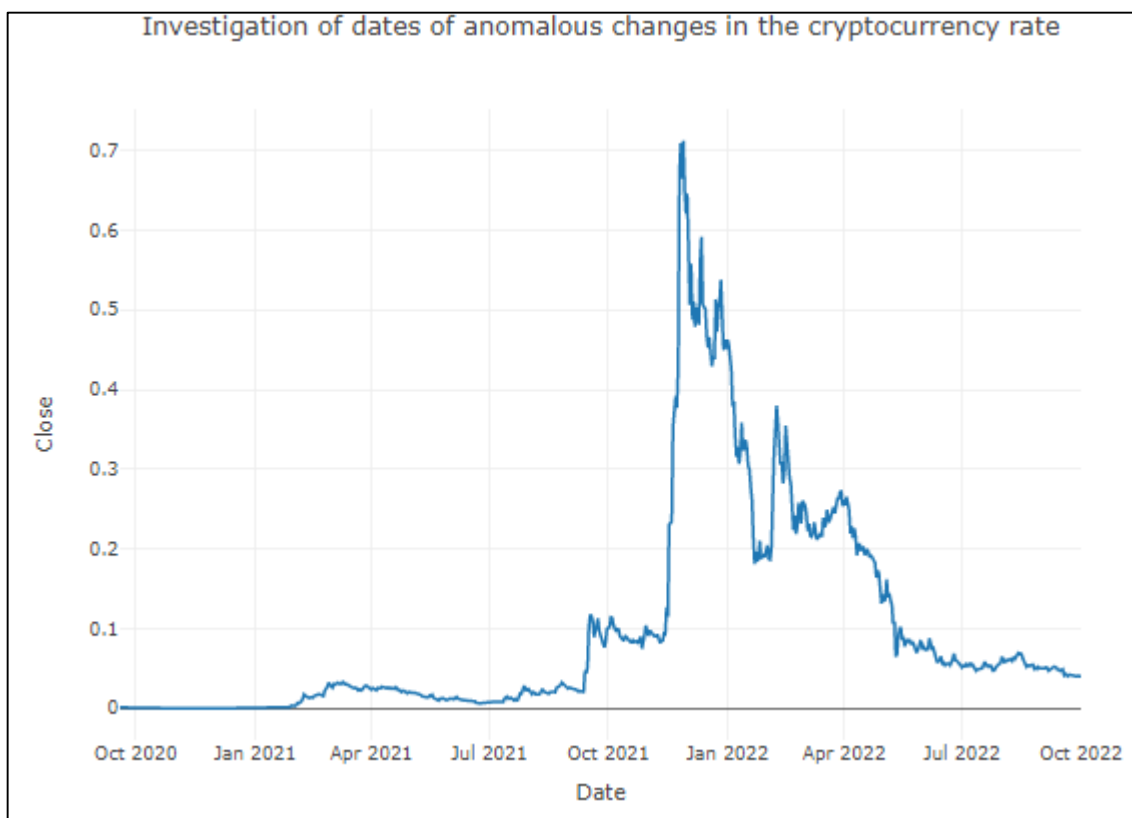


Рисунок 2.25 – Дати аномальних змін курсу криптовалюти та їх параметри

На рисунку 2.26 наведено комбінований графік курсу криптовалюти та аномальних дат з рисунку 2.12

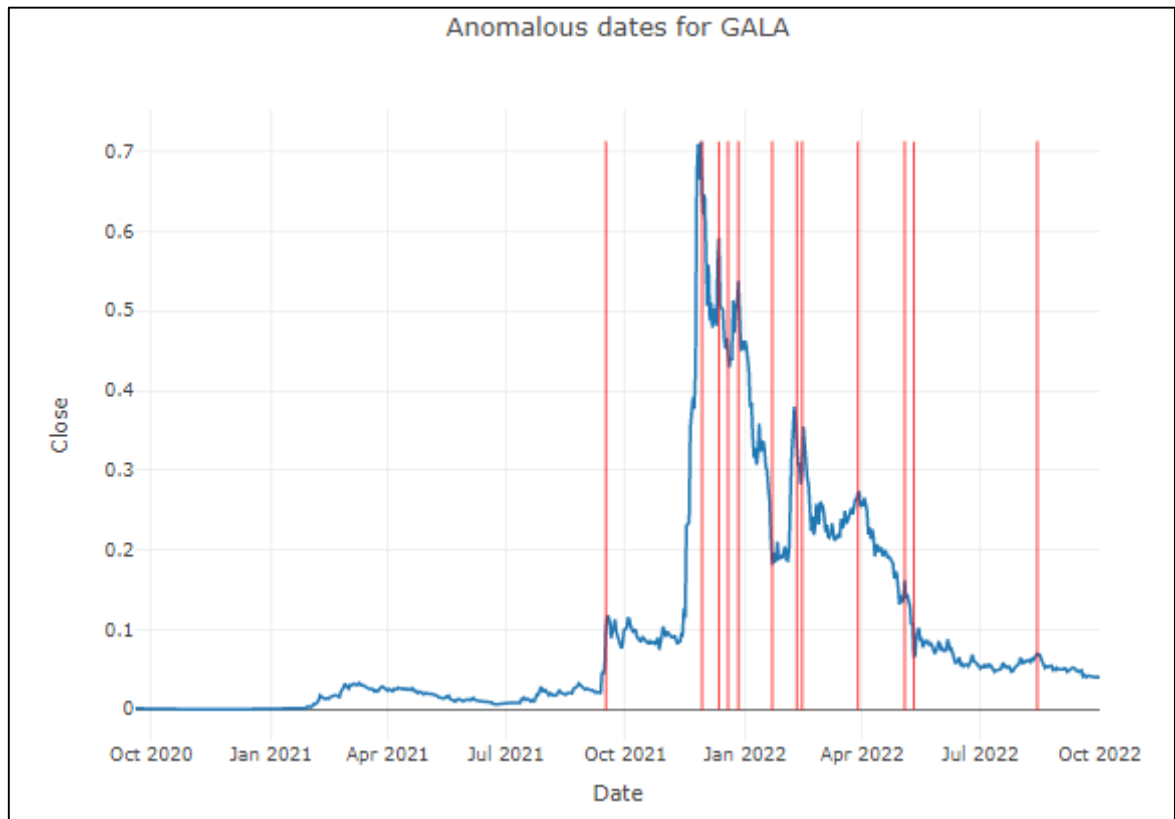


Рисунок 2.26 – Курс криптовалюти (синя товста лінія) та дати аномальних змін (тонкі вертикальні лінії)

На рисунку 2.27 зображено аномальні дати для першої різниці курсу криптовалюти Gala

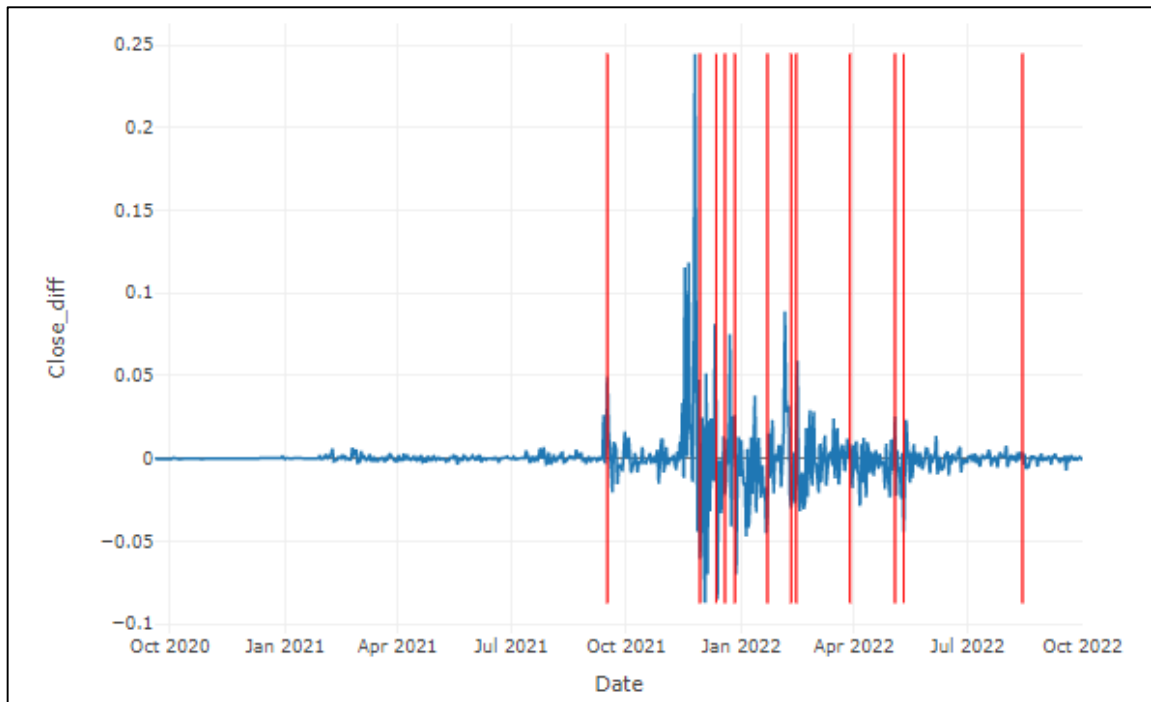


Рисунок 2.27 – Аномальні дати для першої різниці курсу криптовалюти

### 2.3 Синтез нових ознак

Базових ознак для моделювання, зазвичай недостатньо для прогнозування. Тому усі науковці синтезують низку інших. Методика аналізу і прогнозування валютного ринку включає систему генерації сигналів торгової системи (купити, продати, утримувати, закрити поточну позицію).

Для генерації та ідентифікації сигналів, використовується дослідження точки перетину кількох видатних рухомих середніх точок з різними періодами усереднення. Експериментально встановлені правила відкриття позицій на купівлю/продаж обмежують зростання типової кількості всіх поданих сигналів. Основою для закриття позиції є рівень вигоди / збитку, що задається експертом.

Порівняння ефективності методики на різних частотах тимчасового ряду показує, що найбільша ефективність досягається на часових даних. Недоліком є використання ковзних середніх, відомо, що вони часто дають помилкові сигнали, особливо у флеті, запізнювання сигналу під час тренду і розвороту тенденції, що призводить до втрати прибутку. Перевагою методики є простота використання,

недоліком – неточність сигналів опису ринкової ситуації, низька ефективність при бічному тренді.

Моделі прогнозування прибутковості валют (і портфеля інвестора) на ВР FoRex – побудовані на основі виділення періодичних компонент ряду прибутковості методами сингулярного спектрального аналізу. Аналізується ряд прибутковості у вигляді логарифмічних різниць валютного курсу. Для визначення періодичних компонент і їх характеристик застосовується метод 12 передбачення головних компонент (PCLP) Д. Тафтса і Р. Кумаресана. Розрахунки по запропонованим моделям відповідно до розроблених методиками є обчислювальною задачею великої місткості. Зазначений недолік обмежує можливість використання даного підходу для активної стратегії трейдера на ринку Forex в швидко мінливих умовах.

Методика побудови ефективних трендових торгових стратегій з використанням індикатора і осциляторів RSI, Stochastic – моделі являють собою рівняння регресії, в яких керуюча змінна (КЗ) представляється лінійною функцією деяких характеристик осциляторів. Значення КЗ інтерпретуються як критерії достовірності сигналів на вхід в ринок (вихід з ринку).

