

ГУЗЮНЬ Ірина Володимирівна

**Програмний модуль прогнозування ціни біткоїна
засобами машинного навчання / Bitcoin Price
Prediction Software Module Using Machine
Learning**

спеціальність: 122 - Комп'ютерні науки
освітньо-професійна програма - Комп'ютерні науки

Кваліфікаційна робота

Виконала студентка групи КН-41
І. В. Гузюнь

Науковий керівник:
д.т.н., професор, М. П. Комар

Кваліфікаційну роботу
допущено до захисту:

" ____ " _____ 20 ____ р.

Завідувач кафедри
_____ **М. П. Комар**

ТЕРНОПІЛЬ - 2024

Факультет комп'ютерних інформаційних технологій
Кафедра інформаційно-обчислювальних систем і управління
Освітній ступінь «бакалавр»
спеціальність 122 – Комп'ютерні науки
освітньо-професійна програма – Комп'ютерні науки

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ М.П. Комар
«_____» _____ 2023 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ
Гузюнь Ірина Володимирівна
(прізвище, ім'я, по батькові)

1. Тема кваліфікаційної роботи: Програмний модуль прогнозування ціни біткоїна засобами машинного навчання / Bitcoin Price Prediction Software Module Using Machine Learning

керівник роботи Комар Мирослав Петрович, д.т.н., професор

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від 12 грудня 2023 р. № 753.

2. Строк подання студентом закінченої кваліфікаційної роботи 15 травня 2024 р.

3. Вихідні дані до кваліфікаційної роботи: завдання на кваліфікаційну роботу студента, наукові статті, технічна література.

4. Основні питання, які потрібно розробити:

- Провести опис предметної області.
- Зробити аналіз використання глибокого навчання для прогнозування фінансових часових рядів.
- Дослідити стан розвитку методів прогнозування цін криптовалют.
- Зробити постановку задачі дослідження.
- Дослідити поточний стан та аналіз методів прогнозування цін біткоїна.
- Дослідити теоретичні основи застосування глибокого навчання у прогнозуванні фінансових даних.
- Розробити модель прогнозування ціни біткоїна з XGBoost.
- Провести збір наборів даних і налаштування гіперпараметрів.
- Зробити аналіз точності прогнозування цін за різними алгоритмами моделей на навчальному наборі даних.
- Зробити порівняльний аналіз точності прогнозування цін за різними алгоритмами моделей на тестовому наборі даних.

5. Перелік графічного матеріалу в роботі:

- результати експериментальних досліджень.
- Аналоги існуючих засобів прогнозування ціни біткоїна
- Алгоритм роботи програмного модуля

6. Консультанти розділів кваліфікаційної роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		Завдання видав	Завдання прийняв

7. Дата видачі завдання 12 грудня 2023 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів кваліфікаційної роботи	Примітка
1	Затвердження теми кваліфікаційної роботи, ознайомлення з літературними джерелами та складання плану роботи.	до 01.01. 2024 р.	
2	Написання 1 розділу кваліфікаційної роботи	до 01.03. 2024 р.	
3	Написання 2 розділу кваліфікаційної роботи	до 01.04.2024 р.	
4	Написання 3 розділу кваліфікаційної роботи	до 01.05. 2024 р.	
5	Представлення попереднього варіанту кваліфікаційної роботи, перевірка та внесення змін керівником	до 15.05.2024 р.	
6	Опрацювання зауважень та представлення завершеного варіанту кваліфікаційної роботи. Підготовка супроводжуючих документів.	до 20.05.2024 р.	
7	Перевірка кваліфікаційної роботи на оригінальність тексту у системі «Unichesk».	до 10.06.2024 р.	
8	Оформлення кваліфікаційної роботи та отримання допуску до захисту	до 14.06.2024 р.	
9	Подання кваліфікаційної роботи до захисту на засіданні атестаційної комісії.	до 14.06. 2024 р.	

Студент _____ І. В. Гузюнь
(підпис) (прізвище та ініціали)

Керівник кваліфікаційної роботи _____ М.П. Комар
(підпис) (прізвище та ініціали)

АНОТАЦІЯ

Кваліфікаційна робота на тему «Програмний модуль прогнозування ціни біткоїна засобами машинного навчання» на здобуття освітнього ступеня «бакалавр» зі спеціальності 122 «Комп'ютерні науки» освітньої програми «Комп'ютерні науки» написана обсягом в 51 сторінок і містить 24 ілюстрацій, 1 таблицю, 2 додатки та 25 використаних джерел.

Метою роботи є дослідження підходів до прогнозування цін криптовалют та розробка програмного модуля для прогнозування ціни біткоїна з використанням сучасних методів машинного навчання.

Методами розроблення обрано метод аналізу (для дослідження існуючих підходів до прогнозування цін), метод синтезу (для поєднання переваг існуючих методів прогнозування), методи моделювання (для представлення та дослідження процесів зміни цін криптовалют), метод порівняльного аналізу (для оцінювання точності прогнозів різних моделей).

Внаслідок виконання роботи обґрунтовано раціональний підхід до розроблення моделей прогнозування цін криптовалют і створений програмний модуль, який дозволяє прогнозувати ціну біткоїна з високою точністю. Розроблений модуль використовує алгоритм XGBoost, що дозволяє обробляти великі обсяги даних і враховувати складні залежності між змінними.

Результати дослідження можуть бути використані інвесторами, трейдерами, фінансовими аналітиками та дослідниками, які займаються прогнозуванням цін криптовалют.

Ключові слова: БІТКОЇН, ПРОГНОЗУВАННЯ, XGBOOST, МАШИННЕ НАВЧАННЯ, КРИПТОВАЛЮТА.

ANNOTATION

Qualification work on the topic " Software Control Unit for Bitcoin Price Prediction using Machine Learnin" for the degree of bachelor in the specialty 122 "Computer Science" of the educational program "Computer Science" is written in 51 pages and contains 24 illustrations, 1 table, 2 appendices and 25 references.

The aim of the work is to study approaches to forecasting cryptocurrency prices and develop a software module for forecasting the price of bitcoin using modern machine learning methods.

The development methods chosen are the analysis method (to study existing approaches to price forecasting), the synthesis method (to combine the advantages of existing forecasting methods), modeling methods (to represent and study the processes of cryptocurrency price changes), and the comparative analysis method (to assess the accuracy of forecasts of different models).

As a result of this work, a rational approach to the development of cryptocurrency price forecasting models has been substantiated and a software module has been created that allows forecasting the price of bitcoin with high accuracy. The developed module uses the XGBoost algorithm, which allows processing large amounts of data and taking into account complex dependencies between variables.

The results of the study can be used by investors, traders, financial analysts, and researchers involved in cryptocurrency price forecasting.

Keywords: BITCOIN, FORECASTING, XGBOOST, MACHINE LEARNING, CRYPTOCURRENCY.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ.....	6
ВСТУП.....	7
1 АНАЛІЗ ЗАДАЧІ ПРОГНОЗУВАННЯ ЦІНИ БІТКОЇНА.....	9
1.1 Опис предметної області	9
1.2 Огляд і аналіз існуючих аналогів	12
1.3 Постановка задачі дослідження.....	18
2 АЛГОРИТМІЧНЕ ТА ІНФОРМАЦІЙНЕ ЗАБЕЗПЕЧЕННЯ МОДУЛЯ.....	20
2.1 Концептуальні основи вирішення задачі прогнозування ціни біткоїна з використанням машинного навчання	20
2.2 Алгоритмічне забезпечення	22
2.3 Інформаційне забезпечення	27
3 ПРОГРАМНО-ТЕХНОЛОГІЧНЕ ЗАБЕЗПЕЧЕННЯ МОДУЛЯ.....	28
3.1 Реалізація програмного забезпечення.....	28
3.2 Тренування моделі	36
3.3 Тестування розробленої програми	39
ВИСНОВКИ.....	42
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	44
Додаток А Копія публікації	47
Додаток Б Код програмного модуля	50

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

DFD - Data Flow Diagrams

XGBoost - eXtreme Gradient Boosting

STD - State Transition Diagram

ВСТУП

У сучасному світі криптовалюти привертають увагу не лише фінансових експертів та інвесторів, але й академічної спільноти. Їхнє швидке зростання і висока волатильність суттєво впливають на глобальні фінансові ринки та економіку загалом.

За останні кілька років виникла потреба у створенні надійних методів прогнозування курсу криптовалют, щоб інвестори та дослідники могли приймати обґрунтовані рішення та аналізувати ринки криптовалют. Прогнозування курсу біткоїна та інших криптовалют є важливим завданням, але водночас дуже складним через їхню хаотичну та нестабільну природу.

Актуальність проблеми прогнозування курсу криптовалют є очевидною. Світ фінансів, інвестицій та технологій зазнає значних змін завдяки криптовалютам, які стали не лише альтернативним засобом інвестицій, але й основною темою фінансових новин. Навіть незначні події можуть спричинити великі коливання цін на ринках криптовалют. Наприклад, злам біржі Bitfinex в Гонконзі спричинив раптове падіння вартості біткоїна на 20%. Тому розуміння впливу новин, подій та інших факторів на курс криптовалют стає критично важливим для інвесторів та дослідників.

Метою цього дослідження є розробка системи для прогнозування курсів криптовалют на основі моделей машинного навчання, яка в кінцевому результаті буде надавати користувачеві інформацію про те чи варто купувати зараз криптовалюту чи ні. Основна мета полягає у створенні моделі, яка забезпечуватиме точні та корисні прогнози, сприятиме прийняттю обґрунтованих рішень інвесторами та була б достатньо автономною для використання з торговими ботами. Це важливе завдання, оскільки прогнозування курсу криптовалют може визначити успіх та ризики для інвесторів.

Об'єктом дослідження є модуль прогнозування фінансових ринків з акцентом на ринок криптовалют.

Предметом дослідження є методи та алгоритми машинного навчання, які можуть бути використані для прогнозування ціни біткоїна, а також аналіз існуючих рішень у цій області.

Для досягнення цієї мети були визначені наступні основні завдання дослідження:

- дослідження існуючих систем прогнозування та класифікації;
- огляд сучасних методів, що можуть підвищити ефективність

Прогнозування та класифікації;

- аналіз та вибір найкращого методу для підвищення ефективності;
- проєктування системи прогнозування криптовалют;
- програмна реалізація системи;

Практичне значення полягає в тому, що використання розробленої системи допоможе у прогнозуванні можливого курсу криптовалют.

1 АНАЛІЗ ЗАДАЧІ ПРОГНОЗУВАННЯ ЦІНИ БІТКОЇНА

1.1 Опис предметної області

Криптовалюта є підмножиною цифрової валюти, але вона стала основною цифровою валютою з дуже відмінними характеристиками від інших цифрових валют, які централізовано випускаються та обертаються в певному географічному місці. Будучи цифровими активами, базову цінність криптовалют важко зрозуміти. Ці криптовалюти є результатом різноманітних досягнень у багатьох галузях, таких як інформатика, економіка та криптографія.

Зростання популярності криптовалют добре задокументовано. Звіти показують, що в 2017 році кількість активно використовуваних криптовалютних гаманців коливалася від 2,9 до 5,8 мільйонів.[1]

Блокчейн — це технологія, що лежить в основі Ethereum, Bitcoin та інших криптовалют. Це мережа, яка використовує криптографію та децентралізований реєстр для обміну цифровими валютами або токенами. Децентралізація дозволяє кільком організаціям обмінюватися даними та забезпечувати безпеку та консенсус [2].

Ethereum — це технологія блокчейну другого покоління, яка спеціально розроблена як децентралізована платформа смарт-контрактів. [3] Модель створення депозитарію цінних паперів за технологією блокчейн була запропонована в 2016 році.

Криптовалюти надзвичайно корисні для створення валютних токенів, і такі депозити також вимагають моделі з відкритим кодом, децентралізованої системи, повної мови сценаріїв Turing, моделі контролю доступу та інших вимог. Крім того, Blockchain має прийняти галузеві стандарти та позиціонувати себе у фінансовій екосистемі, а також бути сумісним для взаємодії з іншими системами бухгалтерського обліку. [4]

Це важливе застосування, оскільки депозитарій цінних паперів є важливим рівнем, від якого залежить багато інших систем.

Технологія блокчейн має потенціал для покращення поточних процесів обробки інформації та даних фінансових установ, тому багато фінансових установ використовують цю технологію. Цю технологію також можна застосовувати в усій фінансовій системі, а не лише в одній установі. Зараз багато центральних банків розглядають можливість застосування блокчейну у фінансовій інфраструктурі. Наприклад, Південна Корея, Сінгапур та кілька інших країн зробили кроки до створення великомасштабних організацій, орієнтованих на блокчейн, а деякі організації інтегрували блокчейн для покращення своїх поточних бізнес-процесів [5].

Криптовалютний ринок характеризується значними й періодичними арбітражними можливостями між різними біржами. Ці відмінності в ціні більші, якщо розглядати різні країни, і менші між криптовалютами. Такі відмінності також зберігаються протягом кількох годин, а в деяких випадках можуть зберігатися протягом днів або навіть тижнів.