У торговельній стратегії для визначення поточної (довгострокової) тенденції використовується технологія «потрійного вибору» і індикатор RSI, тобто ідентифікація поточного тренду здійснюється поза моделлю. Отримано чотири моделі для чотирьох стратегій. Результати апробації моделей на часових даних курсу GBP / USD для всіх чотирьох стратегій показують, що додавання керуючої змінної (моделі) до стандартних стратегій (без моделі) практично не змінює ситуацію. Методи фрактального аналізу і апарат теорії нечітких множин – використовуються у дослідженні часового ряду валютного курсу пари євродолар і ряду його збільшень з метою виявлення довготривалої пам'яті, її глибини, трендостійкості, циклів.

Для визначення рівня ризику використовується фрактальна характеристика тимчасового ряду – показник Херста. Результати фазового аналізу часових рядів підтверджують наявність довгострокової пам'яті, яка

пояснюється присутністю циклів. Запропонована прогнозна модель, в якій вихідний числовий ряд збільшень валютного курсу перетворюється в лінгвістичний шляхом заміни числових значень термами (лінгвістичними змінними) з урахуванням глибини пам'яті досліджуваного ряду. 13 Результат прогнозу (величина приросту валютного курсу на день) представляється у вигляді нечіткої множини, який потім трансформується в числовий еквівалент за допомогою процедури дефазифікації.

Модель і метод прогнозування призначені для предпрогнозного етапу дослідження ВР. Область застосування отриманих результатів – короткострокове прогнозування критичних тенденцій на валютному ринку в якості додаткового інструменту трейдера.

Методика факторно-регресійного аналізу – відслідковує вплив агрегованих факторів на номінальні і «інваріантно-індексні» курси валют, за допомогою чого розроблені нові економетричні моделі регресії мінової цінності валюти на виділену групу агрегованих факторів. Також запропоноване використання у дослідженнях економіко-математичних моделей впливу макроекономічних факторів на валютні курси, в основу яких покладено поняття індексу мінової цінності валюти.

Показники інваріантних індексів мінової цінності валюти (адитивний і мультиплікативний) використані для модифікації стандартних (структурних) моделей для визначення рівноважного валютного курсу.

Бібліотека TSFRESH дозволяє автоматично синтезувати біля тисячі ознак для заданого часового ряду: середнє, дисперсія, ексцес за тиждень, місяць.

На рисунках 2.28 – 2.32 наведено результат застосування цієї бібліотеки до ознаки «Close» після застосування фільтру, який видаляє ознаки без значень (зі значеннями NaN) та ознаки, середньоквадратичне 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
count	714.000000	714.000000	714.000000	714.000000	714.000000	714.000000
mean	356.500000	169575.166667	356.500000	356.500000	356.500000	356.500000
std	206.258333	151898.463652	206.258333	206.258333	206.258333	206.258333
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	178.250000	31773.250000	178.250000	178.250000	178.250000	178.250000
50%	356.500000	127092.500000	356.500000	356.500000	356.500000	356.500000
75%	534.750000	285957.750000	534.750000	534.750000	534.750000	534.750000
max	713.000000	508369.000000	713.000000	713.000000	713.000000	713.000000

Рисунок 2.28 – Нові ознаки для ознаки «Close» криптовалюти Gala, згенеровані бібліотекою TSFRESH

index_absolute_maximum	index_minimum	index_benford_correlation	index_quantile_q_0.1	...	Close_quantile_q_0.6
714.000000	714.000000	713.000000	714.000000	...	714.000000
356.500000	356.500000	0.109119	356.500000	...	0.101499
206.258333	206.258333	0.365425	206.258333	...	0.137776
0.000000	0.000000	-0.297356	0.000000	...	0.000205
178.250000	178.250000	-0.145280	178.250000	...	0.013596
356.500000	356.500000	-0.064614	356.500000	...	0.048949
534.750000	534.750000	0.295657	534.750000	...	0.113438
713.000000	713.000000	0.864123	713.000000	...	0.712089

Рисунок 2.29 – Нові ознаки для ознаки «Close» криптовалюти Gala, згенеровані бібліотекою TSFRESH (продовження)

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)
714.000000	714.000000	714.000000	714.000000
0.101499	0.101499	0.101499	0.062248
0.137776	0.137776	0.137776	0.084497
0.000205	0.000205	0.000205	0.000126
0.013596	0.013596	0.013596	0.008338
0.048949	0.048949	0.048949	0.030020
0.113438	0.113438	0.113438	0.069570
0.712089	0.712089	0.712089	0.436718

Рисунок 2.30 – Нові ознаки для ознаки «Close» криптовалюти Gala, згенеровані бібліотекою TSFRESH (продовження)

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)
714.000000	714.000000	714.000000
0.039369	0.027838	0.019685
0.053441	0.037788	0.026720
0.000080	0.000056	0.000040
0.005274	0.003729	0.002637
0.018986	0.013425	0.009493
0.044000	0.031113	0.022000
0.276205	0.195306	0.138102

Рисунок 2.31 – Нові ознаки для ознаки «Close» криптовалюти Gala, згенеровані бібліотекою TSFRESH (продовження)

Close_fft_coefficient_attr_“real”_coeff_0	Close_fft_coefficient_attr_“abs”_coeff_0
714.000000	714.000000
0.101499	0.101499
0.137776	0.137776
0.000205	0.000205
0.013596	0.013596
0.048949	0.048949
0.113438	0.113438
0.712089	0.712089

Рисунок 2.32 – Нові ознаки для ознаки «Close» криптовалюти Gala, згенеровані бібліотекою TSFRESH (продовження)

Таким чином, відібрано 561 ознак на основі базової інформації крипторинків. Розглянемо зовнішні фактори впливу.

## 2.4 Побудова моделей методами машинного навчання

Побудовані тренувальний, валідаційний та тестовий датасети з усіма ознаками курсу криптовалюти Gala за даними 2022 року наведено на рисунку 2.33.



```

Origin dataset has 340 rows and 60 features
Get training dataset with 330 rows
Get validation dataset with 5 rows
Get test dataset with 5 rows

```

Рисунок 2.33 – Розмір побудованих тренувального, валідаційного та тестового датасетів з усіма ознаками курсу криптовалюти Gala за даними 2022 року

Побудуємо різні моделі, використовуючи методи штучного інтелекту. Використовуємо елементи, вибрані раніше. Як щодо нормалізації їх за допомогою стратегії "sklearn.preprocessing.StandardScaler".

Ми визначаємо моделі які зображено на рисунку 2.34

	name_model	type_data	r2_score	rmse	mape
9	Linear Regression	valid	0.543857	0.030552	5.171395
12	Linear SVR	valid	0.356824	0.036279	6.867392
0	Prophet_4_days_3_order	valid	-0.657371	0.058237	11.300432
4	Prophet_7_days_3_order	valid	-1.084688	0.065314	12.102852
2	Prophet_5_days_3_order	valid	-1.060643	0.064937	12.846515
7	Prophet_14_days_12_order	valid	-1.544065	0.072153	14.056655
3	Prophet_5_days_12_order	valid	-1.600915	0.072954	14.676709
6	Prophet_14_days_3_order	valid	-1.71488	0.074536	14.912129
5	Prophet_7_days_12_order	valid	-2.080754	0.079399	15.474591
1	Prophet_4_days_12_order	valid	-2.035468	0.078814	15.893748
16	MLP Regressor	valid	-5.044017	0.111212	22.385564
10	KNeighbors Regressor	valid	-7.391326	0.13104	26.252296
13	Random Forest Regressor	valid	-10.13803	0.150971	30.008292
8	ARIMA_auto	valid	-13.441008	0.171904	34.138191
15	XGB Regressor	valid	-31.958058	0.259698	50.787187
14	Bagging Regressor	valid	-37.101729	0.279229	55.8096
11	Support Vector Machines	valid	-69.607196	0.380113	73.92856

Рисунок 2.34 – Моделі машинного навчання та їх параметри

## 2.5 Вибір оптимальної моделі

Після обчислення усіх моделей на валідаційних даних здійснюється порівняння прогнозів з відомими цільовими значеннями «Close» у валідаційному датасеті за 3-ма метриками:  $r2\_score$ , RMSE, MSE.

Середня квадратична помилка (Mean Squared Error, MSE):

$$MSE = \frac{1}{N} \sum_{i=1}^N e_i^2 \quad (2.1)$$

Дана метрика враховує наявність рідкісних екстремальних помилок.

RMSE (Root Mean Squared Error):

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2} \quad (2.2)$$

Дана метрика має схожі з MSE характеристиками. Результат наведено на рисунку 2.37.

	name_model	r2_score	rmse	mape	params
9	Linear Regression	0.543857	0.030552	5.171395	{'fit_intercept': False}

Рисунок 2.37 – Похибки моделі Linear Regression, побудованих для багатьох ознак для валідаційних даних курсу криптовалюти

Як видно, похибка оптимальної моделі за усіма трьома метриками «Linear Regression» має низьку відносну похибку, порівняно малу похибку RMSE і допустиму точність за метрикою r2\_score, що означає, що самі значення прогножуються достатньо точно, але напрямок зростання чи падіння курсу прогнозується доволі непогано (коефіцієнт кореляції r2\_score є додатнім і становить 0.54). Отже, оптимальна модель є гарною також за метрикою MAPE становить 5.17%, тобто за відносною похибкою.

## 2.6 Висновки

У другому розділі зроблено розвідувальний аналіз даних, здійснений і описаний синтез нових ознак за яким відібрано 561 ознаку на основі базової інформації крипторинків, досліджено 12 аномальних дат які вагомо впливають

на подальше прогнозування курсу криптовалюти, побудована модель машинного навчання, також була вибрана оптимальна модель, а саме «Linear Regression», з'ясовано, що похибка оптимальної моделі за метрикою MAPE становить 5.17%, за метрикою RMSE становить 0,03 USD що означає, що самі значення прогнозуються достатньо точно.

## **3 РОЗРОБЛЕННЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ КУРСУ КРИПТОВАЛЮТИ GALA У 2021-2022 РОКАХ**

### **3.1 Розроблення інформаційної технології прогнозування курсу Gala**

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

- Аналіз даних та синтез ознак;
- Формування цільової ознаки;
- Передоброблення даних – здійснити стандартизацію ознак методом «sklearn.preprocessing.StandardScaler»;
- Формування тренувального, валідаційного та тестових датасетів для різних типів моделей;
- Вибір і тренування моделей на тренувальному датасеті (train).  
Прогнозування за кожною ідентифікованою моделлю валідаційних даних (valid);
- Вибір оптимальної моделі;
- Тренування оптимальної моделі за розширеним датасетом (train+valid).  
Прогнозування за ідентифікованою моделлю тестових даних (test);
- Оцінювання точності прогнозу тестових даних.