Криптовалютні ринки значно відрізняються від традиційних ринків тим, що на них немає жодних правил, які б гарантували, що інвестори отримають найкращі ціни під час торгівлі. Ось чому арбітражери настільки важливі на ринку криптовалют, оскільки вони можуть торгувати на всіх ринках. Але контроль капіталу ускладнює арбітражникам перевірку різних торгових стратегій. [6]

Кореляція між відмінностями між країнами є більшою для країн із суворим контролем капіталу та меншою для країн із відносно відкритою економікою. Це може означати, що інвестори надають більше значення криптовалютам у країнах із менш розвиненими ринками капіталу. Аналіз взаємозв'язку між чистим потоком замовлень і цінами на різні криптовалюти проливає світло на можливі причини цих цінових відмінностей. [7]

З кожним роком криптовалюти стають все більш популярними. Цьому сприяє розвиток комп'ютерних технологій. Сьогодні дистанційно працюють люди. Покупки в Інтернеті та оплата товарів в Інтернеті є звичайним явищем. Все це призвело до появи криптовалюти, яка була більш зручною в платежах. Ними користуються фізичні та юридичні особи з різними цілями. При цьому, якщо

раніше існувало варіанти криптовалюти, то зараз з'явилася досить велика кількість – варіанти. При виборі важливо вивчити характеристики, переваги та недоліки криптовалюти. Зараз багато людей мають потребу в покупці криптовалюти. [8]

Існує три основних способи отримання доходу з його допомогою:

- торгівля на спеціалізованих біржах;
- майніть криптовалюту за допомогою власної ферми;
- майнінг криптовалюти за допомогою хмарного майнінгу.

Майнінг сам по собі є дуже складним обчислювальним процесом. Це дозволило комп'ютерам, можливо розташованим у різних місцях, розв'язати математичні задачі, що призвело до створення біткоіна. У випадку гарантується висока безпека. Багато людей вирішують купити машину для майнінгу.

На сьогоднішній день такі пристрої популярні і бувають різних різновидів, але вони є кращим рішенням, ніж ферми на відеокарті, тому що приносять більше прибутку власнику. Необхідно скласти список валют, які будуть генеруватися за певним алгоритмом. [9]

На сьогоднішній день в секторі криптовалют найбільш затребуваним є хмарний майнінг. Він має такі важливі переваги:

- простий у використанні навіть новачкам.
- немає потреби у спеціальних знаннях, оскільки вам потрібно просто заплатити та почати заробляти;
- не потребує придбання відповідного обладнання та аксесуарів.

Загальні результати аналізу показали, що прогнозування ціни біткоіна є складною та багатоаспектною задачею, яка потребує врахування широкого спектра факторів, а саме прогнозування ціни біткоіна може бути ефективно здійснене за допомогою комплексного підходу, що поєднує технічний аналіз, фундаментальний аналіз та сучасні технології машинного навчання. Такий підхід дозволяє врахувати різноманітні фактори, що впливають на ринок криптовалют, та підвищити точність прогнозів. [10]

Таким чином, опис предметної області дозволив визначити основні чинники, які необхідно враховувати при побудові моделей прогнозування ціни біткоіна. Це

створює міцну основу для подальшої розробки ефективних прогнозних моделей, що можуть бути використані як інвесторами, так і фінансовими аналітиками для прийняття обґрунтованих рішень на ринку криптовалют.

1.2 Огляд і аналіз існуючих аналогів

WalletInvestor - Компанія пропонує коротко або довгостроковий прогноз курсу криптовалют на сайті <https://walletinvestor.com/forecast> (див рисунок 1.1). Методи, використані для прогнозу не розголошуються та є комерційними.

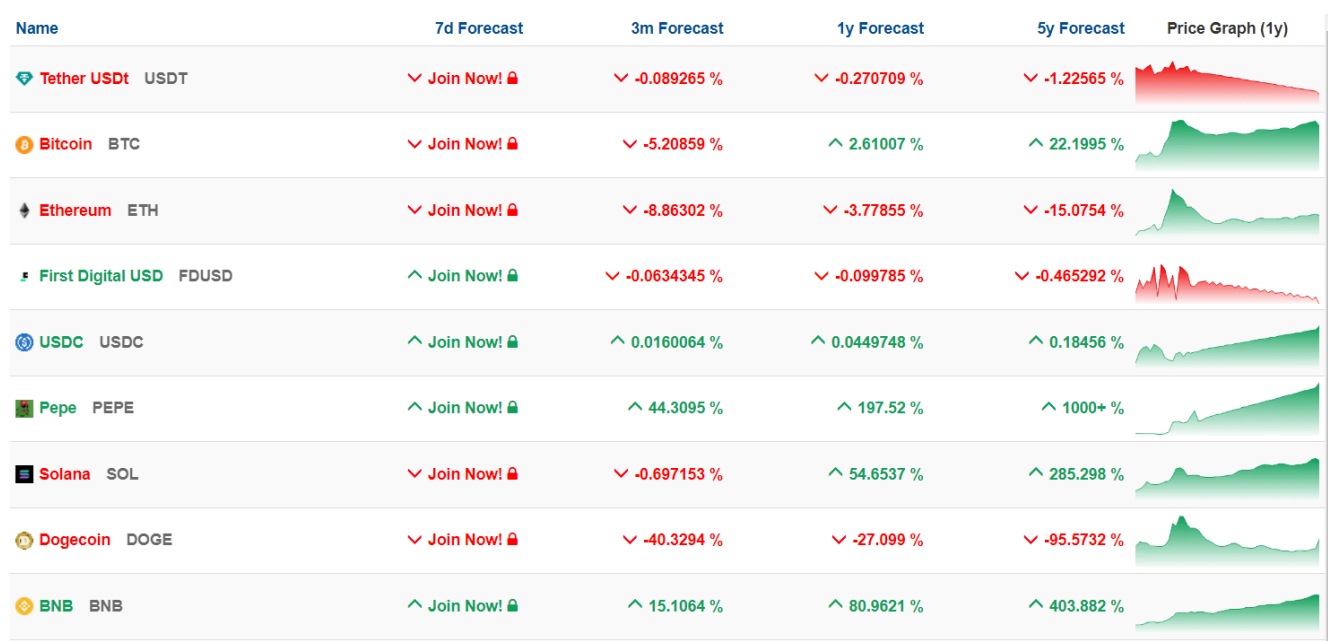


Рисунок. 1.1. – Вид сайту с прогнозами Walletinvestor [11]

На сторінці WalletInvestor (див рисунок 1.2) представлено прогнози вартості різних фінансових активів, таких як криптовалюти, акції, та інші інвестиційні інструменти. Користувачі можуть знайти коротко- та довгострокові прогнози, цінові цілі, а також підтримку та опір для різних активів.

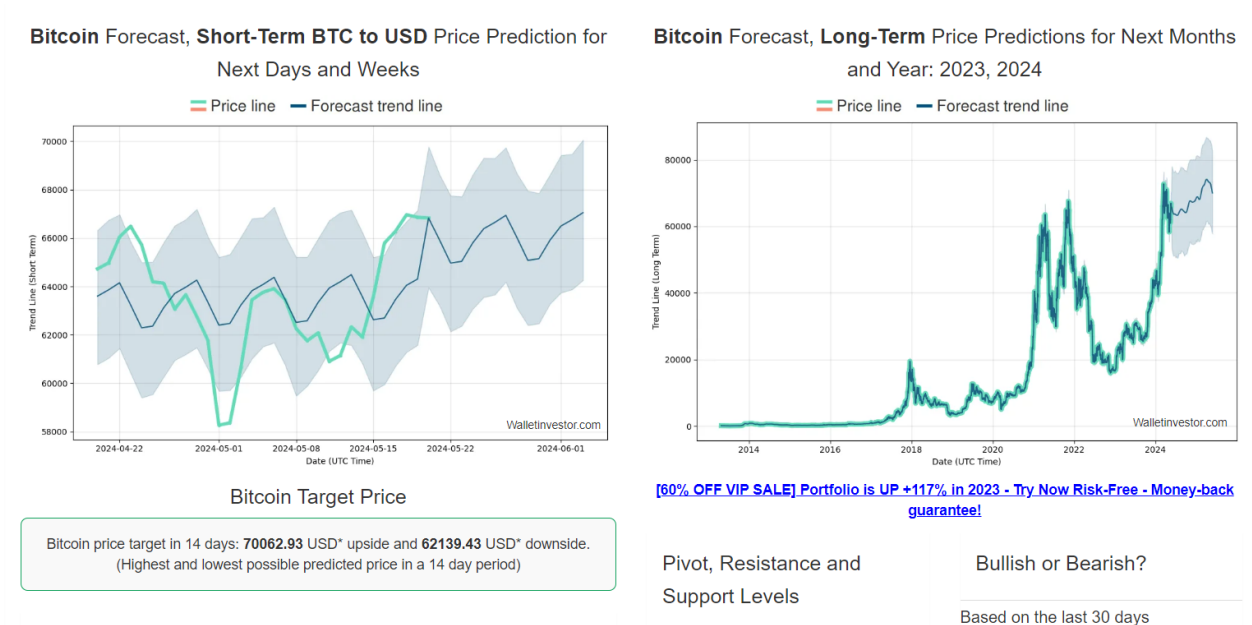


Рисунок. 1.2 – Вид сайту Walletinvestor з прогнозами стосовно Bitcoin [11]

Крім криптовалюти, сайт пропонує прогнози для акцій американського ринку, з використанням смарт-графіків та аналізом патернів (Walletinvestor.com). Також доступні функції конвертера валют та калькулятора, які дозволяють порівнювати ціни різних криптовалют відносно основних світових валют.

CoinCheckup - На сторінці CoinCheckup представлені актуальні ціни, графіки та ринкова капіталізація різних криптовалют (див рисунок 1.3).

#	Name	Last Price	24H	7D	Market Cap ↓	24h Volume	Supply	7D Chart
1	BTC Bitcoin	\$62,033.99	-1.01%	-2.46%	\$1.22 T	\$89.08 B	19.70 M 19.70 M	
2	ETH Ethereum	\$2,915.34	-1.94%	-5.27%	\$350.18 B	\$11.94 B	120.11 M 120.11 M	
3	USDT Tether	\$1.00	-0.04%	-0.05%	\$110.88 B	\$87.71 B	110.85 B 67.92 B	
4	BNB Binance Coin	\$581.60	-2.19%	-1.93%	\$85.84 B	\$1.36 B	147.59 M 147.59 M	
5	SOL Solana	\$146.63	0.79%	-5.60%	\$65.79 B	\$4.77 B	448.65 M 576.02 M	
6	USDC USD Coin	\$1.00	0.01%	0.01%	\$33.13 B	\$4.79 B	33.12 B 48.88 B	
7	XRP XRP	\$0.5062	0.03%	-5.68%	\$28.02 B	\$1.26 B	55.36 B 99.99 B	
8	TON Toncoin	\$6.76	-6.71%	14.09%	\$23.48 B	\$399.20 M	3.47 B 5.11 B	
9	DOGE Dogecoin	\$0.1533	3.65%	-2.25%	\$22.11 B	\$54.41 B	144.29 B 144.29 B	
10	ADA Cardano	\$0.4344	-2.53%	-3.62%	\$15.50 B	\$526.17 M	35.67 B 36.87 B	

Рисунок 1.3. – Вид сайту с прогнозами CoinCheckup [12]

Користувачі можуть знайти інформацію про найпопулярніші криптовалюти, нові монети, топ-отримувачі та втрати, а також прогнози та аналітику. Coincheckur також пропонує розділ з прогнозом на майбутнє, який ґрунтується на ситуації на ринку та оглядах з інших платформ.

CoinCheckur пропонує вичерпну інформацію про біткоїн, включаючи оновлення цін в режимі реального часу, історичні дані та детальний аналіз. Платформа надає користувачам інструменти для фундаментального та інвестиційного аналізу, а також прогнози щодо майбутніх показників біткоїна. Користувачі також можуть отримати доступ до новин, тенденцій та інформації, пов'язаної з біткоїном та іншими криптовалютами.

Крім того, CoinCheckur містить графіки та метрики, які допомагають інвесторам приймати обґрунтовані рішення. (див рисунок 1.4).