OHLC-діаграма криптовалюти Gala за даними сайту Yahoo зображено на рисунку 3.1

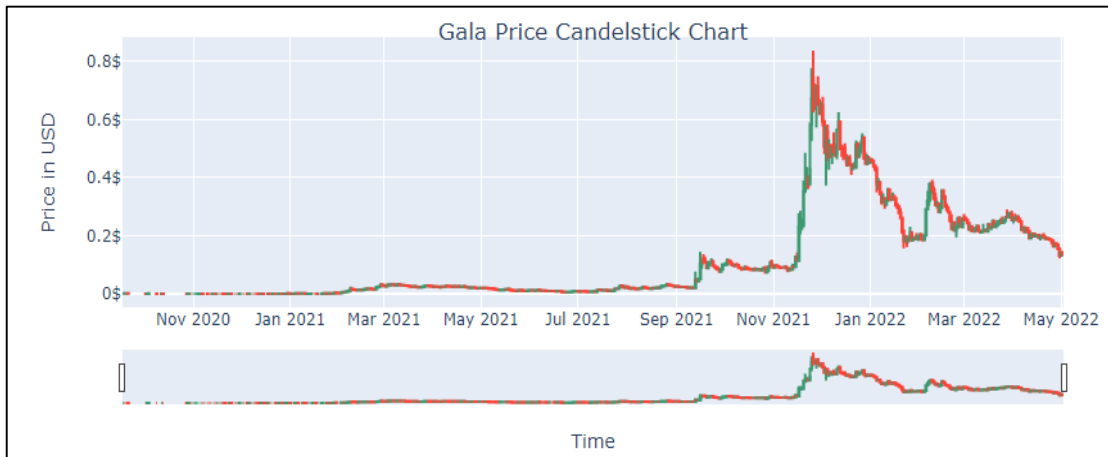


Рисунок 3.1 – OHLC-діаграма криптовалюти Gala за даними сайту Yahoo

Перелік синтезу ознак зображено на рисунку 3.2 – 3.3.

```
[ '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)

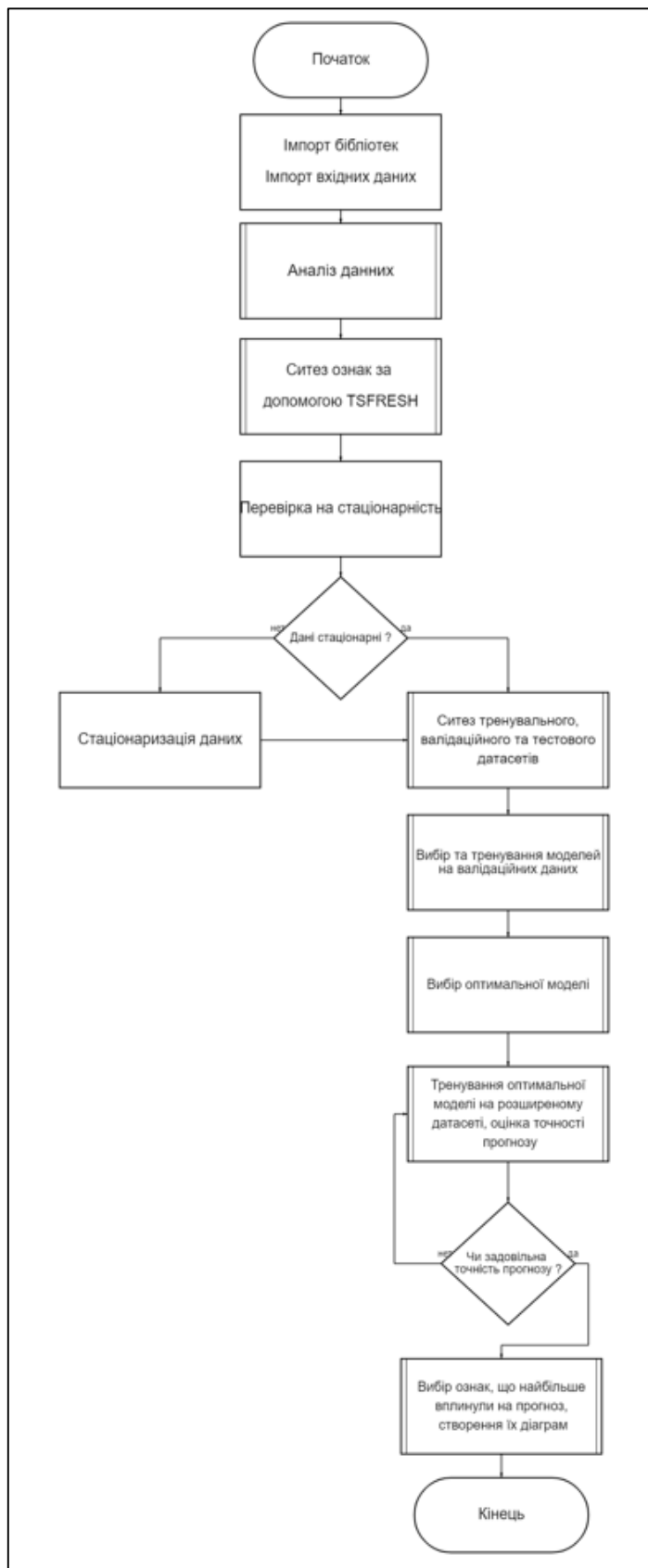


Рисунок 3.4 – Блок-схема алгоритму інформаційної технології прогнозування курсу Gala у 2021-2022 роках

### 3.2 Прогнозування тестових даних

Вибрана у розділі 2 оптимальна модель «Linear Regression» була заново ідентифікована на вже на розширеному датасеті «train+valid», після чого було зроблено прогнозування вже тестових даних та порівняння з реальними даними (рис. 3.5 - 3.8).

```
Optimal model by metrics "mape" is "Linear Regression" with type "Linear Regression" parameters {'fit_intercept': False}
Tuning model 'Linear Regression'
```

Рисунок 3.5 – Похибки і оптимальні параметри моделі Linear Regression оптимальної за метрикою  $r2\_score$

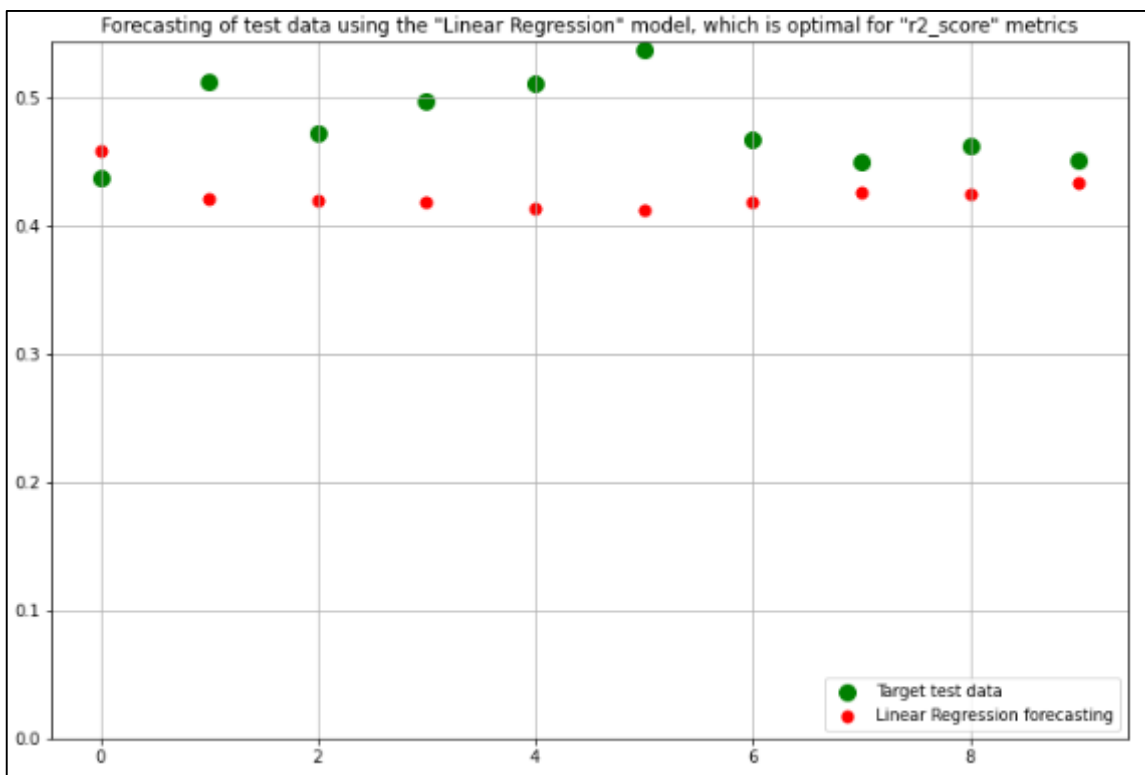


Рисунок 3.6 – Реальні тестові дані курсу криптовалюти Gala та їх прогноз за моделлю Linear Regression, оптимальної за метрикою  $r2\_score$



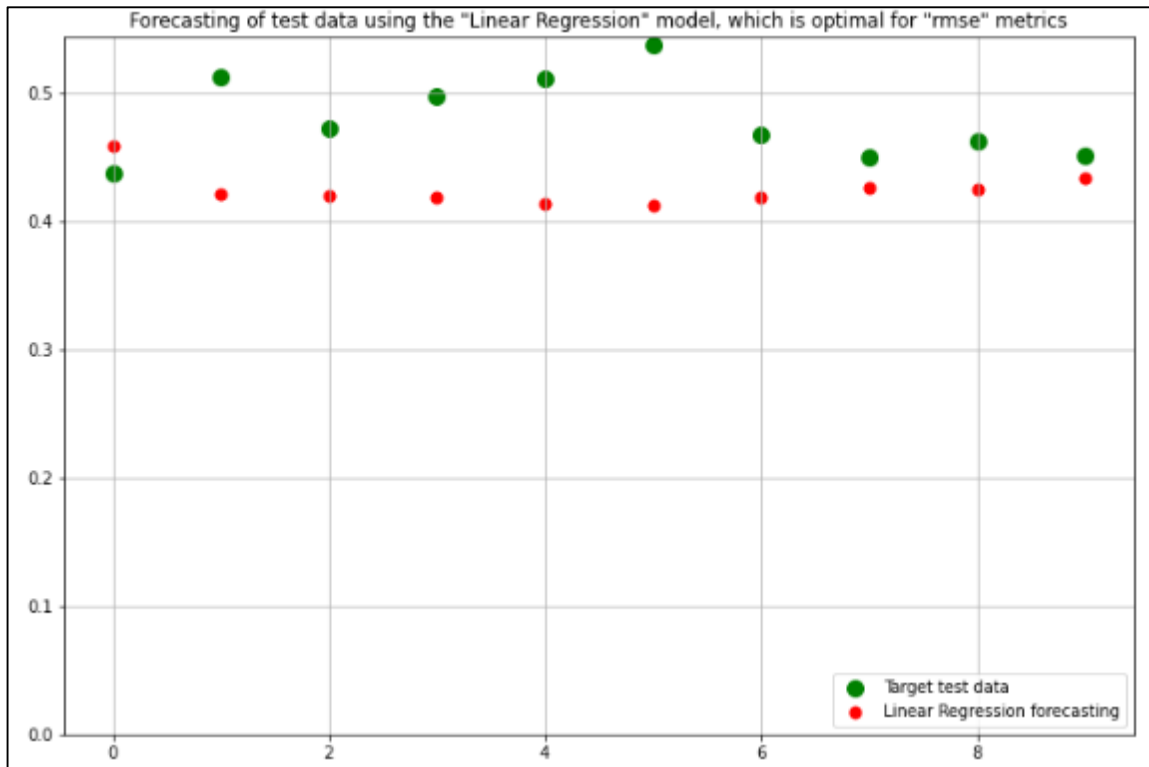


Рисунок 3.7 – Реальні тестові дані курсу криптовалюти Gala та їх прогноз за моделлю Linear Regression, оптимальної за метрикою rmse

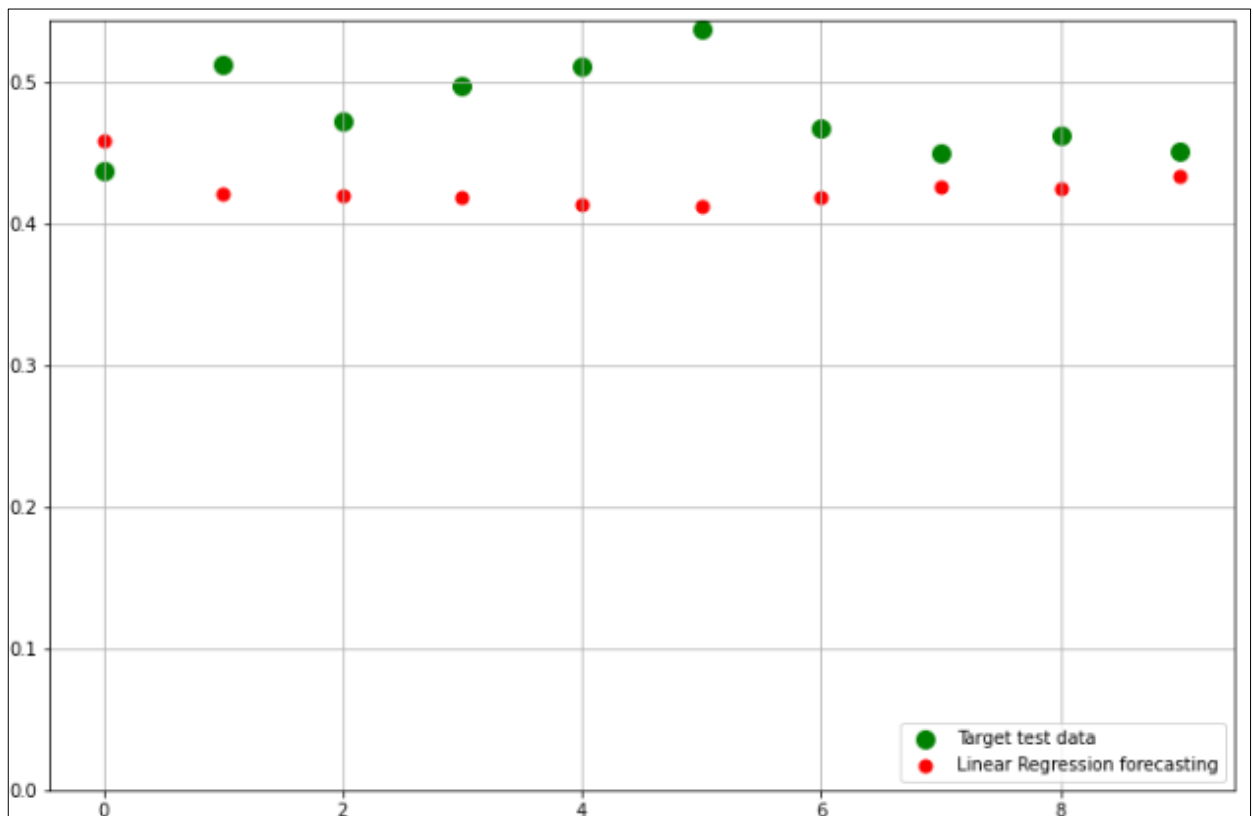


Рисунок 3.8 – Реальні тестові дані курсу криптовалюти Gala та їх прогноз за моделлю Linear Regression, оптимальної за метрикою mape

Діаграма важливості ознак зображена на рисунку 3.9.

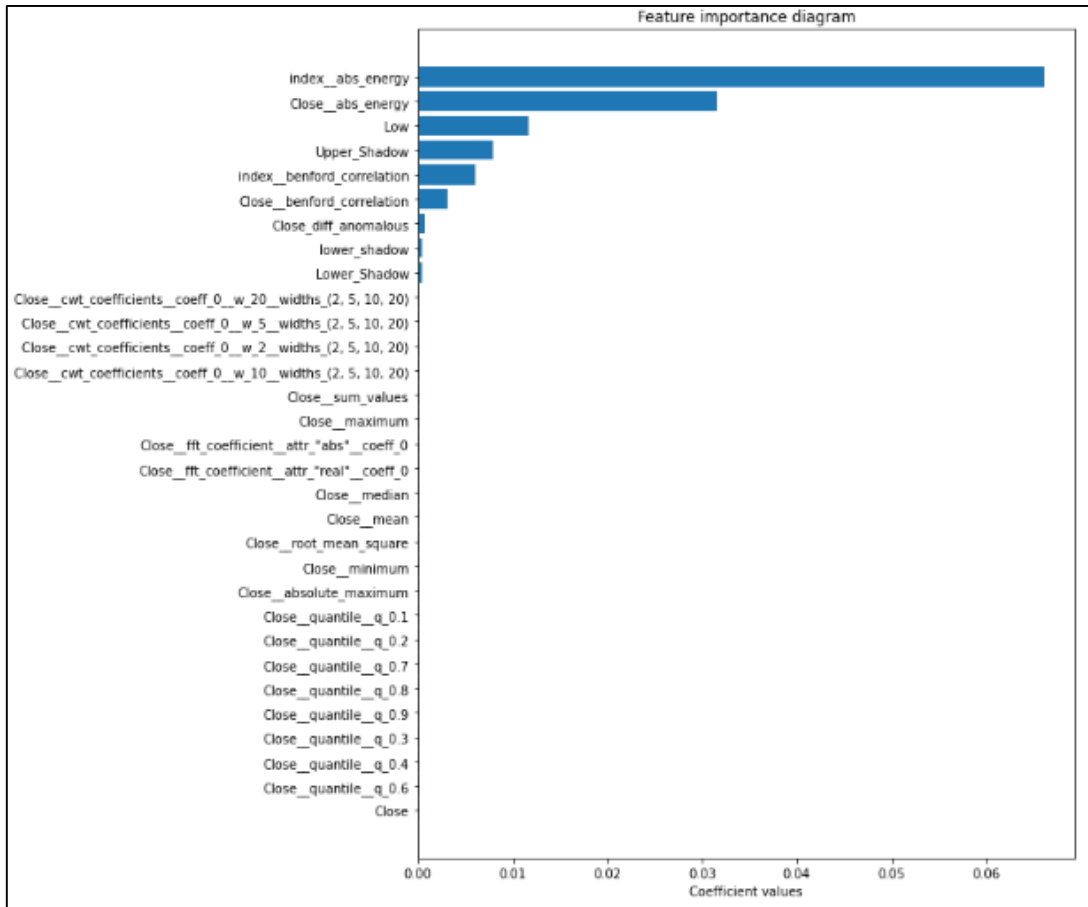


Рисунок 3.9 – Діаграма важливості ознак

Побудуємо деталізовану діаграму важливості ознак за допомогою бібліотеки SHAP для оптимальної моделі, побудованої за методом машинного навчання «Linear Regression», для моделювання курсу криптовалюти (рис. 3.10) [15].

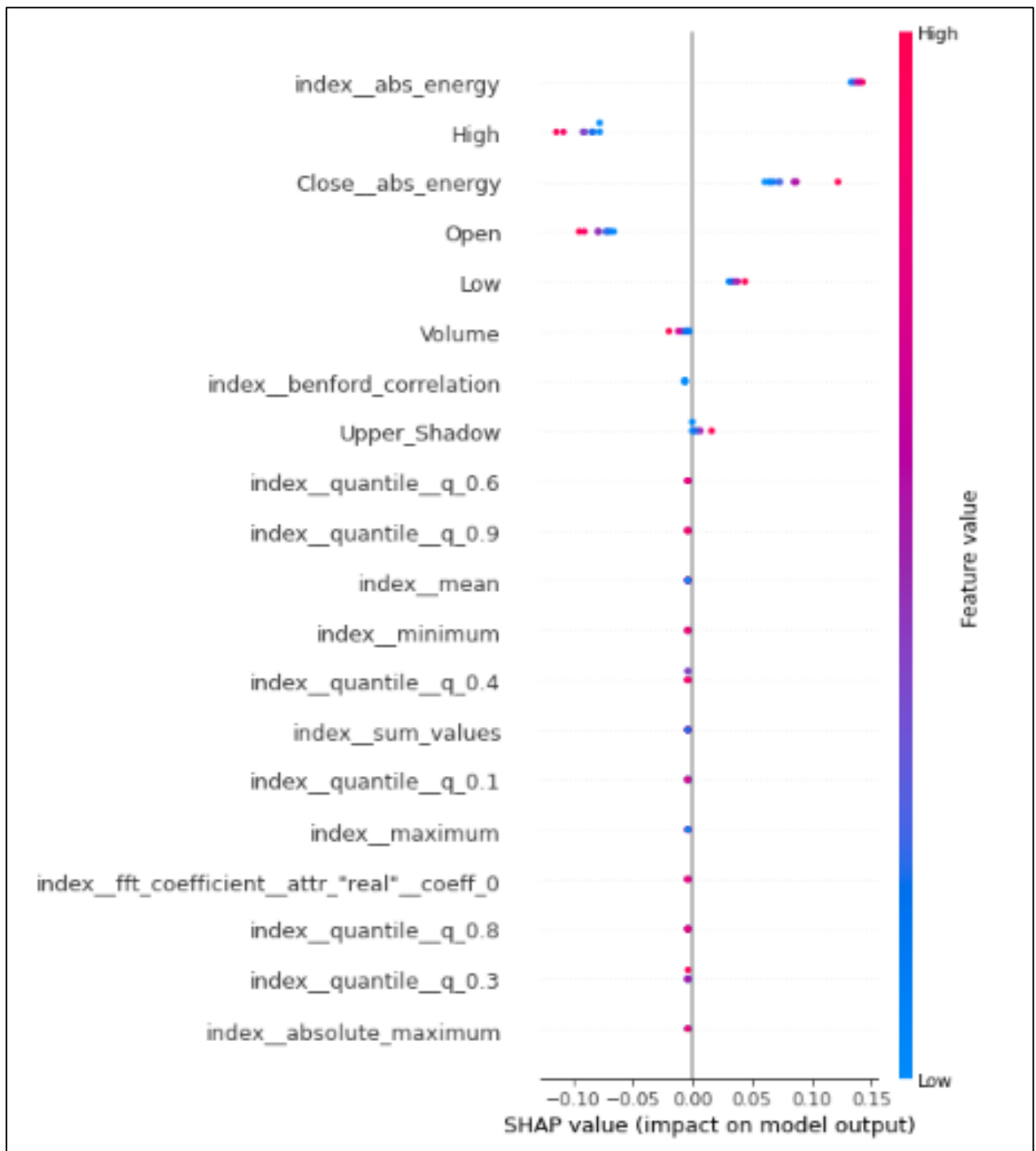


Рисунок 3.10 – Деталізована діаграма важливості ознак

Побудуємо діаграму важливості ознак за допомогою модулю `inspection` бібліотеки `sklearn` для оптимальної моделі, побудованої за методом машинного навчання для «Linear Regression» моделювання курсу криптовалюти (рис. 3.11).

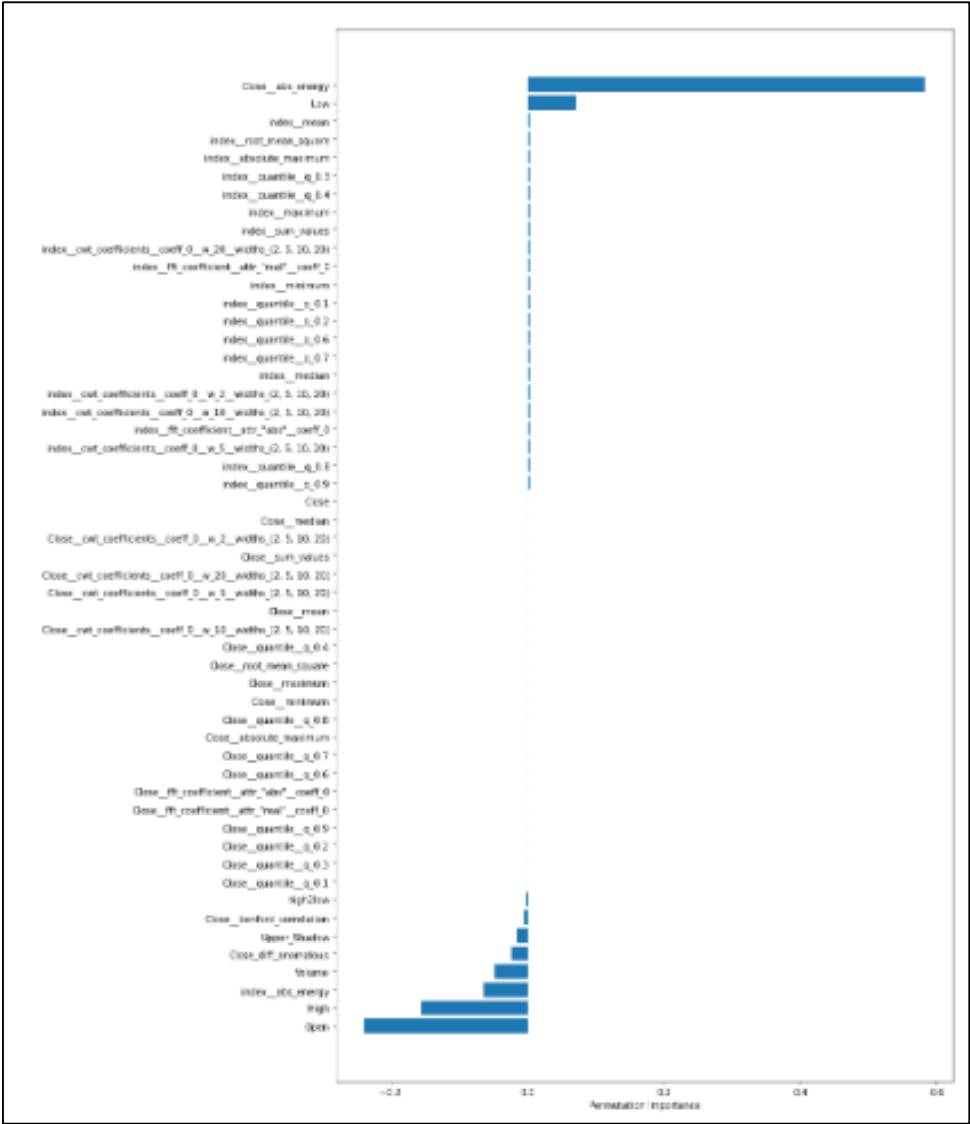


Рисунок 3.11 – Діаграма перестановок із важливістю ознак

Побудувано діаграму типу «Force plot» із важливістю ознак, які вплинули на прогноз значення у першу дату тестового датасету, за допомогою бібліотеки SHAP для оптимальної моделі, побудованої за методом машинного навчання «Linear Regression», для моделювання курсу криптовалюти (рис. 3.12).

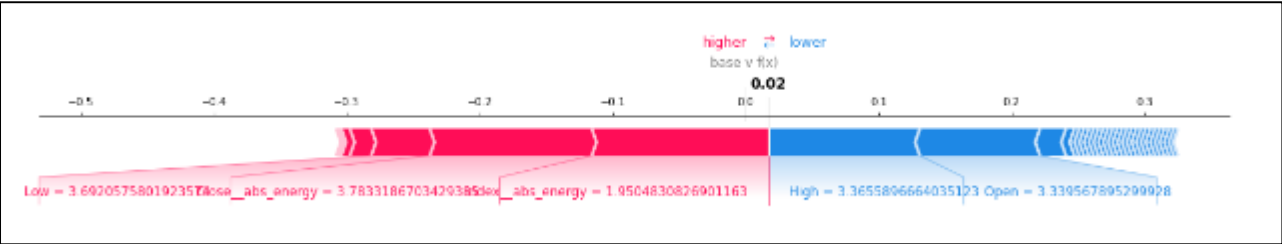


Рисунок 3.12 – Діаграма типу «Force plot»

Побудовано діаграму перестановок із важливістю ознак, за допомогою бібліотеки ELI5 для оптимальної моделі, побудованої за методом машинного навчання «Linear Regression» для моделювання курсу криптовалюти (рис. 3.13).

Weight	Feature
0.5885 ± 0.6819	Close_abs_energy
0.0397 ± 0.1781	Low
0.0223 ± 0.2403	Volume
0.0048 ± 0.0153	lower_shadow
0.0026 ± 0.0006	index_root_mean_square
0.0026 ± 0.0026	index_sum_values
0.0022 ± 0.0029	index_maximum
0.0020 ± 0.0031	index_fft_coefficient_attr_ "real" _coeff_0
0.0019 ± 0.0038	index_cwt_coefficients_coeff_0_w_10_widths_(2, 5, 10, 20)
0.0018 ± 0.0017	index_quantile_q_0.2
0.0016 ± 0.0030	index_quantile_q_0.6
0.0016 ± 0.0026	index_quantile_q_0.1
0.0016 ± 0.0017	index_cwt_coefficients_coeff_0_w_2_widths_(2, 5, 10, 20)
0.0015 ± 0.0021	index_cwt_coefficients_coeff_0_w_20_widths_(2, 5, 10, 20)
0.0014 ± 0.0031	index_median
0.0014 ± 0.0019	index_quantile_q_0.3
0.0014 ± 0.0018	index_quantile_q_0.9
0.0012 ± 0.0030	index_absolute_maximum
0.0010 ± 0.0019	index_minimum
0.0009 ± 0.0020	index_mean
... 40 more ...	

Рисунок 3.13 – Діаграма перестановок із важливістю ознак

Отже, як видно з побудованих діаграм на рисунках 3.10-3.13 можна зробити висновок щодо впливу ознак для прогнозу. Виділено ознаки, які мають найбільшу важливість, це такі як: «Close», «Open», «Close\_bemford\_correlation», «index\_quantile\_q\_0.6», «Close\_abs\_energy», «index\_sum\_values», «Close\_diff\_anomalius», «index\_abs\_energy», «High2Low» «index\_quatil\_q\_0.9 », «index\_mean», «Low», «High», «Lower\_Shadow».

Різні ознаки мають різний вплив на прогнозування, це зв'язано з різною важливістю ознак сформованих на основі різних даних та різної вагомості на результат. Таким чином ознаки: «Open», «Close», «Low», «High», які є базовими для кожної криптовалюти, мають досить високу вагомість впливу на результат, можна виділити, що ознаки, які сформовані на основі «Close», «Low», «High», також є важливими, проте мають менший вплив.

Для збільшення точності прогнозу можна збільшити обсяг даних. Це можна зробити зменшивши інтервал з денного до чотиригодинного або годинного, таким чином, це збільшить об'єм даних та дасть більше даних, саме ознаки «Open», яка є найважливішою.

### **3.3 Висновки**

У третьому розділі сформовано алгоритм роботи технології. Побудовано графік з реальними тестовими даними курсу криптовалюти Gala та їх прогнозом за оптимальними моделями по різних метрикам. Під час прогнозування даних на розширеному датасеті отримана похибка за метрикою RMSE у розмірі 0,03 USD, що свідчить про те, що модель добре навчена і дає досить точний прогноз курсу криптовалюти Gala. Для оптимальної моделі побудовано діаграму перестановок із важливістю ознак за допомогою бібліотеки ELI5. Для моделювання курсу криптовалюти Gala оптимальною моделлю виявилась «Linear Regression». На основі діаграми ELI5 була здійснена оптимізація ознак. Найважливішими виявились ознаки: «Open», «Close», «Low», «High», «Index\_\_abs\_energy» які є базовими для кожної криптовалюти та мають досить високий вплив на результат прогнозування. Також можна виділити, що ознаки, які сформовані на основі «Close», «Low», «High» мають менший вплив на результат прогнозування.

## 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 із використанням технологій машинного навчання для прогнозування курсу криптовалюти Gala. Особливістю розробки є підвищення точності інформаційної технології для прогнозування курсу криптовалюти Gala. Розроблювана технологія матиме високий показник точності прогнозування курсу криптовалюти Gala, а також буде універсальною так як датасет містить топ-100 криптовалют і будь-яку із них можна використовувати для прогнозування.

Аналогом може бути Walletinvestor, Belinvestor, NeuroShell орієнтовною вартістю 65000\$.

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 таблиці 4.1.

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

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 таблиці 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 аналогів
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
Практик на 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 таблиці 4.1

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		
	Пасека Б. В.	Богачук А. Р.	Гусак С. В.
	Бали		
Технічна здійсненність концепції	3	4	4
Наявність аналогів на ринку	4	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	4
Термін реалізації ідеї	4	4	4
Супровідна документація	3	4	4
Сума	44	43	45
Середньоарифметична сума балів	$(44+43+45) / 3 = 44$		

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 розробляється інформаційна технологія із використанням технологій машинного навчання для прогнозування курсу криптовалюти Gala. Особливістю розробки є підвищення точності інформаційної технології для прогнозування курсу криптовалюти Gala. Розроблювана технологія матиме високий показник точності прогнозування курсу криптовалюти Gala, а також буде універсальною так як датасет містить топ-100 криптовалют і будь-яку із них можна використовувати для прогнозування.