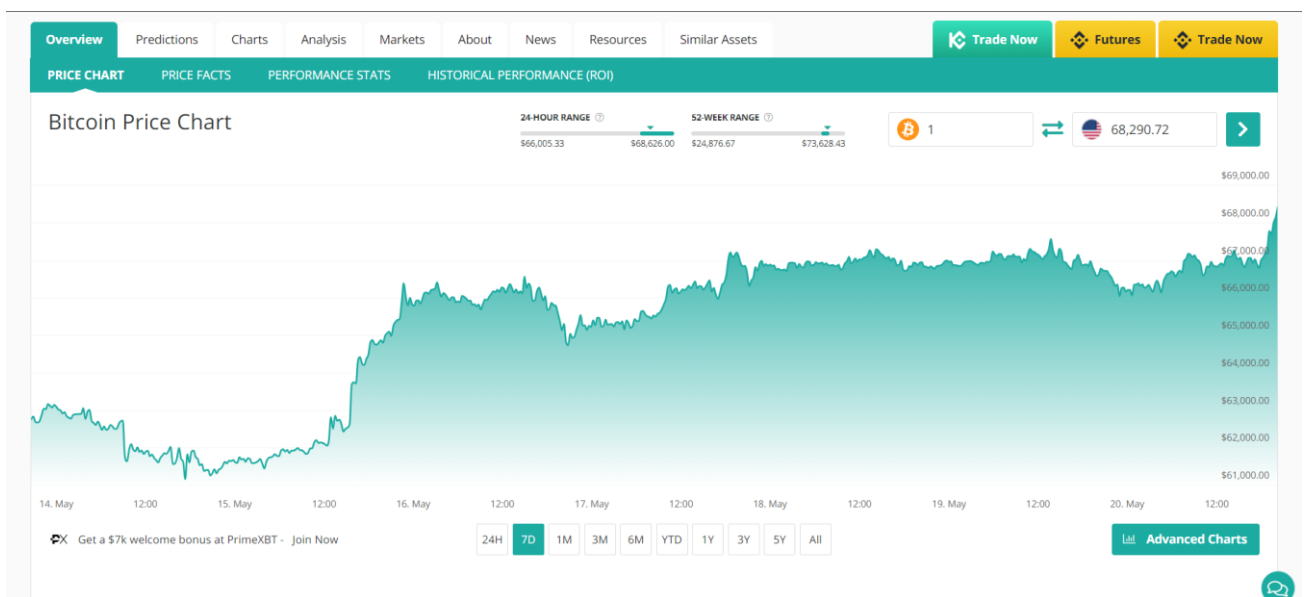
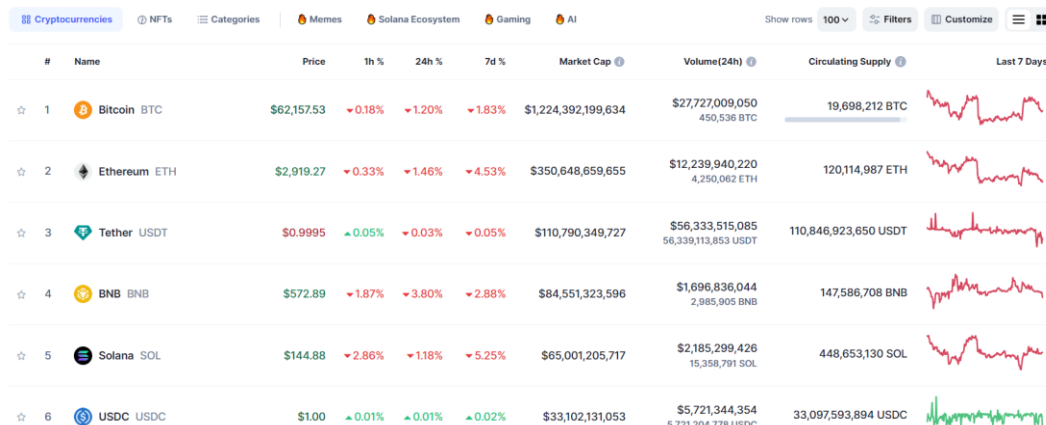


Рисунок. 1.4 – Вид сайту CoinCheckur з прогнозами стосовно Bitcoin [12]

CoinMarketCap - На сторінці CoinMarketCap представлена детальна інформація про криптовалюти, включаючи їх поточні та історичні ціни, ринкову капіталізацію, обсяги торгів. Сайт ранжує криптовалюти за ринковою капіталізацією і пропонує дані про трендові монети, найбільших зростаннях і падіннях (див рисунок 1.5). Також є розділи про криптовалютні біржі, NFTs, DeFi,

та смарт-контракти. Користувачі можуть знайти аналітику ринку, освітні ресурси та інструменти для конвертації цін.



#	Name	Price	1h %	24h %	7d %	Market Cap	Volume(24h)	Circulating Supply	Last 7 Days
1	Bitcoin BTC	\$62,157.53	▼0.18%	▼1.20%	▼1.83%	\$1,224,392,199,634	\$27,727,009,050 450,536 BTC	19,698,212 BTC	
2	Ethereum ETH	\$2,919.27	▼0.33%	▼1.46%	▼4.53%	\$350,648,659,655	\$12,239,940,220 4,250,062 ETH	120,114,987 ETH	
3	Tether USDT	\$0.9995	▲0.05%	▼0.03%	▼0.05%	\$110,790,349,727	\$56,333,515,085 56,339,113,853 USDT	110,846,923,650 USDT	
4	BNB BNB	\$572.89	▼1.87%	▼3.80%	▼2.88%	\$84,551,323,596	\$1,696,836,044 2,985,905 BNB	147,586,708 BNB	
5	Solana SOL	\$144.88	▼2.86%	▼1.18%	▼5.25%	\$65,001,205,717	\$2,185,299,426 15,358,791 SOL	448,653,130 SOL	
6	USDC USDC	\$1.00	▲0.01%	▲0.01%	▲0.02%	\$33,102,131,053	\$5,721,344,354 5,721,204,778 USDC	33,097,593,894 USDC	

Рисунок. 1.5. Вид сайту с прогнозами CoinMarketCap [13]

Cryptowatch - На сторінці Cryptowatch представлений професійний інтерфейс для торгівлі криптовалютами, зокрема парою BTC/USD. Інтерфейс включає живий графік цін на біткоїн з можливістю налаштування індикаторів та різних таймфреймів, реєстр замовлень (Order Book) з відображенням поточних заявок на купівлю та продаж, вікно для розміщення ринкових, лімітних та інших типів замовлень, огляд останніх угод (Trade History) та інформація про баланс та відкриті замовлення користувача (див рисунок 1.6).

Це дозволяє трейдерам проводити детальний технічний аналіз і здійснювати операції в реальному часі.

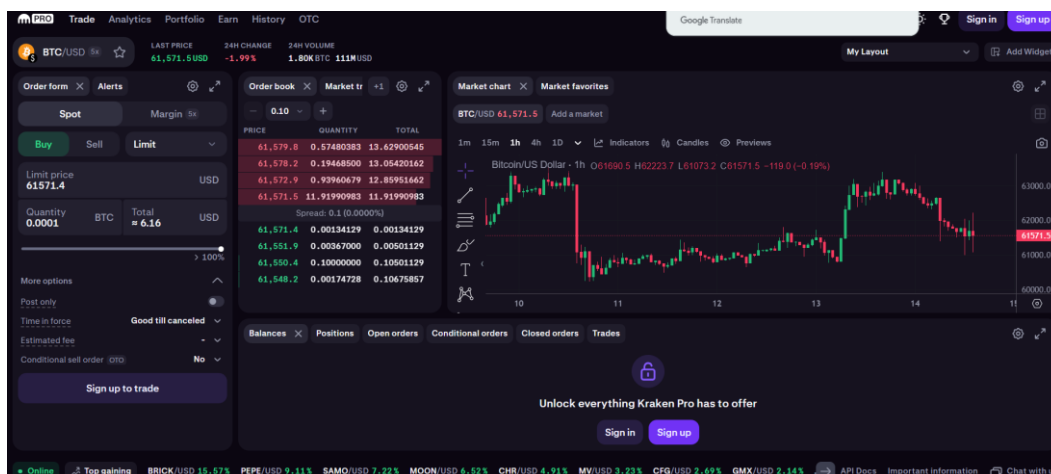


Рисунок. 1.6. - Вид сайту с прогнозами Cryptowatch [14]

CryptoCompare - це комплексна платформа для аналізу криптовалютних даних. Вона надає актуальні та історичні ринкові дані про різні криптовалюти, включаючи графіки цін, ринкову капіталізацію, обсяги торгів та біржові рейтинги (див рисунок 1.7).

Сайт також пропонує інструменти для управління портфелем, калькулятори для майнінгу та оновлення новин. Крім того, на CryptoCompare є форум для обговорення спільноти, посібники для початківців та детальні огляди гаманців, бірж і обладнання для майнінгу. На платформі є безкоштовне API з унікальними даними. Всі дані платформа обробляє в онлайн режимі з сайтів найбільших криптовалютних бірж.

Це цінний ресурс як для початківців, так і для досвідчених ентузіастів криптовалют, інвесторів та аналітиків.

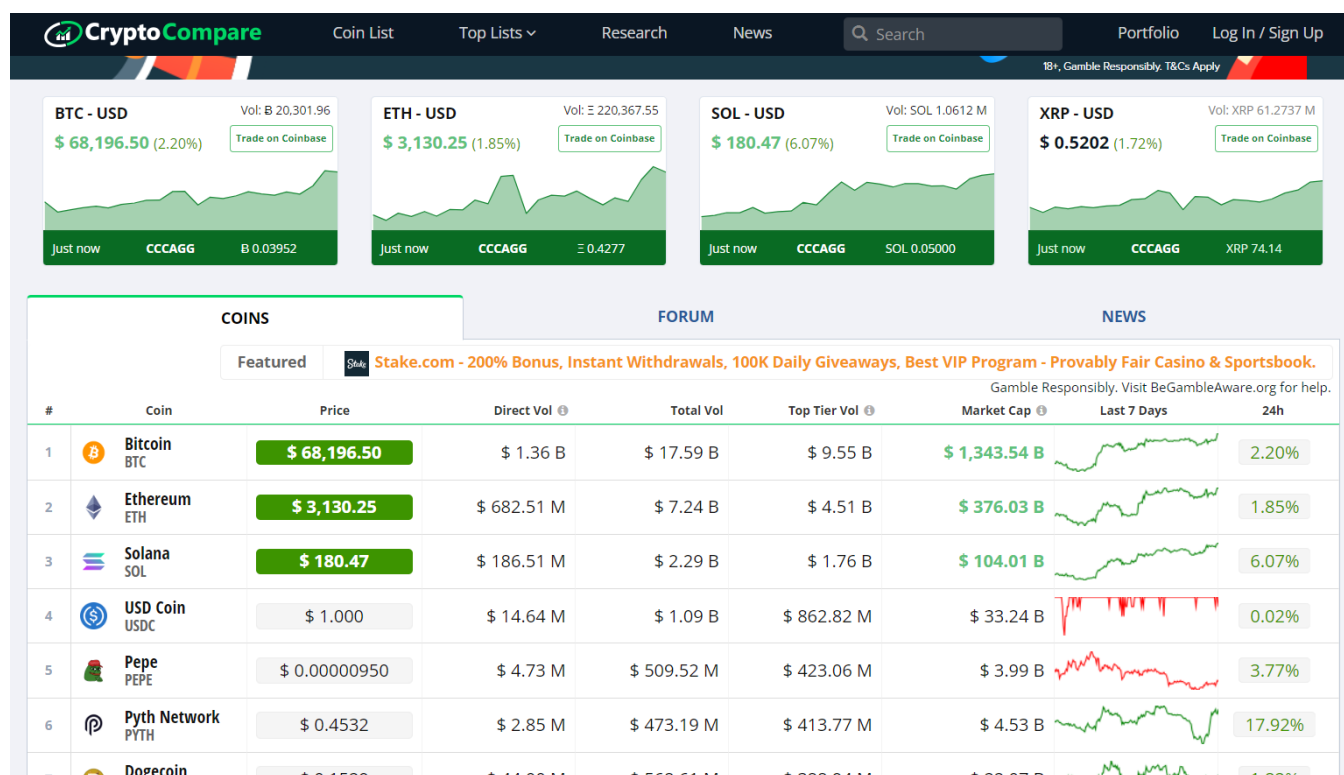


Рисунок. 1.7 - Вид сайту с прогнозами CryptoCompare [15]

Для кращого розуміння функціональних можливостей та відмінностей доступних інструментів для прогнозування цін на криптовалюту, у таблиці 1.1 наведено порівняльну характеристику деяких із них.

Таблиця 1.1 – Аналіз існуючих сервісів для прогнозування цін на криптовалюту

Критерій	WalletInvestor	CoinCheckup	CoinMarketCap	Cryptowatch	CryptoCompare
Прогнозування цін (в доларах)	Так	Так	Ні (прогнозує в процентному відношенні)	Так	Так
Аналіз монет	Ні	Так, ґрунтовний аналіз з оцінками за різними факторами.	Так, ґрунтовний аналіз з рейтингами та оцінками.	Так, ґрунтовний аналіз з рейтингами та оцінками.	Так, ґрунтовний аналіз з рейтингами та оцінками.
Дані про ринок	Так, дані про ціни, обсяги торгів, ринкову капіталізацію та домінування.	Так, дані про ціни, обсяги торгів, ринкову капіталізацію та домінування.	Так, дані про ціни, обсяги торгів, ринкову капіталізацію та домінування.	Так, дані про ціни, обсяги торгів, ринкову капіталізацію та домінування.	Так, дані про ціни, обсяги торгів, ринкову капіталізацію та домінування.
Інструменти та ресурси	Ні	Так, калькулятор прибутку, конвертер валют та інші інструменти.	Так, калькулятор прибутку, конвертер валют та інші інструменти.	Так, калькулятор прибутку, конвертер валют та інші інструменти.	Так, включають інструменти для порівняння бірж та трейдингу
Безкоштовні та платні функції	Безкоштовні функції.	Безкоштовні та платні функції.	Безкоштовні та платні функції.	Безкоштовні та платні функції.	Безкоштовні та платні функції.
Мови	Англійська.	Англійська.	Англійська та інші мови.	Англійська та інші мови.	Англійська та інші мови.

Провівши аналіз існуючих сервісів для прогнозування цін на криптовалюту, було виведено наступні висновки:

WalletInvestor: Якщо вам потрібні прогнози цін на криптовалюту, WalletInvestor може бути хорошим вибором.

CoinCheckup: Якщо вам потрібен ґрунтовний аналіз монет з оцінками та рейтингами.

CoinMarketCap: Якщо вам потрібні дані про ринок, такі як ціни, обсяги торгів та ринкова капіталізація.

Cryptowatch: Якщо вам потрібні дані про ринок, ґрунтовний аналіз монет, новини та події, а також інструменти та ресурси.

CryptoCompare: Якщо вам потрібне порівняння цін та бірж, а також калькулятори

Важливо зазначити, що жоден з цих сайтів не гарантує точні прогнози цін.

Інвестори та трейдери повинні завжди проводити власні дослідження, використовувати різні методи прогнозування та диверсифікувати свої портфелі, щоб мінімізувати ризики.

1.3 Постановка задачі дослідження

Прогнозування цін на Bitcoin та інші криптовалюти є складною задачею через їхню високу волатильність та динамічний характер. Існуючі методи прогнозування часто не беруть до уваги широкий спектр інформаційних факторів, які можуть впливати на курс криптовалют, таких як історичні ціни, обсяги торгів, новини та соціальні мережі.

Ця робота пропонує новий підхід до прогнозування цін на Bitcoin, який ґрунтується на комплексному аналізі інформаційних факторів. Дослідження буде охоплювати наступні аспекти:

Визначення інформаційних факторів: Визначити та дослідити широкий спектр інформаційних факторів, які можуть впливати на курс Bitcoin, включаючи історичні ціни, обсяги торгів, новини, твіти, настрої в соціальних мережах та інші дані.

Збір та обробка даних: Розробити та реалізувати методи збору та обробки даних з різних джерел, включаючи веб-сайти, API та соціальні мережі.

Аналіз даних: Застосувати методи аналізу даних, такі як візуалізація, статистичний аналіз та машинне навчання, для виявлення закономірностей та зв'язків між інформаційними факторами та курсом Bitcoin.

Розробка моделей прогнозування: Розробити та оцінити різні моделі машинного навчання для прогнозування цін на Bitcoin, беручи до уваги інформаційні фактори.

Інтерпретація та оцінка результатів: Проаналізувати та інтерпретувати результати моделей прогнозування, оцінити їхню точність та стійкість.

Визначення напрямків для подальших досліджень: Визначити напрямки для подальших досліджень, спрямованих на вдосконалення методів прогнозування цін на Bitcoin та інші криптовалюти.

Ця робота відрізняється від інших досліджень прогнозування цін на Bitcoin тим, що вона:

- Використовує комплексну модель інформаційних факторів: Дослідження буде охоплювати широкий спектр інформаційних факторів, які не завжди враховуються в інших методах прогнозування.
- Застосовує передові методи аналізу даних: Дослідження буде використовувати передові методи аналізу даних, такі як машинне навчання, для виявлення закономірностей та зв'язків між інформаційними факторами та курсом Bitcoin.
- Фокусується на інтерпретації та оцінці результатів: Дослідження буде не лише розробляти моделі прогнозування, але й інтерпретувати їхні результати, оцінювати їхню точність та стійкість.

2 АЛГОРИТМІЧНЕ ТА ІНФОРМАЦІЙНЕ ЗАБЕЗПЕЧЕННЯ МОДУЛЯ

2.1 Концептуальні основи вирішення задачі прогнозування ціни біткоїна з використанням машинного навчання

У цьому розділі розглядаються концептуальні основи вирішення завдання щодо створення модуля прогнозування ціни біткоїна з використанням XGBoost. Основну увагу буде приділено структурно-орієнтованому підходу, який дозволяє детально аналізувати і декомпонувати систему на функціональні частини. В даному розділі буде представлена функціональна структура модуля, наведено перелік функцій та їх характеристики. Також будуть описані інформаційні взаємозв'язки між елементами системи та зв'язки з іншими системами і зовнішніми об'єктами, що слугують джерелами або споживачами інформації.

Спроектowana діаграма DFD описує структуру модуля для проекту прогнозування ціни біткоїна з використанням алгоритму XGBoost (eXtreme Gradient Boosting) (див рис 2.1).

Цей модуль буде використовуватися для збору та обробки історичних даних про ціну біткоїна, навчання та оцінки моделі XGBoost, а також для прогнозування майбутніх цін.

Для того щоб користувач отримав всій прогноз потрібно натренувати модель на наявних історичних даних. Для того щоб мати змогу використовувати ці дані з найбільшою ефективністю дані потрібно обробити, цей процес включає в себе очистку не корисних даних, виявлення аномалій та інші кроки. Після чого потрібно витягнути основну інформацію яка характеризує надані дані а також розділити їх на тренувальні, тестові та валідаційні, для виконання наступного кроку який включатиме перевірку точності прогнозування навченої моделі. Після чого перевірену модель з хорошою якістю можна буде використовувати для створення прогнозів на ціну для Біткоїна та надавати ці дані користувачеві. [16]

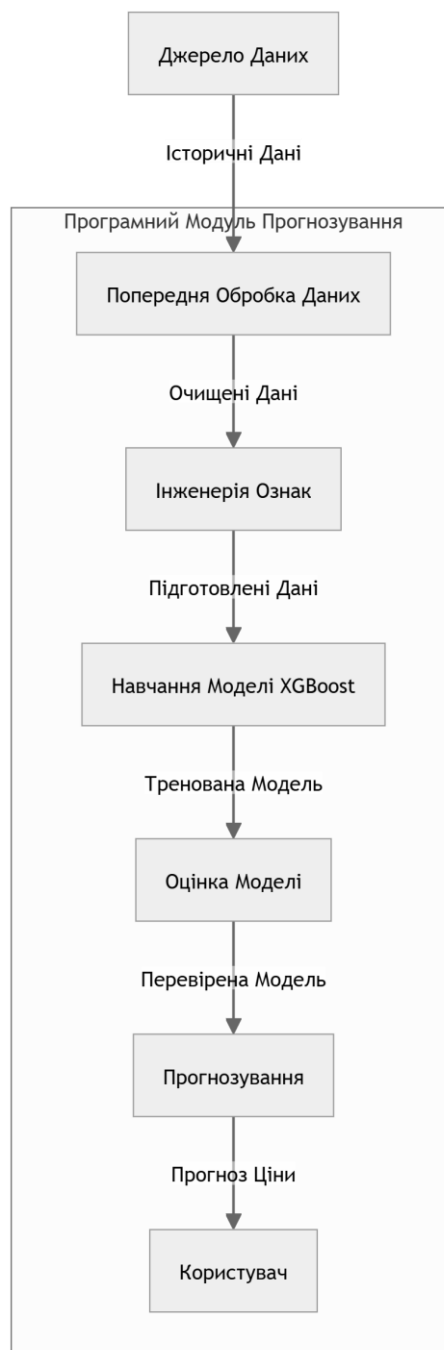


Рисунок 2.1 – Діаграма потоків даних (DFD)

Поведінка системи в часі для прогнозування ціни біткоїна з XGBoost:

Алгоритм прогнозування ціни біткоїна з використанням XGBoost демонструє динамічну поведінку протягом часу, залежно від кількох факторів.

Навчання та оновлення моделі

На початковому етапі алгоритм навчається на історичних даних про ціну біткоїна. Цей процес може бути тривалим, тому прогнози на початкових етапах

можуть бути не дуже точними. З часом, коли модель буде навчатися на більшій кількості даних, прогнози стануть більш достовірними.

Змінність ринку біткоїна

Ринок біткоїна характеризується високою мінливістю, ціна може різко змінюватися протягом короткого періоду часу. Алгоритм прогнозування намагається адаптуватися до цих змін, але може знадобитися деякий час, щоб модель оновилася та почала робити точні прогнози в нових умовах.

Зовнішні фактори.

На ціну біткоїна можуть впливати й інші фактори, крім історичних даних, такі як економічні події, регуляторні зміни та новини. Алгоритм прогнозування не може враховувати всі ці фактори, тому прогнози можуть бути неточними, якщо відбуваються значні зовнішні події.

Приблизний опис поведінки системи:

- Короткостроковий: Прогнози можуть бути менш точними протягом перших кількох днів або тижнів після навчання моделі або після значних змін на ринку.
- Середньостроковий: З часом, коли модель буде навчатися на більшій кількості даних та адаптуватися до змін ринку, прогнози стануть більш точними.
- Довгостроковий: Точність прогнозів буде залежати від стійкості ринку біткоїна, а також від здатності алгоритму адаптуватися до нових умов та вчитися на нових даних.

2.2 Алгоритмічне забезпечення

Алгоритмічне забезпечення відіграє ключову роль у процесі створення ефективної системи прогнозування цін на Біткойн. Даний підрозділ присвячений детальному опису алгоритму, що використовується для тренування моделі на основі історичних даних про ціну криптовалюти (див. рис. 2.2) та алгоритму передбачення ціни на основі даних користувача (див. рис. 2.3).

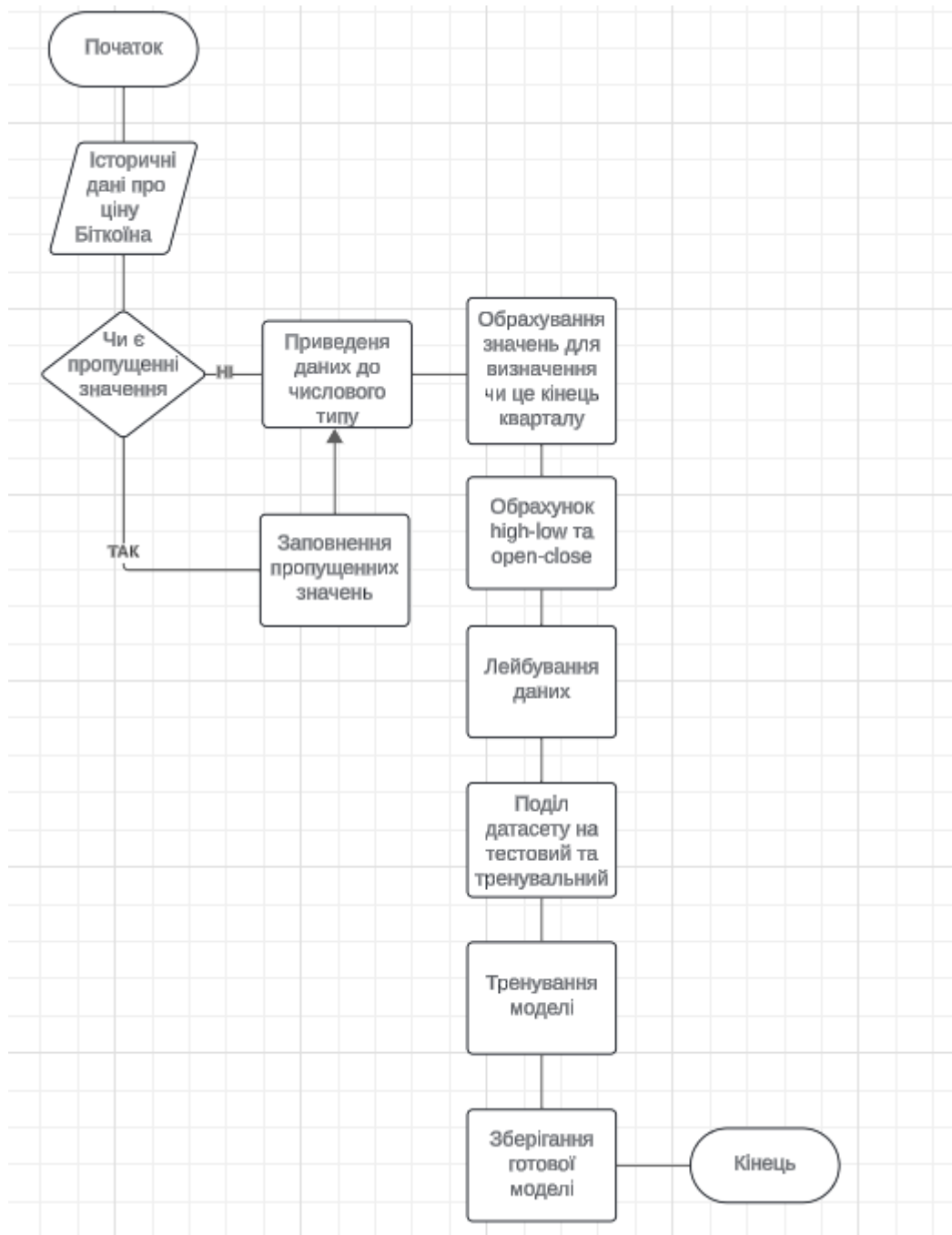


Рисунок 2.2 – Схема алгоритму тренування моделі

Алгоритм починається з завантаження історичних даних про ціну Біткойна з CSV-файлу які були отримані з відкритих джерел. Для попередньої обробки даних будуть використовуватися бібліотеки `pandas` і `numpy`, які дозволять ефективно очистити та нормалізувати дані. Далі виконується перетворення стовпця 'timestamp' до формату дати та додавання нового стовпця 'date', що містить дату в строковому форматі. Всі дані сортуються за датою у висхідному порядку для подальшого аналізу.