#### **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, 20 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	35000	1750,00	32	56000,000
Програміст	30000	1500,00	32	48000,000
Всього				104000,00

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 брати участь в розробці обладнання/програмного продукту.

Додаткову заробітну плату прийнято розраховувати як 13 % від основної заробітної плати розробників та робітників:

$$Z_d = Z_o \cdot 13 \% / 100 \% \quad (4.2)$$

$$Z_d = (104000,00 \cdot 13 \% / 100 \% ) = 13520,00 \text{ (грн).}$$

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

$$H_3 = (Z_o + Z_d) \cdot 22 \% / 100\% \quad (4.3)$$

$$H_3 = (104000,00 + 13520,00) \cdot 22 \% / 100 \% = 25854,40 \text{ (грн).}$$

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:

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

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 30000 грн., термін його корисного використання згідно податкового законодавства – 2 роки, а термін його фактичного використання – 1,60 міс.

$$A_{\text{обл}} = \frac{30000}{2} \times \frac{1,6}{12} = 2000 \text{ (грн)}.$$

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  $B_{\text{нем.ак.}} = 0$  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 (Lenovo IdeaPad 3 15IAU7)	30000	2	1,60	2000,000
Офісне обладнання (меблі)	25000	4	1,60	833,333
Приміщення	1050000	20	1,60	7000,000
Всього				9833,33

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:

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

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

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

$\Phi$  – фактична кількість годин роботи обладнання, годин;

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

$$V_e = 0,9 \cdot 0,4 \cdot 8 \cdot 32 \cdot 6,2 = 571,392 \text{ (грн).}$$

Інші витрати та загальновиробничі 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:

$$I_{\text{в}} = (Z_{\text{o}} + Z_{\text{p}}) \cdot \frac{H_{\text{ів}}}{100\%}, \quad (4.6)$$

де  $H_{\text{ів}}$  – норма нарахування за статтею «Інші витрати»;

$$I_{\text{в}} = 104000,00 * 75\% / 100\% = 78000 \text{ (грн)}.$$

До статті «Накладні (загальновиробничі) витрати» належать: витрати, пов'язані з управлінням організацією; витрати на винахідництво та раціоналізацію; витрати на підготовку (перепідготовку) та навчання кадрів; витрати, пов'язані з набором робочої сили; витрати на оплату послуг банків; витрати, пов'язані з освоєнням виробництва продукції; витрати на науково-технічну інформацію та рекламу та ін. Витрати за статтею «Накладні (загальновиробничі) витрати» розраховуються як 100...150% від суми основної заробітної плати дослідників:

$$H_{\text{нзв}} = (Z_{\text{o}} + Z_{\text{п}}) \cdot \frac{H_{\text{нзв}}}{100\%}, \quad (4.7)$$

де  $H_{\text{нзв}}$  – норма нарахування за статтею «Накладні (загальновиробничі) витрати».

$$H_{\text{нзв}} = 104000,00 * 120\% / 100\% = 124800 \text{ (грн)}.$$

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

$$V_{\text{заг}} = 104000,00 + 13520,00 + 25854,40 + 9833,33 + 571,39 + 78000 +$$

$$+124800 = 356579,13 \text{ (грн)}.$$

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:

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

де  $\eta$  – 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:

$$ЗВ = 356579,13 / 0,5 = 713158 \text{ (грн)}.$$

#### **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 тієї 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:

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

б) 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);

в) 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;

г) 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;

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

$\Pi_6$  – вартість програмного продукту у році до впровадження результатів розробки;

$\Delta N$  – збільшення кількості споживачів продукту, в аналізовані періоди часу, від покращення його певних характеристик;

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

$\rho$  – коефіцієнт, який враховує рентабельність продукту;

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

Припустимо, що при прогнозованій ціні 9500 грн. за одиницю виробу, термін збільшення прибутку складе 3 роки. Після завершення розробки і її вдосконалення, можна буде підняти її ціну на 500 грн. Кількість одиниць реалізованої продукції також збільшиться: протягом першого року – на 8000 шт., протягом другого року – на 3000 шт., протягом третього року на 1000 шт. До моменту впровадження результатів наукової розробки реалізації продукту не було:

$$\Delta\Pi_1 = (0 \cdot 500 + (9500 + 500) \cdot 8000) \cdot 0,8333 \cdot 0,2 \cdot (1 - 0,18) = 10386666,251 \text{ (грн).}$$

$$\Delta\Pi_2 = (0 \cdot 500 + (9500 + 500) \cdot (8000 + 3000)) \cdot 0,8333 \cdot 0,2 \cdot (1 - 0,18) = 15033332,732 \text{ (грн).}$$

$$\Delta\Pi_3 = (0*500 + (9500 + 500) * (8000 + 3000 + 1000) * 0,8333 * 0,2) * (1 - 0,18) = 16399999,344 \text{ (грн)}.$$

Отже, комерційний ефект від реалізації результатів розробки за три роки складе 41819998,33 грн.

Розрахунок ефективності вкладених інвестицій та 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:

$$ПП = \sum_1^T \frac{\Delta\Pi_i}{(1 + \tau)^t}, \quad (5.10)$$

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 якого виявляються результати впровадженої науково-дослідної (науково-технічної) роботи, роки;

$\tau$  – 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:

$$ПП = (10386666,251 / (1 + 0,1)^1) + (15033332,732 / (1 + 0,1)^2) + (16399999,344 / (1 + 0,1)^3) = 9442423,86 + 12424241,93 + 12321562,24 = 34188228,03 \text{ (грн)}.$$

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

$$PV = k_{инв} * ЗВ, \quad (4.11)$$

де  $k_{инв}$  – 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  $k_{инв}=2...5$ , але може бути і більшим;

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

$$PV = 2 * 713158 = 1426316,50 \text{ (грн)}.$$

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

$$E_{абс} = ПП - PV, \quad (4.12)$$

$$E_{абс} = 34188228,03 - 1426316,50 = 32761911,53 \text{ (грн)}.$$

Оскільки  $E_{абс} > 0$  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  $E_g$ . Для цього використаємо формулу:

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

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

$$E_g = \sqrt[3]{(1 + 32761911,53/1426316,50) - 1} = 1,883.$$

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  $d = (0,09...0,14)$ ;

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

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

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

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

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

$$T_{ок} = 1 / 1,883 = 0,53 \text{ (роки).}$$

Оскільки  $T_{ок} < 3$ -х років, а саме термін окупності рівний 0,53 роки, то фінансування даної наукової розробки є доцільним.

#### **4.4 Висновки**

Економічна частина даної роботи 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 конкурентоспроможним. Період окупності складе близько 0,53 роки.