Якщо є пропущені значення у даних, вони заповнюються середніми значеннями. Після цього, дані перетворюються до числового типу, зокрема, стовпці 'open', 'high', 'low', 'close' конвертуються до числового формату. Для визначення кінця кварталу, додається нова характеристика 'is_quarter_end', яка вказує, чи є місяць кінцем кварталу.

Після цього, виконується обчислення додаткових характеристик: 'open-close', яка обчислює різницю між відкриттям та закриттям ціни, та 'low-high', яка обчислює різницю між мінімальною та максимальною ціною. Дані лейбуються, тобто створюється новий стовпець 'target', який вказує, чи зросте ціна наступного дня, порівняно з поточним днем.

Для подальшого аналізу, дані розділяються на тестовий та тренувальний датасети з використанням функції `train_test_split`, при цьому використовуються 80% даних для тренування і 20% для тестування. Перед поділом дані нормалізуються за допомогою `StandardScaler`, що допомагає зробити моделі більш точними.

Далі виконується тренування моделі. Основним алгоритмом прогнозування було обрано модель XGBoost з використанням бібліотеки `xgboost`. XGBoost (Extreme Gradient Boosting) є потужним інструментом для задач регресії та класифікації, який заснований на методі градієнтного бустінгу. Градієнтний бустінг комбінує результати декількох слабких моделей, переважно дерев рішень, для формування єдиної сильної прогностичної моделі. Цей підхід дозволяє значно підвищити точність прогнозів. Навчання моделі XGBoost включатиме налаштування різних гіперпараметрів, таких як кількість дерев, глибина дерев, швидкість навчання та інші параметри, які впливають на продуктивність моделі. [17-19]

Після налаштування гіперпараметрів модель буде тренуватися на навчальній вибірці. Процес навчання XGBoost полягає у послідовному побудові дерев рішень, де кожне нове дерево коригує помилки попередніх. На кожній ітерації алгоритм намагається зменшити залишкову помилку за рахунок побудови нового дерева, яке прогнозує ці залишки. Це дозволяє моделі поступово покращувати свої прогнози.

А також оцінюється точність моделі як на тренувальному, так і на тестовому наборі даних з використанням метрики `roc_auc_score`.

Попередній алгоритм описував тренування моделі на історичних даних про ціну Біткойна (див. рис. 2.2). Тепер, коли модель натренована і готова до використання, настає черга її застосування для реального прогнозування цін на основі нових даних, які вводить користувач. Перехід від етапу тренування моделі до етапу її застосування є критично важливим, оскільки саме на цьому етапі результати моделі стають корисними для кінцевих користувачів. Тож на наступному рисунку ми можемо розглянути блок-схему цього алгоритму (див. рис. 2.3).



Рисунок 2.3 – Алгоритм прогнозування ціни на основі даних користувача

На початку алгоритму відбувається завантаження та попередня обробка даних. Зокрема, введення даних, таких як close, open, high, low, та data, є важливим етапом, де всі необхідні параметри збираються для подальшого аналізу.

Наступним кроком є обчислення нових характеристик, які включають різницю між цінами відкриття та закриття (close-open) і різницю між найвищою та найнижчою цінами за день (high-low). Ці характеристики допомагають краще зрозуміти внутрішню динаміку ринку. Також проводиться визначення на основі введеної дати, чи є кінець кварталу, що може мати значний вплив на коливання цін.

Після попередньої обробки даних здійснюється завантаження наперед натренованої моделі. Детальний опис алгоритму тренування цієї моделі зображений на рис. 2.2. Завантажена модель потім використовується для класифікації і прогнозування ціни Біткойна на основі вхідних даних.

Завершується алгоритм виведенням результатів користувачеві. Модель класифікує, чи зросте ціна Біткойна, чи впаде, і цей результат виводиться користувачеві для прийняття подальших рішень. Використання комплексного підходу до обробки та аналізу даних забезпечує високу точність моделі, що є критично важливим для успішного прогнозування та прийняття обґрунтованих рішень на основі отриманих результатів.

Переваги та недоліки алгоритму XGBoost.

Переваги:

- Алгоритм XGBoost є точним та ефективним методом прогнозування.
- Він може обробляти великі набори даних з високою швидкістю.
- Він може бути налаштований за допомогою різних гіперпараметрів для покращення його продуктивності.

Недоліки:

- Алгоритм XGBoost може бути складним для налаштування.
- Він може бути чутливим до шуму в даних.

Діаграма переходів станів (STD - State Transition Diagram) наочно демонструє послідовність і взаємозв'язок між цими етапами (див рис. 2.2). Вона є ключовим елементом проектування системи, оскільки дозволяє візуалізувати

процеси і стани, через які проходять дані, починаючи від їхнього завантаження і закінчуючи отриманням прогнозів. Нижче подано опис структури STD діаграми, яка відображає алгоритмічне забезпечення модуля прогнозування ціни біткоїна.

2.3 Інформаційне забезпечення

Алгоритм використовує історичні дані Bitcoin для навчання моделі машинного навчання з метою передбачення майбутніх цін. Вхідні дані для навчання моделі включають історичні дані Bitcoin, які отримуються з файлу CSV під назвою "bitcoin.csv". Вхідними даними для тренування моделі є історичні дані про торгівлю біткоїнами, які включають такі параметри, як відкриття (open), закриття (close), найвища (high) та найнижча (low) ціни за певний період часу, а також дата (date) кожного запису. Ці дані використовуються для обчислення додаткових показників, таких як різниця між ціною відкриття і закриття (open-close) та різниця між найвищою і найнижчою цінами (low-high). Крім того, вхідні дані включають показник, чи є дата кінцем кварталу (is_quarter_end).

Такі ж дані користувач має ввести для того щоб отримати прогноз щодо майбутньої ціни Біткоїна.

Вихідними даними є прогноз зміни ціни біткоїна на основі введених користувачем параметрів. Модель класифікує, чи буде ціна падати чи рости, тобто вона повертає значення 1 або 0. На основі цього результату надається текстове повідомлення про прогнозовану зміну ціни криптовалюти. Якщо модель повертає 1, це означає, що ціна зросте, і виводиться повідомлення "The price of bitcoin will increase". Якщо модель повертає 0, це означає, що ціна впаде, і виводиться повідомлення "Bitcoin price will fall".

3 ПРОГРАМНО-ТЕХНОЛОГІЧНЕ ЗАБЕЗПЕЧЕННЯ МОДУЛЯ

3.1 Реалізація програмного забезпечення

У цьому розділі буде розглянуто процес розробки архітектури для програмного модуля, призначеного для прогнозування ціни біткойна за допомогою машинного навчання на мові програмування Python з використанням бібліотеки XGBoost та розроблену UML-діаграму класів.

На наступному рисунку 3.1 зображена UML-діаграмі класів в якій представлена структурна організація програмної системи для прогнозування ціни біткойна з використанням XGBoost. Діаграма включає чотири основні класи: HistoricalData, Feature, Model та Prediction.

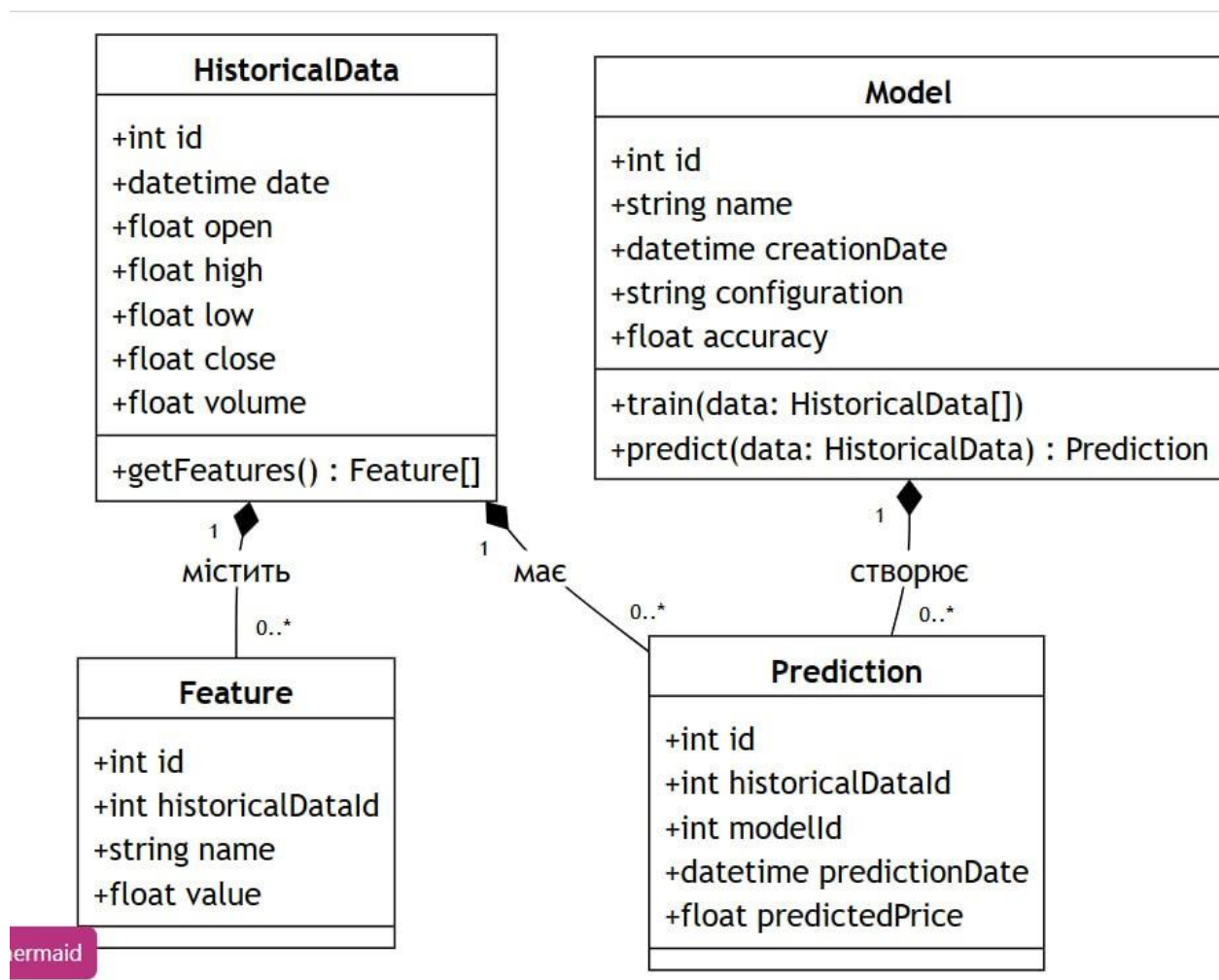


Рисунок 3.1 - UML-діаграмі класів

Клас `HistoricalData` представляє історичні дані про біткоїн. Він містить такі атрибути, як `id` (ідентифікатор), `date` (дата у форматі `datetime`), `open` (початкова ціна), `high` (найвища ціна), `low` (найнижча ціна), `close` (ціна на закриття) та `volume` (обсяг торгів). Цей клас також містить метод `getFeatures()`, який повертає масив об'єктів типу `Feature`. Цей клас використовується для зберігання і обробки історичних даних про біткоїн.

Клас `Feature` представляє окрему ознаку, яка буде використовуватися для моделювання. Він містить такі атрибути, як `id` (ідентифікатор), `historicalDataId` (ідентифікатор відповідного запису з класу `HistoricalData`), `name` (назва ознаки) та `value` (значення ознаки). Цей клас дозволяє деталізувати та зберігати окремі ознаки, які будуть використовуватися для тренування моделі.

Клас `Model` представляє модель для прогнозування ціни біткоїна. Він містить такі атрибути, як `id` (ідентифікатор), `name` (назва моделі), `creationDate` (дата створення у форматі `datetime`), `configuration` (конфігурація моделі) та `accuracy` (точність моделі). Цей клас також містить два методи: `train(data: HistoricalData[])`, який використовується для тренування моделі на основі масиву історичних даних, та `predict (data: HistoricalData): Prediction`, який використовується для прогнозування ціни на основі історичних даних. Цей клас є основним для реалізації алгоритму `XGBoost` та забезпечує всі необхідні функції для тренування і використання моделі.

Клас `Prediction` представляє прогнозовані дані. Він містить такі атрибути, як `id` (ідентифікатор), `historicalDataId` (ідентифікатор відповідного запису з класу `HistoricalData`), `modelId` (ідентифікатор моделі, яка здійснила прогноз), `predictionDate` (дата прогнозу у форматі `datetime`) та `predictedPrice` (прогнозована ціна). Цей клас дозволяє зберігати результати прогнозування, забезпечуючи можливість аналізу та візуалізації прогнозованих даних.

Зв'язки між класами на діаграмі відображають їхні взаємодії. Клас `HistoricalData` містить один або більше об'єктів класу `Feature`, що відображає процес екстракції ознак з історичних даних. Клас `Model` використовує дані класу `HistoricalData` для тренування і створює об'єкти класу `Prediction` на основі

прогнозування. Таким чином, ця діаграма UML класів відображає структурну організацію компонентів програмної системи, їхні атрибути та методи, а також інформаційні зв'язки між ними, що дозволяє забезпечити ефективне прогнозування цін на біткоїн з використанням XGBoost. Такий підхід забезпечує чітке уявлення про архітектуру програмного забезпечення та взаємодію між різними компонентами системи. Наступним кроком є перехід до структурної схеми програмної системи, яка надасть нам глибше розуміння про взаємодію елементів модуля на більш високому рівні.

Структурна схема програмної системи для реалізації моделі прогнозування ціни біткоїна з використанням XGBoost включає декілька основних елементів: збирання даних, попередня обробка даних, поділ даних на тренувальну та тестову вибірки, побудова та тренування моделі, оцінка моделі, прогнозування та візуалізація результатів.

В цьому розділі розглянемо детальніше основні елементи реалізації які стосуються роботи з даними.

Збирання даних (Data Collection) є першим етапом системи і включає збір історичних даних про ціну біткоїна з відкритих джерел, таких як фінансові API або CSV файли. Цей етап є критичним для забезпечення наявності достатнього обсягу даних для моделювання. В даному випадку у нас є наступний датасет з [Kaggle](#) за 2017-2023 роки. Який містить 3 126 000 рядків даних (див рис. 3.2) [20]

	timestamp	open	high	low	close	volume	date
0	2023-08-01 13:19:00	28902.48	28902.49	28902.48	28902.49	4.686580	2023-08-01
1	2023-08-01 13:18:00	28902.48	28902.49	28902.48	28902.49	4.775890	2023-08-01
2	2023-08-01 13:17:00	28908.52	28908.53	28902.48	28902.49	11.522630	2023-08-01
3	2023-08-01 13:16:00	28907.41	28912.74	28907.41	28908.53	15.896100	2023-08-01
4	2023-08-01 13:15:00	28896.00	28907.42	28893.03	28907.41	37.746570	2023-08-01
...	...	...	...	...	...	...	...
3125995	2017-08-17 04:04:00	4261.48	4261.48	4261.48	4261.48	0.140796	2017-08-17
3125996	2017-08-17 04:03:00	4261.48	4261.48	4261.48	4261.48	0.012008	2017-08-17
3125997	2017-08-17 04:02:00	4280.56	4280.56	4280.56	4280.56	0.261074	2017-08-17
3125998	2017-08-17 04:01:00	4261.48	4261.48	4261.48	4261.48	0.000000	2017-08-17
3125999	2017-08-17 04:00:00	4261.48	4261.48	4261.48	4261.48	1.775183	2017-08-17
3126000 rows × 7 columns							

Рисунок 3.2 – огляд датасету для тренування

Значення кожної колонки в датасеті:

- Timestamp – часова мітка (дата та час логуювання зміни ціни)
- Open - це ціна відкриття часового періоду
- High- це найвища ціна за період часу
- Low - це найнижча ціна за період часу
- Close - це ціна закриття часового періоду
- Volume - скільки одиниць Біткоїна було куплено або продано за цей період
- Date – день логуювання

Для кращого розуміння даних які у нас є варто побудувати графік змін ціни Біткоїна за цей період. (див рис. 3.3).

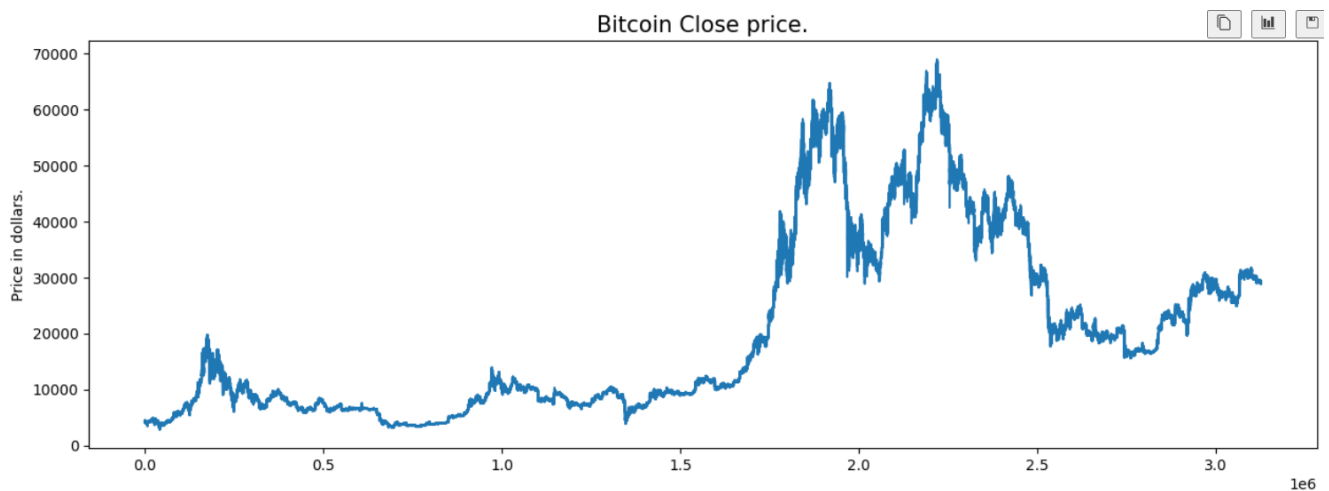


Рисунок 3.3 – графічне подання зміни ціни Біткоїна за 2017-2023 роки

Для визначення року з найвищим зростанням ціни Bitcoin за цей період 2017-2023 років можна розрахувати середнє значення ціни за кожен рік. Цей аналіз допоможе чітко ідентифікувати період, коли спостерігалось найбільш стрімке зростання. (див рис. 3.4).

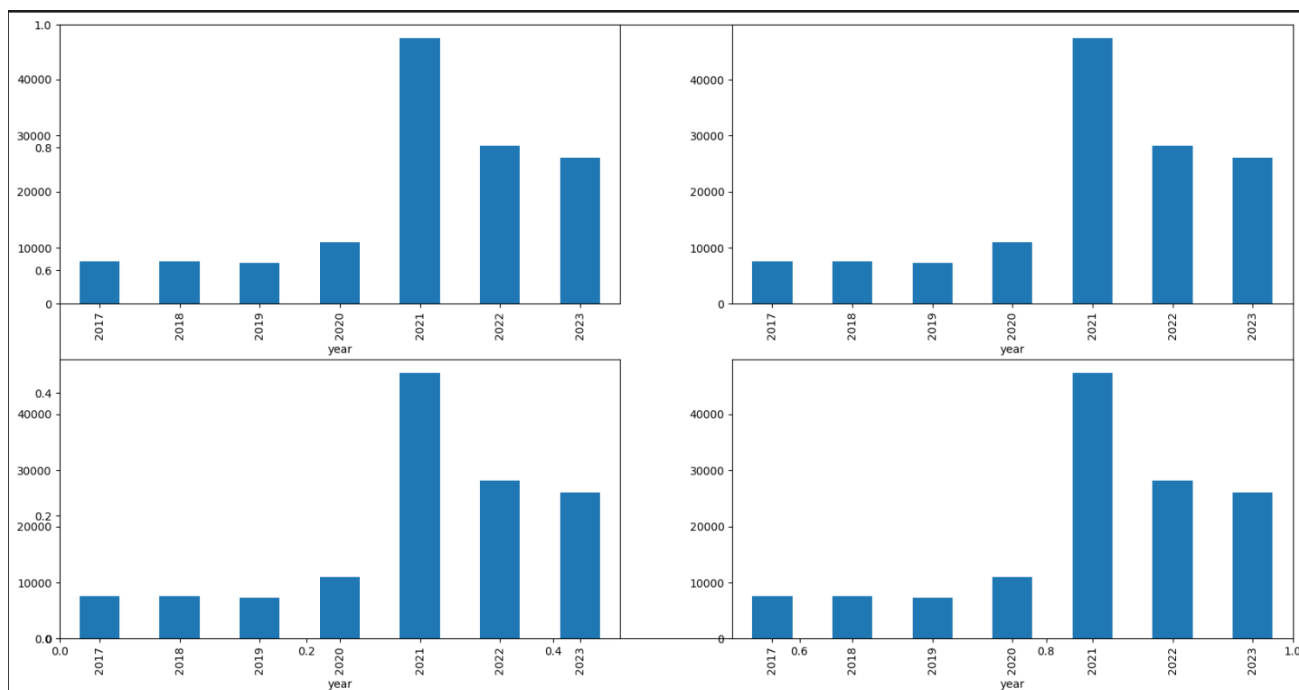


Рисунок 3.4 – середні значення ціни Біткоїна по кожному році.

Як ми можемо помітити з даних графіка то цей скачок ціни відбувся в 2021 року. І якщо ми перевіримо інформацію з відкритих джерел, то дізнаємось що дійсно в 2021 році ціна Біткоїна стрімко зросла в зв'язку з пом'якшенням грошово-кредитної політики Китаєм, через що збільшуються покупки активів, зокрема криптовалюти. [21]

На основі наявних даних створимо 3 нові колонки: 'open-close' та low-high' нові колонки в датафреймі, які мають важливе значення для аналізу та моделювання даних. Ось пояснення для кожної з цих колонок:

1. 'open-close': Ця колонка містить різницю між ціною відкриття (Open) та ціною закриття (Close) для кожного торгового дня. Значення цієї колонки показує, наскільки ціна змінилася протягом дня. Якщо значення додатне, то ціна закриття нижча за ціну відкриття, що може вказувати на зниження ціни протягом дня. Якщо значення від'ємне, то ціна закриття вища за ціну відкриття, що може свідчити про зростання ціни.
2. 'low-high': Ця колонка містить різницю між найнижчою (Low) та найвищою (High) ціною протягом торгового дня. Значення цієї колонки показує діапазон коливання цін за день. Це може бути корисним для аналізу

волатильності ринку, оскільки більші значення вказують на більші коливання цін.

3. Колонка `is_quarter_end` використовується для позначення, чи є даний запис кінцем кварталу. Кварталами вважаються періоди по три місяці, і, відповідно, кінець кварталу припадає на останній місяць кожного кварталу (березень, червень, вересень і грудень).

Ці три нові колонки надають важливу інформацію для аналізу та моделювання, допомагаючи краще розуміти поведінку цін на ринку та робити точніші прогнози.

Додатково, зробимо візуалізацію кореляції між різними змінними в датасеті для виявлення можливих взаємозв'язків, які можуть бути корисними для побудови моделі прогнозування (див. рис. 3.5). Ці кроки допоможуть забезпечити надійну та точну підготовку даних для подальшого етапу побудови та навчання моделі прогнозування цін.

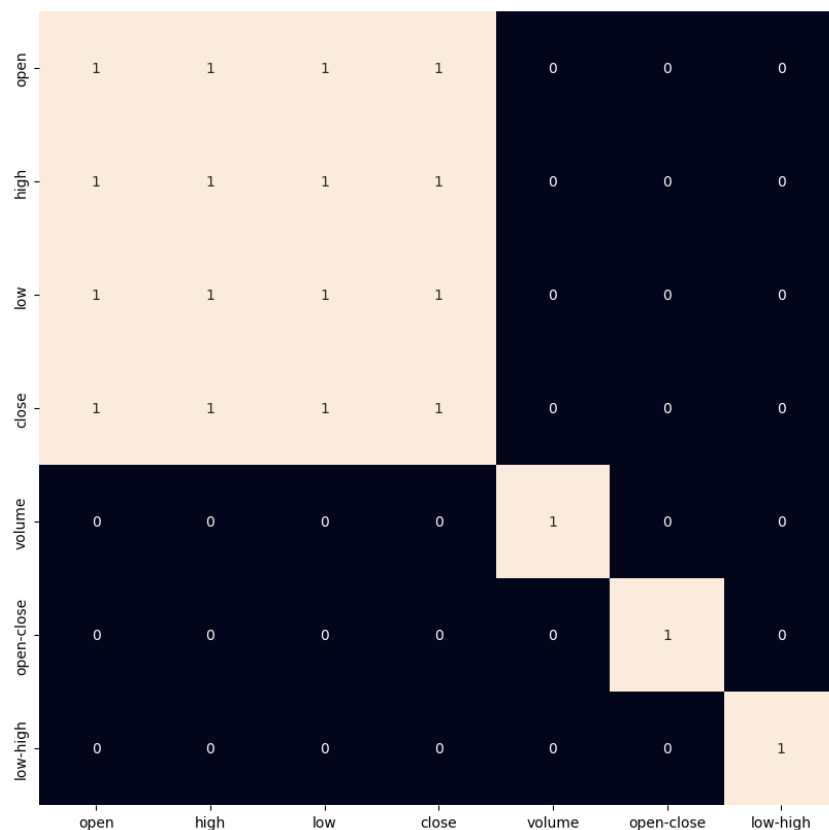


Рисунок 3.5 – кореляція між даними

Попередня обробка даних (Data Preprocessing) полягає в очищенні та підготовці даних до моделювання. Це включає видалення пропущених значень, нормалізацію, перетворення форматів даних та створення додаткових ознак (features). Цей етап важливий для зменшення шуму в даних і підвищення точності прогнозування, забезпечуючи перетворення вхідних даних у формат, зручний для аналізу.