## ВИСНОВКИ

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

Розглянуто проблеми прогнозування курсу криптовалюти Gala, існуючі технології прогнозування, отримано базові ознаки датасету по криптовалюти Gala за даними сайту Yahoo. Для даної роботи обрано методи машинного навчання для прогнозування та виділення важливості кожного з елементів інформаційного впливу, оскільки в загальному вони дають кращі результати ніж статистичні.

Зроблено розвідувальний аналіз даних, здійснено і описано синтез нових ознак за яким відібрано 561 ознаку на основі базової інформації крипторинків, досліджено 12 аномальних дат, які вагомо впливають на подальше прогнозування курсу криптовалюти Gala, побудовано моделі машинного навчання для прогнозування курсу криптовалюти Gala та обрано оптимальну модель за різними метриками, якою виявилась модель, побудована на основі методу «Linear Regression». З'ясовано, що похибка оптимальної моделі за метрикою MAPE становить 5.17%, за метрикою RMSE становить 0,03 USD, що означає, що самі значення прогнозуються достатньо точно.

Розроблено інформаційну технологію прогнозування курсу криптовалюти Gala на основі методу машинного навчання Linear Regression. Розроблено блок-схему алгоритму роботи інформаційної технології прогнозування курсу криптовалюти Gala. Також побудовано діаграму перестановок із важливістю ознак за допомогою бібліотеки ELI5, на основі якої здійснено оптимізацію ознак. Найважливішими виявились ознаки: «Open», «Close», «Low», «High», «Index\_\_abs\_energy», а також ознаки, які сформовані на основі «Close», «Low», «High» які мають досить високий вплив на результат.

Економічна частина даної роботи 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 конкурентоспроможним. Період окупності складе близько 0,53 роки.

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

Опубліковано тези з результатами роботи на Всеукраїнській науково-практичній інтернет-конференції «Молодь в науці: дослідження, проблеми, перспективи» (Вінниця, 2022-2023 рр.).

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Mokin V.B. Kaggle Dataset «Forecasting Top Cryptocurrencies». URL: <https://www.kaggle.com/datasets/vbmokin/forecasting-top-cryptocurrencies>
2. Жук В.Л., Жуков С.О. Вибір оптимальної моделі для прогнозування курсу криптовалюти Gala у 2021 – 2022 роках. *Молодь в науці: дослідження, проблеми, перспективи (МН-2022)*. Вінниця, 2022. URL: <https://conferences.vntu.edu.ua/index.php/mn/mn2023/paper/view/16826>
3. Crypto gaming. Gala games. URL: <https://venturebeat.com/2022/06/06/gala-games-blockchain-game-grit-will-debut-on-the-epic-games-store/>
4. Криптовалюта Gala. Основні поняття. URL: <https://coinmarketcap.com/uk/currencies/gala/>
5. Mokin V.B. Kaggle Notebook «Crypto - BTC: Advanced Analysis & Forecasting». URL: <https://www.kaggle.com/vbmokin/crypto-btc-advanced-analysis-forecasting>
6. Mokin V.B. Kaggle Notebook «Cryptocurrency: Years data comparing with AutoViz». URL: <https://www.kaggle.com/code/vbmokin/cryptocurrency-years-data-comparing-with-autoviz>
7. Kharpal A., Newburger E. drops below \$33,000 to hit its lowest level since July 2021 after stock sell-off. 8 Jan 2021. URL: <https://www.cnbc.com/2022/05/08/bitcoin-drops-below-35000-over-the-weekend-extending-fridays-losses.html>
8. Mokin V.B. Kaggle Notebook «BTC & COVID-19 in USA : EDA & Prediction». URL: <https://www.kaggle.com/code/vbmokin/btc-covid-19-in-usa-eda-prediction>
9. Facebook Prophet Documentation. Quick Start. URL: [https://facebook.github.io/prophet/docs/quick\\_start.html](https://facebook.github.io/prophet/docs/quick_start.html)
10. Statsmodels.tsa Documentation. Statsmodels.tsa.stattools.adfuller. URL: <https://www.statsmodels.org/dev/generated/statsmodels.tsa.stattools.adfuller.html>



11. Statsmodels.tsa Documentation. Statsmodels.tsa.seasonal.seasonal\_decompose. URL: [https://www.statsmodels.org/dev/generated/statsmodels.tsa.seasonal.seasonal\\_decompose.html](https://www.statsmodels.org/dev/generated/statsmodels.tsa.seasonal.seasonal_decompose.html)
12. Major world market indices. URL: <https://www.investing.com/indices/major-indices> (Accessed on: Oct. 7, 2022).
13. Mokin Vitalii. Kaggle Notebook «BTC & COVID-19 in USA : EDA & Prediction», URL: <https://www.kaggle.com/vbmokin/btc-covid-19-in-usa-eda-prediction> (Accessed on: Oct. 7, 2022).
14. 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. DOI: <https://doi.org/10.31649/1997-9266-2022-161-2-81-93>.
15. «Facebook Prophet» Documentation. Quick Start. URL: [https://facebook.github.io/prophet/docs/quick\\_start.html](https://facebook.github.io/prophet/docs/quick_start.html)
16. Машинне навчання. URL: <https://www.it.ua/knowledge-base/technology-innovation/machine-learning>
17. Statsmodels.tsa Documentation. Statsmodels.tsa.stattools.adfuller. URL: <https://www.statsmodels.org/dev/generated/statsmodels.tsa.stattools.adfuller.html>
18. Statsmodels.tsaDocumentation.Statsmodels.tsa.seasonal.seasonal\_decompose. URL: [https://www.statsmodels.org/dev/generated/statsmodels.tsa.seasonal.seasonal\\_decompose.html](https://www.statsmodels.org/dev/generated/statsmodels.tsa.seasonal.seasonal_decompose.html)
19. Бібліотека Shape. URL: <https://habr.com/en/post/428213/>
20. Бібліотека Sklearn. URL: <https://scikit-learn.org/stable/>

## 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 GALA У 2021-2022 РОКАХ»

08-53.МКР.002.02.000.ТЗ

Керівник: к.т.н., доц. каф. САІТ

\_\_\_\_\_ Жуков С. О.

«\_19\_» \_\_\_\_\_ 09 \_\_\_\_\_ 2022 р.

Розробив: студент гр. 2ІСТ-21м

\_\_\_\_\_ Жук В. Л.

«\_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.
- 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 Gala 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 Gala 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 – 25.09.2022
2. Ідентифікація та вибір оптимальної моделі для прогнозування криптовалюти Gala ..... 26.09 – 31.09.2022
3. Розроблення інформаційної технології для аналізу та прогнозування курсу криптовалюти Gala ..... 01.10 – 10.10.2022
4. Економічна частина ..... 15.10 – 30.10.2022
5. Оформлення пояснювальної записки ..... 10.11 – 30.11.2022

### 7. Очікувані результати та порядок реалізації:

Отримані результати по прогнозуванню курсу криптовалюти Gala 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» грудня 2022 р.  
Початок розробки ..... «20» вересня 2022 р.  
Граничні терміни виконання МКР ..... «30» листопада 2022 р.

Розробив студент групи 2ІСТ-21м \_\_\_\_\_ Жук В. Л.

## Додаток Б

Протокол перевірки кваліфікаційної роботи на наявність текстових запозичень

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

Тип роботи: магістерська кваліфікаційна робота

Підрозділ: кафедра САІТ

Науковий керівник: Жуков С. О. к.т.н., доц. каф. САІТ

**Показники звіту подібності Unicheck**

Оригінальність	87,4 %
Схожість	12,6 %

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



Жук В. Л.

Керівник роботи



Жуков С. О.

## 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)
```