Для цього етапу спершу потрібно перевірити чи немає у нас відсутніх значень в датасеті (див рис. 3.6).

```
df_sorted.isnull().sum()
```

✓ 0.5s

open	0
high	0
low	0
close	0
volume	0
date	0
open-close	0
low-high	0
target	0
dtype:	int64

Рисунок 3.6 – перевірка на відсутні значення в датасеті.

Та також перевіримо на відповідність тип наших даних:

```
df_sorted.info()
```

✓ 0.0s

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 3126000 entries, 0 to 3125999  
Data columns (total 9 columns):  
#   Column      Dtype  
---  ---  
0   open        float64  
1   high        float64  
2   low         float64  
3   close       float64  
4   volume      float64  
5   date        object  
6   open-close  float64  
7   low-high    float64  
8   target      int32  
dtypes: float64(7), int32(1), object(1)  
memory usage: 202.7+ MB
```

Рисунок 3.7 – перевірки типів даних в датасеті

Відповідно до отриманих результатів можемо сказати що наш датасет готовий для подальшого тренування, та не потрібно виконувати зміну типів, видалення або заповнення відсутніх значень.

Наша модель буде вчитись передбачати ціну Біткоїна та надавати нам інформацію про те чи варто нам купувати цю криптовалюту.

Для цього нам потрібно додати відповідну колонку даних (targets) в якому буде зазначено чи варто купувати Біткоїн чи ні. Для цього ми додамо колонку буде використана для позначення того, чи буде ціна в колонці Close зростати (1) чи падати (0) у наступному рядку. Відповідно якщо далі ціна зростає то краще купити Біткоїн, якщо спадає то краще почекати кінця спаду.

Для забезпечення ефективного навчання модель потребує приблизно однакову кількість різних маркерів (targets) у датасеті. Це гарантує, що модель отримує рівномірну кількість даних, що дозволяє їй навчатися на збалансованому наборі даних. Такий підхід запобігає перевазі одного класу над іншим, що підвищує точність та надійність класифікації. Для цього можна використати кругову діаграму для відображення розподілу маркерів (pie-chart) (див. рис. 3.8).

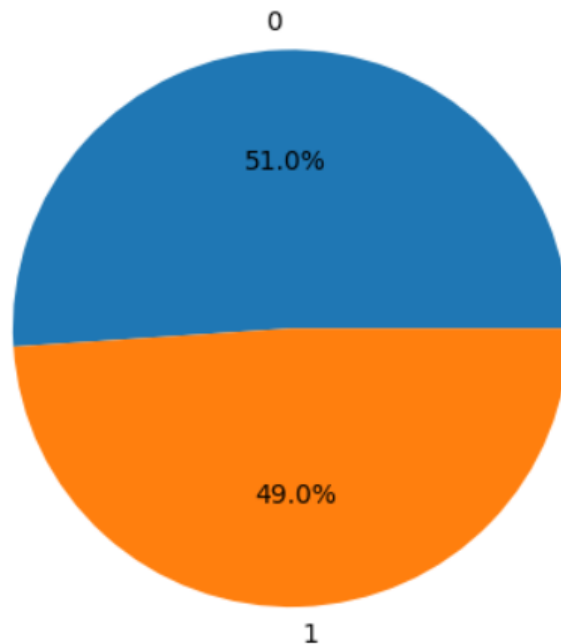


Рисунок 3.8 - Кругова діаграма з розподілом маркерів

По результатах зображених на круговій діаграмі (див рис. 3.6) можна визначити що у нас є приблизно рівномірний розподіл маркерів, що забезпечить нашій моделі хорошу якість тренування.

3.2 Тренування моделі

Перед початком тренування необхідно поділити наш датасет на тренувальний та тестовий. Поділ даних на тренувальну та тестову вибірки (Data Splitting) є необхідним для забезпечення незалежної оцінки моделі. Дані розділяються на дві частини: одна використовується для тренування моделі, інша – для її оцінки (див. рис. 3.9). Це важливо для оцінки продуктивності моделі та уникнення перенавчання, оскільки забезпечує незалежну перевірку точності моделі, на даних які модель іще не бачила та не мала можливості вивчити. В цьому випадку наш розподіл буде 20% на тестовий датасет, а також ми задамо значення змінної `random_state=2022` це забезпечить фіксоване розділення для відтворюваних результатів.

```
x_train, x_valid, y_train, y_valid = train_test_split(
    features, target, test_size=0.2, random_state=2022)
print(x_train.shape, x_valid.shape)
```

✓ 0.6s

(2500800, 3) (625200, 3)

Рисунок 3.9 – поділ даних на тренувальну та тестову вибірку

Наступний етап це побудова та тренування моделі (Model Building and Training). Він включає створення та тренування моделі машинного навчання з використанням алгоритму XGBoost. Чому ж саме цей алгоритм був обраний для цього завдання?

XGBoost (eXtreme Gradient Boosting) - це потужний алгоритм машинного навчання, який став дуже популярним у багатьох областях, включаючи

прогнозування цін на біткойн. XGBoost відомий своєю високою точністю прогнозування. Це досягається завдяки техніці градієнтного бустінгу, яка дозволяє моделі уточнювати прогнози, фокусуючись на областях, де вона робить помилки. Одною із переваг XGBoost є те що він відмінно працює з великими обсягами даних, і його ефективність зазвичай не падає навіть при великій кількості ознак і спостережень. [22, 23]

Це робить його ідеальним вибором для аналізу складних фінансових даних, таких як ціни на біткойн. Також цей алгоритм має різні методи регуляризації, такі як L1 та L2 регуляризація, що допомагають уникнути перенавчання моделі та підвищити її загальну здатність до узагальнення. Крім того він оптимізований для швидкої роботи, завдяки своїм ефективним алгоритмам обчислень та можливості паралельної обробки даних. Це дозволяє ефективно використовувати його навіть для великих датасетів.

На цьому етапі тренувальна вибірка використовується для налаштування параметрів моделі. Цей етап включає налаштування параметрів і оптимізацію моделі для досягнення найкращих результатів.

Перед початком тренування зазначимо кілька параметрів для XGBoost класифікатора, які ми можемо побачити на рисунку 3.10. Ці параметри були вибрані після кількох тренувань моделі та вибрані як найкращі.

```
param = {  
    'eta': 0.2,  
    'max_depth': 7,  
    'num_class': 3,  
    'n_estimators': 20}  
✓ 0.0s
```

Рисунок 3.10 – параметри для XGBoost

Розглянемо що кожен з цих параметрів означає:

- `eta (learning rate)`: Параметр швидкості навчання (`eta` або `learning rate`) контролює вагу нових дерев, доданих до моделі. Чим менше значення, тим більш обережним буде процес навчання, і тим менше ризику перенавчання моделі. Значення 0.5 вказує на досить швидке навчання, де кожне нове дерево має значний вплив на загальну модель.
- `max_depth`: Цей параметр визначає максимальну глибину дерев, що використовуються в моделі. Глибші дерева можуть уловлювати більш складні закономірності в даних, але також можуть призвести до перенавчання. Значення 10 означає, що дерево може мати до 10 рівнів, що дозволяє моделі уловлювати досить складні взаємозв'язки.
- `num_class`: Параметр `num_class` використовується для багатокласової класифікації і вказує кількість класів, які модель повинна передбачити. Значення 3 означає, що задача класифікації має три можливі класи результатів.
- `n_estimators`: Цей параметр визначає кількість дерев, які будуть побудовані під час тренування моделі. Кожне дерево намагається виправити помилки попередніх дерев. Значення 20 означає, що модель складатиметься з 20 дерев, що може бути достатньо для захоплення складних патернів, але може вимагати налаштування для уникнення перенавчання.

Також додатково був заданий параметр `objective='multi:softprob'` який дозволяє моделі вирішувати задачі багатокласової класифікації та повертати ймовірності належності до кожного класу. Це особливо корисно, коли потрібно мати не лише прогнозований клас, але й розподіл ймовірностей для прийняття більш інформованих рішень.

3.3 Тестування розробленої програми

Оцінка моделі (Model Evaluation) полягає в оцінці ефективності моделі на тестових даних. Ось кілька способів, якими ми використаємо для тестування та валідації отриманої моделі:

Для оцінки результатів цих моделей ми використовуватимемо криву ROC-AUC. Це обумовлено тим, що нам необхідно мати можливість передбачати м'які ймовірності, які представляють собою безперервні значення від 0 до 1, замість простого бінарного прогнозу, що дорівнює 0 або 1. Крива ROC-AUC ідеально підходить для оцінки точності таких прогнозів на основі м'яких ймовірностей, оскільки вона вимірює відповідність між дійсними та передбаченими ймовірностями. Результати моделі ми можемо побачити на рисунку 3.11.

```
XGBClassifier(base_score=None, booster=None, callbacks=None,
               colsample_bylevel=None, colsample_bynode=None,
               colsample_bytree=None, device=None, early_stopping_rounds=None,
               enable_categorical=False, eta=0.2, eval_metric=None,
               feature_types=None, gamma=None, grow_policy=None,
               importance_type=None, interaction_constraints=None,
               learning_rate=None, max_bin=None, max_cat_threshold=None,
               max_cat_to_onehot=None, max_delta_step=None, max_depth=7,
               max_leaves=None, min_child_weight=None, missing=nan,
               monotone_constraints=None, multi_strategy=None, n_estimators=20,
               n_jobs=None, num_class=3, ...) :
Training Accuracy : 0.8603715429391633
Validation Accuracy : 0.8595370462857757
```

Рисунок 3.11 – Результати точності моделей після завершення тренування

На рисунку 3.11 показано що велика кількість параметрів не задіяні. Якщо підібрати їх найкращим чином то точність результатів моделі можна буде покращити і ще на кілька значень. Проте наш результат є достатнім для того щоб дозволити використання цієї моделі для майбутнього використання.

Також побудуємо матрицю невідповідностей (див. рис. 3.12). Матриця невідповідностей - це матриця, яка підсумовує продуктивність моделі машинного

навчання на наборі тестових даних. Це засіб відображення кількості точних і неточних випадків на основі прогнозів моделі. Вона часто використовується для вимірювання продуктивності моделей класифікації, які мають на меті передбачити категоріальну мітку для кожного вхідного екземпляра.

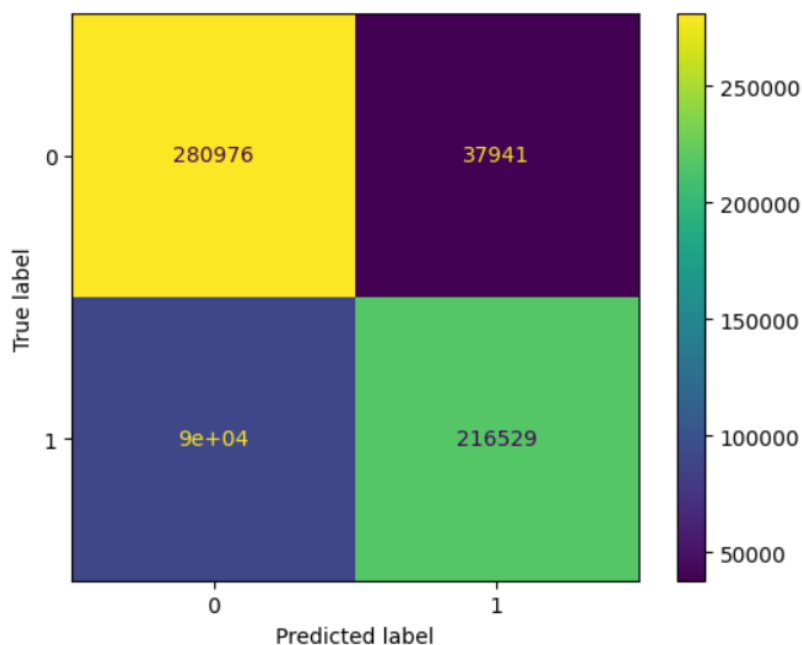


Рисунок 3.12 – Матриця не відповідності

Прогнозування (Prediction) є основною функцією програмного модуля, яка дозволяє використовувати натреновану модель для прогнозування майбутніх цін на основі нових даних. Нові дані вводяться в модель, яка генерує прогнози цін. Це забезпечує можливість прогнозувати значення цін на біткоїн та надавати користувачам інформацію по тому чи варто закупити криптовалюту чи почекати до переломного моменту коли ціна досягне кінцевого різня низької ціни після якого має бути зростання.

Тож давайте протестуємо нашу модель на кількох прикладах вхідних даних.

Та перевіримо прогноз. Для цього використаємо дані з сайтів які були представлені в розділі 1.2

Для цієї перевірки було вибрано зі сайту CoinCheckUp дані за 9 Травня (див. рис. 3.13). Для цього нам потрібно розрахувати різницю між ціною відкриття (open) і закриття (close) та найменшою (low) і найбільшою (high) за цей період ціною

ціною, та вказати чи є даний запис кінцем кварталу. Отримані значення показані на рисунку 3.14

Date	Open	High	Low	Close	Volume
Sun, May 12th 2024	\$60,823.19	\$61,838.51	\$60,600.90	\$61,459.22	\$502.85 M
Sat, May 11th 2024	\$60,790.82	\$61,489.54	\$60,474.73	\$60,823.19	\$505.61 M
Fri, May 10th 2024	\$63,077.18	\$63,469.13	\$60,181.24	\$60,790.82	\$1.57 B
Thu, May 9th 2024	\$61,179.81	\$63,429.98	\$60,623.94	\$63,077.18	\$1.14 B

Рисунок 3.13 – Дані по ціні Біткоїна за кілька днів

Ось результат прогнозу про подальшу ціну Біткоїна тренованої моделі.