Додаток Г

## **ІЛЮСТРАТИВНА ЧАСТИНА**

### **ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПРОГНОЗУВАННЯ КУРСУ КРИПТОВАЛЮТИ GALA У 2021-2022 РОКАХ**

Виконав: студент гр. 2ІСТ-21м

\_\_\_\_\_ Жук В. Л.

«\_ 01 \_» \_\_\_\_\_ 12 \_\_\_\_\_ 2022 р.

Керівник: к.т.н., доц. каф. САІТ

\_\_\_\_\_ Жуков С. О.

«\_ 02 \_» \_\_\_\_\_ 12 \_\_\_\_\_ 2022 р.

Нормоконтроль: к.т.н., доцент

\_\_\_\_\_ Жуков С. О.

«\_ 02 \_» \_\_\_\_\_ 12 \_\_\_\_\_ 2022 р.

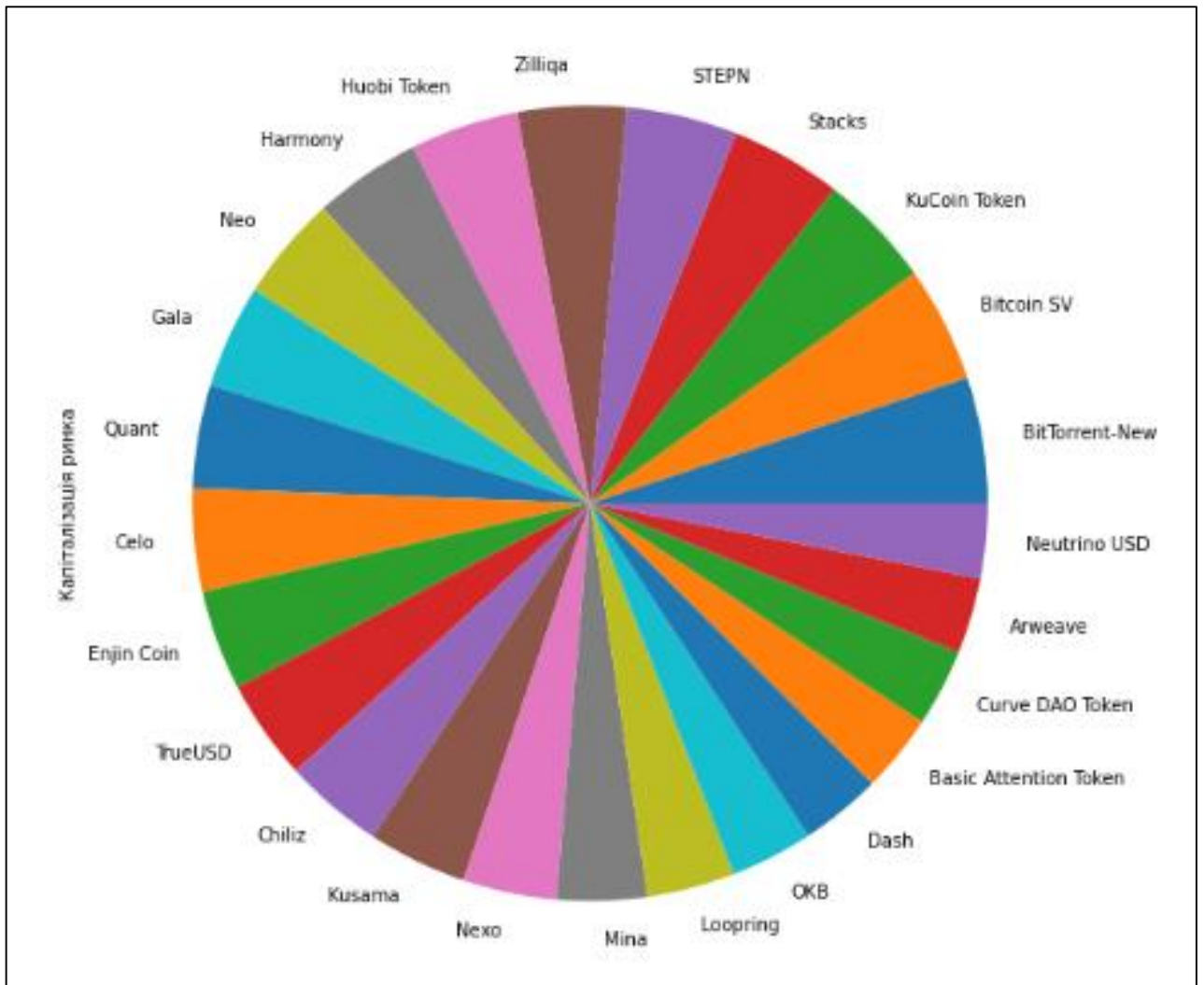


Рисунок Г.1 – Капіталізація ринку криптовалюти

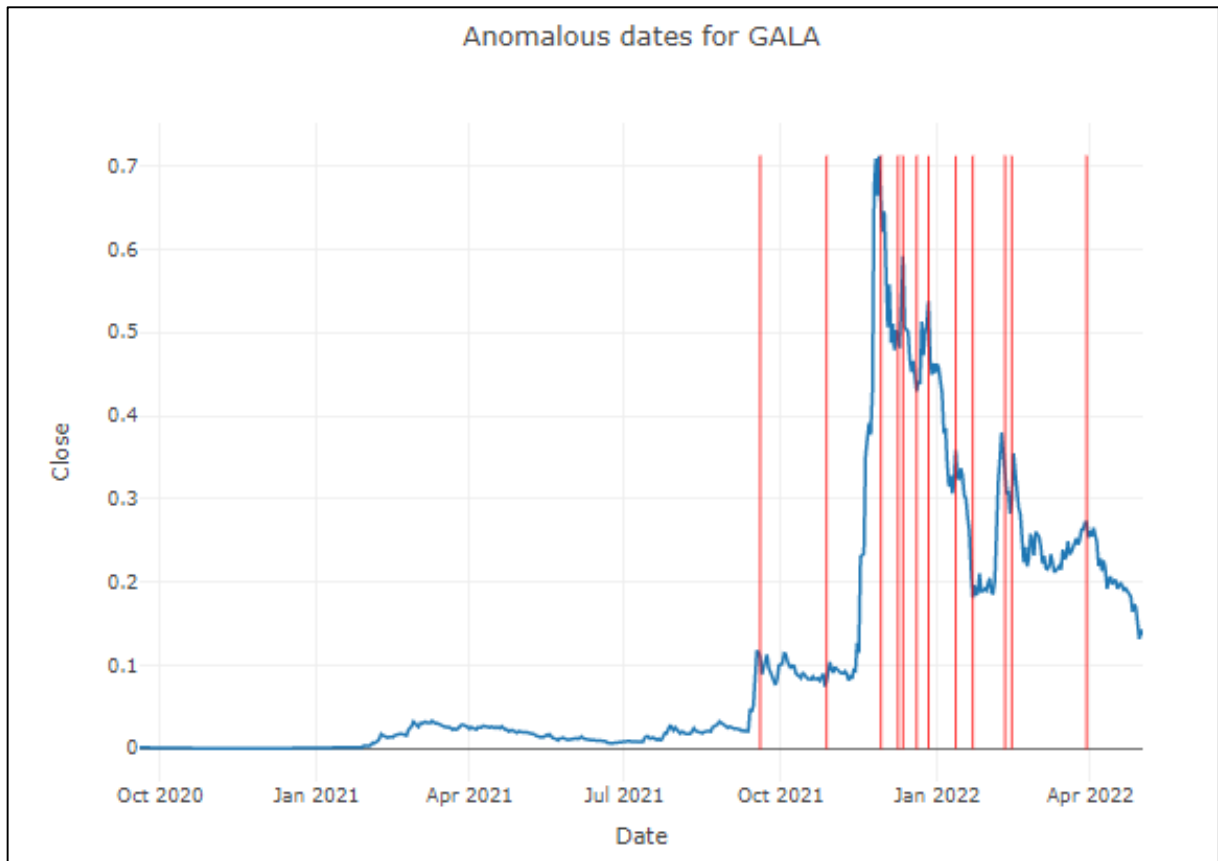


Рисунок Г.2 – Курс криптовалюти (синя товста лінія) та дати аномальних змін (тонкі вертикальні лінії)

	name_model	type_data	r2_score	rmse	mape
9	Linear Regression	valid	0.543857	0.030552	5.171395
12	Linear SVR	valid	0.356824	0.036279	6.867392
0	Prophet_4_days_3_order	valid	-0.657371	0.058237	11.300432
4	Prophet_7_days_3_order	valid	-1.084688	0.065314	12.102852
2	Prophet_5_days_3_order	valid	-1.060643	0.064937	12.846515
7	Prophet_14_days_12_order	valid	-1.544065	0.072153	14.056655
3	Prophet_5_days_12_order	valid	-1.600915	0.072954	14.676709
6	Prophet_14_days_3_order	valid	-1.71488	0.074536	14.912129
5	Prophet_7_days_12_order	valid	-2.080754	0.079399	15.474591
1	Prophet_4_days_12_order	valid	-2.035468	0.078814	15.893748
16	MLP Regressor	valid	-5.044017	0.111212	22.385564
10	KNeighbors Regressor	valid	-7.391326	0.13104	26.252296
13	Random Forest Regressor	valid	-10.13803	0.150971	30.008292
8	ARIMA_auto	valid	-13.441008	0.171904	34.138191
15	XGB Regressor	valid	-31.958058	0.259698	50.787187
14	Bagging Regressor	valid	-37.101729	0.279229	55.8096
11	Support Vector Machines	valid	-69.607196	0.380113	73.92856

Рисунок Г.3 – Таблица похибки моделей

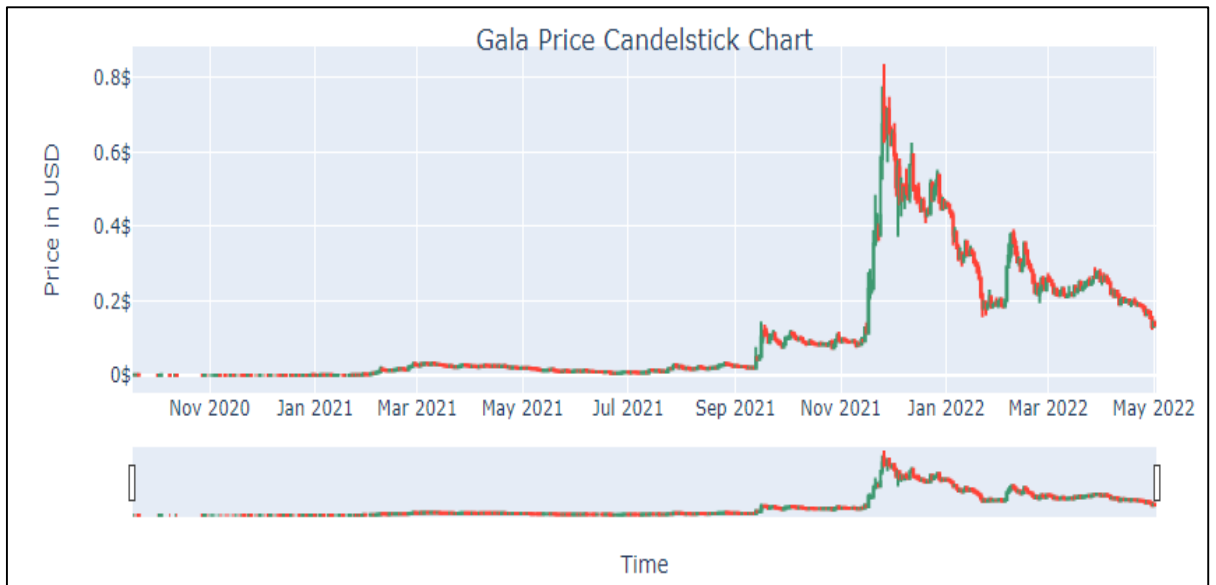


Рисунок Г.4 – OHLC-діаграма криптовалюти Gala за даними сайту Yahoo

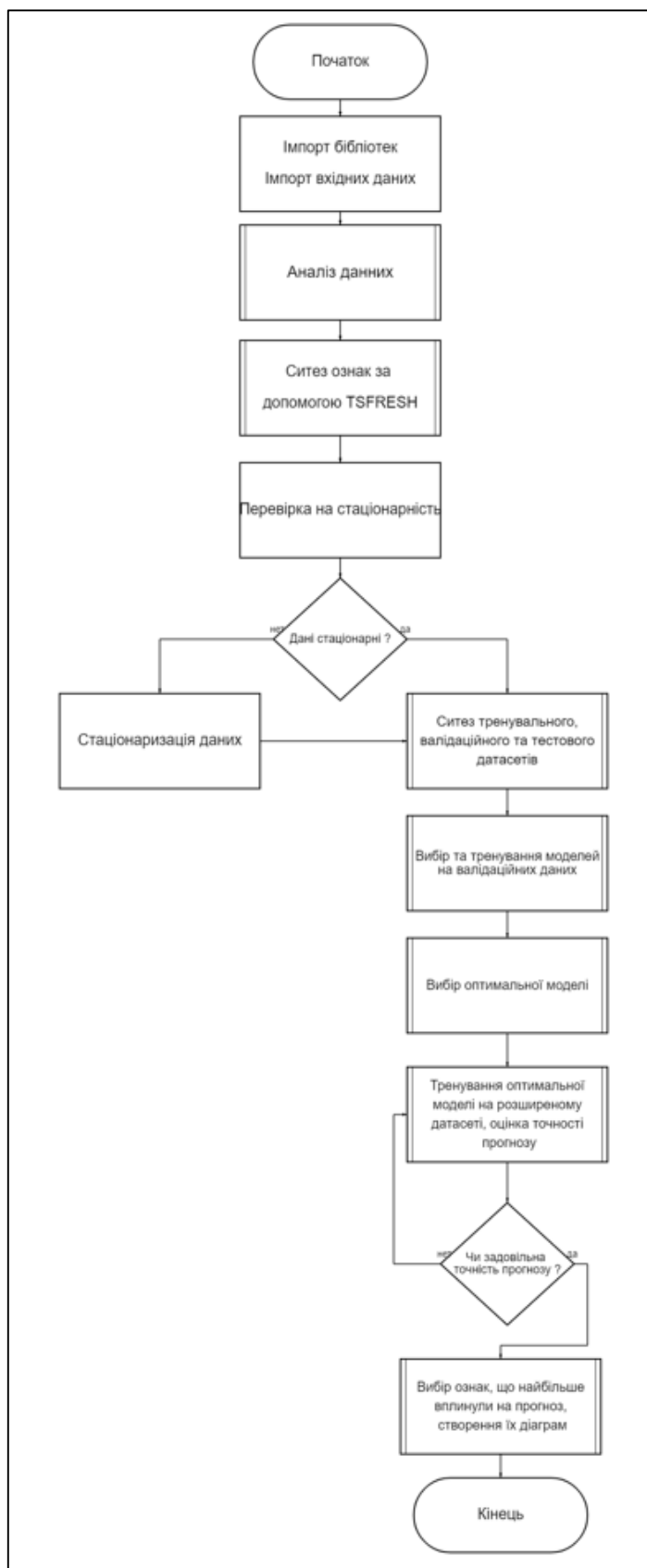


Рисунок Г.5 – Блок-схема алгоритму інформаційної технології прогнозування курсу Gala у 2021-2022 роках

Weight	Feature
0.5885 ± 0.6819	Close_abs_energy
0.0397 ± 0.1781	Low
0.0223 ± 0.2403	Volume
0.0048 ± 0.0153	lower_shadow
0.0026 ± 0.0006	index__root_mean_square
0.0026 ± 0.0026	index__sum_values
0.0022 ± 0.0029	index__maximum
0.0020 ± 0.0031	index__fft_coefficient__attr_"real"__coeff_0
0.0019 ± 0.0038	index__cwt_coefficients__coeff_0__w_10__widths_(2, 5, 10, 20)
0.0018 ± 0.0017	index__quantile__q_0.2
0.0016 ± 0.0030	index__quantile__q_0.6
0.0016 ± 0.0026	index__quantile__q_0.1
0.0016 ± 0.0017	index__cwt_coefficients__coeff_0__w_2__widths_(2, 5, 10, 20)
0.0015 ± 0.0021	index__cwt_coefficients__coeff_0__w_20__widths_(2, 5, 10, 20)
0.0014 ± 0.0031	index__median
0.0014 ± 0.0019	index__quantile__q_0.3
0.0014 ± 0.0018	index__quantile__q_0.9
0.0012 ± 0.0030	index__absolute_maximum
0.0010 ± 0.0019	index__minimum
0.0009 ± 0.0020	index__mean
... 40 more ...	

Рисунок Г.6 – Діаграма перестановок із важливістю ознак



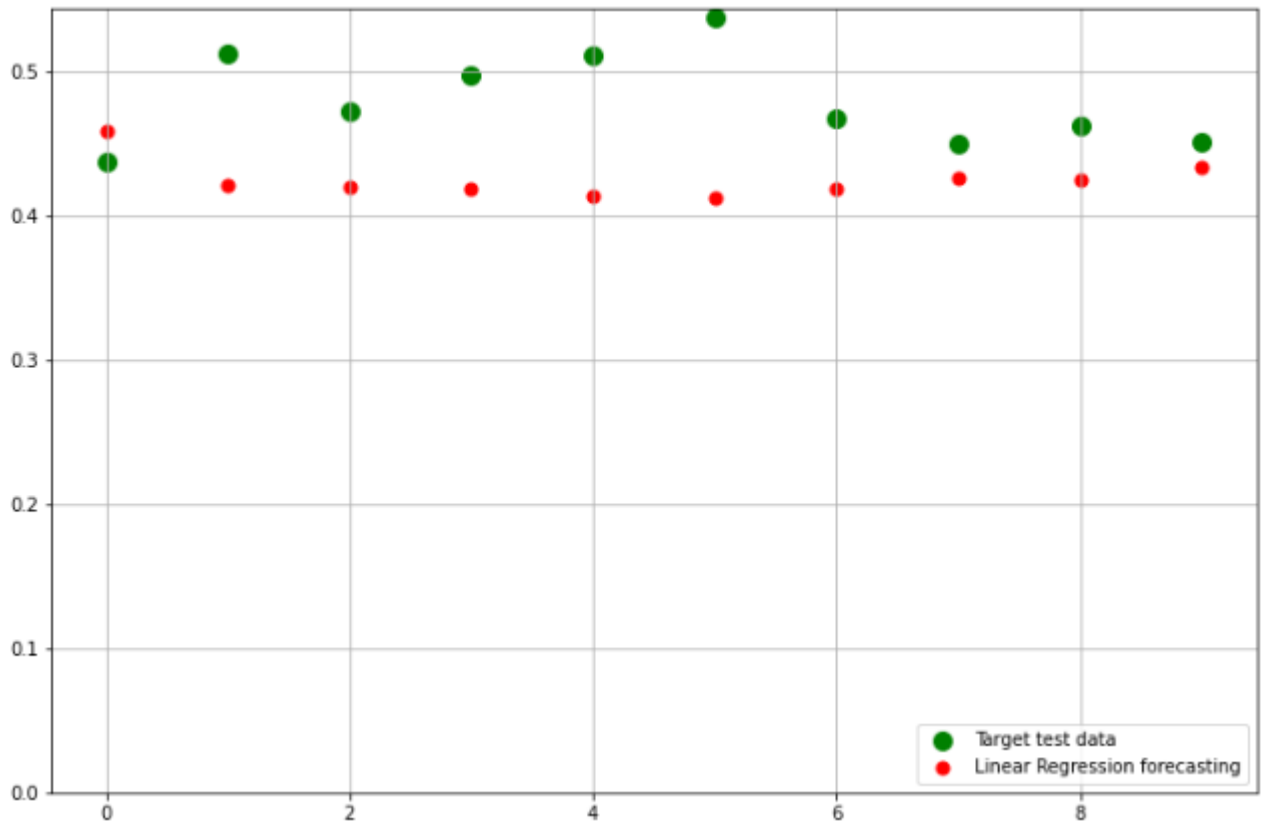


Рисунок Г.7 – Реальні тестові дані курсу криптовалюти Gala та їх прогноз за оптимальною моделлю - Linear Regression