```
test_preds = models[0].predict([[-1.89727, -2.13708, 0]])
✓ 0.0s

if test_preds == 0:
    print("Bitcoin price will fall")
else:
    print("The price of bitcoin will increase")
✓ 0.0s

Bitcoin price will fall
```

Рисунок 3.14 – Прогноз тренованої моделі

Як результат модель прогнозує спад ціни біткоїна. І як ми можемо побачити на рисунку 3.13, на наступний день ціна біткоїна до кінця дня впала на 3 000 доларів.

ВИСНОВКИ

Прогнозування курсу біткоїна має велике значення в сучасному фінансовому світі, оскільки дозволяє інвесторам ухвалювати обґрунтовані рішення та мінімізувати ризики. У світі з великою кількістю непередбачуваних факторів, що впливають на вартість криптовалют, точне прогнозування курсу стає надзвичайно важливим завданням для стабільного інвестування.

Під час цієї роботи розроблено програмний модуль для прогнозування ціни біткоїна. Перші етапи роботи включали загальний аналіз проблематики прогнозування цін криптовалют та методів їхнього аналізу. Проведено комплексний аналіз існуючих моделей машинного навчання та інструментів, які використовуються для прогнозування.

Практичні дослідження включали аналіз існуючих методів та моделей прогнозування, та готових реалізацій, зокрема з використанням XGBoost. Цей алгоритм дозволяє створити потужні прогностичні моделі завдяки своїй здатності обробляти великі набори даних і враховувати складні залежності між змінними. Ця технологія надає можливість створити модель, яка враховує складність та динамічність ринку криптовалют, що забезпечує точні прогнози та підтримку інвесторів у прийнятті рішень.

Успішна реалізація проекту з прогнозування курсу біткоїна за допомогою XGBoost передбачала комплексний підхід, що включає збір та аналіз великого обсягу історичних даних, тренування моделі XGBoost на цих даних, імплементацію алгоритмів прогнозування та постійне вдосконалення моделі на основі нових даних і методів машинного навчання.

Важливим етапом розробки є тестування та валідація ефективності моделі. Після впровадження програмного забезпечення було ретельно перевірено його на реальних даних, щоб переконатися в надійності та точності прогнозів.

Основними користувачами програмного модуля для прогнозування курсу біткоїна з використанням XGBoost можуть бути інвестори, трейдери, фінансові

аналітики та дослідники. Такий модуль допоможе їм приймати обґрунтовані рішення, знижуючи ризики та покращуючи точність прогнозів.

Програмний модуль для прогнозування курсу біткоїна є актуальним та перспективним напрямком у сфері фінансових технологій та машинного навчання. Його впровадження сприятиме підвищенню ефективності торгових стратегій, зменшенню фінансових втрат та покращенню результатів інвестування в криптовалюту.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Швидке зростання ринку криптовалют: детальний аналіз. URL: <https://www.binance.com/uk-UA/square/post/2024-05-30-rapid-growth-in-cryptocurrency-market-a-detailed-analysis-8775644743409>
2. Що таке блокчейн і як він працює? URL: <https://academy.binance.com/uk/articles/what-is-blockchain-and-how-does-it-work>
3. What Are Crypto Tokens, and How Do They Work? URL: <https://www.investopedia.com/terms/c/crypto-token.asp>
4. What is Ethereum in simple words. How does Ether differ from Bitcoin. URL: <https://minfin.com.ua/ua/invest/articles/ethereum-prostymi-slovami/>
5. Блокчейн в економіці та бізнесі. URL: https://economyandsociety.in.ua/journals/15_ukr/142.pdf
6. Benefits of blockchain <https://www.ibm.com/topics/benefits-of-blockchain>
7. Mikhaylov, A. (2020). Cryptocurrency market analysis from the open innovation perspective. Journal of Open Innovation: Technology, Market, and Complexity, 6(4), 197.
8. The Crypto Question: Bitcoin, Digital Dollars, and the Future of Money URL: <https://www.cfr.org/backgrounder/crypto-question-bitcoin-digital-dollars-and-future-money>
9. Що таке майнінг криптовалют? URL:
10. What Is Bitcoin? How To Mine, Buy, and Use It URL: <https://www.investopedia.com/terms/b/bitcoin.asp>
<https://blog.whitebit.com/uk/what-is-cryptocurrency-mining/>
11. WalletInvestor URL: <https://walletinvestor.com/forecast>
12. CoinCheckup URL: <https://coincheckup.com/>
13. CoinMarketCap URL: <https://coinmarketcap.com/uk/>
14. Cryptowatch URL: <https://pro.kraken.com/app/trade/btc-usd>
15. CryptoCompare URL: <https://www.cryptocompare.com/>

16. What is exploratory data analysis (EDA)? URL:
<https://www.ibm.com/topics/exploratory-data-analysis>
17. Cryptocurrency price prediction based on Xgboost, LightGBM and BNN
ULR:
https://www.researchgate.net/publication/379180753_Cryptocurrency_price_prediction_based_on_Xgboost_LightGBM_and_BNN
18. BITCOIN VALUE PREDICTION URL:
<https://ijcrt.org/papers/IJCRT2311100.pdf>
19. XGBoost Mathematics Explained URL:
<https://towardsdatascience.com/xgboost-mathematics-explained-58262530904a>
20. Bitcoin Price Dataset (2017-2023) URL:
<https://www.kaggle.com/datasets/jkraak/bitcoin-price-dataset/data>
21. Біткоїн перетнув позначку у 51 тисячу доларів URL:
<https://www.unian.ua/economics/finance/bitkojin-rizko-zletiv-rekord-z-grudnya-2021-roku-12543462.html>
22. Bitcoin price prediction with Python, when the past does not repeat itself/
URL: <https://cienciadedatos.net/documentos/py41-forecasting-cryptocurrency-bitcoin-machine-learning-python>
23. Data Analysis and Classification using XGBoost URL:
<https://www.kaggle.com/code/lucidlenn/data-analysis-and-classification-using-xgboost>
24. Островерхов В. М., Біловус Л. І., Возьний К. З., Луцишин О. О., Монастирський Г. Л., Надвиничний С. А., Питель С. В., Шандрук С. К. Загальні методичні рекомендації з підготовки, оформлення, захисту та оцінювання кваліфікаційних робіт здобувачів вищої освіти першого (бакалаврського) і другого (магістерського) рівнів / Укладачі: Тернопіль: ЗУНУ, 2024. 83 с.
25. Комар М.П., Саченко А.О., Васильків Н.М., Гладій Г.М., Коваль В.С., Ліп'яніна-Гончаренко Х.В. Методичні рекомендації до виконання кваліфікаційної роботи з освітньо-професійної програми «Комп'ютерні науки» спеціальності 122

«Комп'ютерні науки» за першим (бакалаврським) рівнем вищої освіти. Тернопіль: ЗУНУ, 2024. 52 с.

26. Гузюнь І. В. Програмний модуль прогнозування ціни біткоїна засобами машинного навчання. Наука, освіта та технології: актуальні проблеми теорії та практики: тези доп. Міжнародна науково-практична конференція (Тампере, Фінляндія, 12 червня 2024 р.). С. 24-25. [Додаток А]

Додаток А
Копія публікації

Інформаційні технології

УДК 004.93

Гузюнь І. В.

студентка 4 курсу факультету
комп'ютерних інформаційних технологій,
Західноукраїнський національний університет

**ПРОГРАМНИЙ МОДУЛЬ ПРОГНОЗУВАННЯ ЦІНИ БІТКОЇНА
ЗАСОБАМИ МАШИННОГО НАВЧАННЯ.**

У сучасному світі криптовалюти привертають увагу не лише фінансових експертів та інвесторів, але й академічної спільноти. Їхнє швидке зростання і висока волатильність суттєво впливають на глобальні фінансові ринки та економіку загалом. За останні кілька років виникла потреба у створенні надійних методів прогнозування курсу криптовалют, щоб інвестори та дослідники могли приймати обґрунтовані рішення та аналізувати ринки криптовалют. Прогнозування курсу біткоїна та інших криптовалют є важливим завданням, але водночас дуже складним через їхню хаотичну та нестабільну природу. [1]

Актуальність проблеми прогнозування курсу криптовалют є очевидною. Світ фінансів, інвестицій та технологій зазнає значних змін завдяки криптовалютам, які стали не лише альтернативним засобом інвестицій, але й основною темою фінансових новин. Навіть незначні події можуть спричинити великі коливання цін на ринках криптовалют. Наприклад, злам біржі Bitfinex в Гонконзі спричинив раптове падіння вартості біткоїна на 20%. Тому розуміння впливу новин, подій та інших факторів на курс криптовалют стає критично важливим для інвесторів та дослідників.

Біткоїн, перша та найвідоміша криптовалюта, яка здобула значну популярність протягом останнього десятиліття. Її ціна демонструвала динамічне зростання, супроводжуючись різкими коливаннями, що робить її предметом активних дискусій та досліджень.

Ціна біткоїна визначається комплексом факторів, поділених на дві категорії:

1. Попит і пропозиція. Зростання популярності біткоїна, його сприйняття як інвестиційного активу, використання як засобу платежу та інші фактори, що стимулюють бажання володіти ним, сприяють зростанню попиту. Крім того кількість біткоїнів обмежена 21 мільйоном, що робить його дефіцитним активом. Зростання складності майнінгу, уповільнення темпів випуску нових біткоїнів також впливає на пропозицію. [2]

2. Спекулятивні фактори. Позитивні новини, пов'язані з біткоїном, такі як легалізація в нових країнах, інновації в екосистемі, можуть призвести до зростання ціни. Навпаки, негативні новини, збої в роботі бірж, хакерські атаки можуть призвести до падіння. Не менш важливою є поведінка інвесторів, схильність до ризику, очікування щодо майбутнього біткоїна значно впливають на ціноутворення. [2]

Зростання біткоїна несе в собі як потенційні вигоди, так і ризики для економіки та фінансової системи.

До вигоди можна додати те що біткоїн може сприяти децентралізації фінансової системи, роблячи платежі більш доступними, швидкими та дешевими, також криптовалюта може стати новим класом активів, пропонуючи інвесторам альтернативні інвестиційні стратегії і що не менш важливо стимулювати інновації в фінансових технологіях та інших сферах.

Щодо ризиків, то вони можуть бути наступними:

- Різкі коливання ціни біткоїна можуть створювати ризики для інвесторів та фінансових установ.
- Деанонізований характер біткоїна може сприяти його використанню в незаконних цілях, таких як відмивання грошей або фінансування тероризму.

- Зростання біткоїна може призвести до втрати контролю та впливу традиційними фінансовими інститутами.

Прогнозування ціни криптовалюти - це складне завдання, яке потребує розуміння динаміки ринку, технічного аналізу та економічних факторів. Хоча жоден метод не може дати 100% гарантії точності, методи прогнозування, такі як технічний аналіз, машинне навчання та фундаментальний аналіз, можуть допомогти трейдерам та інвесторам приймати більш обґрунтовані рішення. [3]

Методи прогнозування ціни криптовалюти, зокрема біткоїна, є надзвичайно важливими для фінансових ринків через високу волатильність і непередбачуваність цих активів. Програмний модуль для прогнозування ціни біткоїна є важливим інструментом для фінансових ринків, ЗМІ, дослідників та інвесторів. Він дозволяє приймати обґрунтовані рішення, знижувати ризики і підвищувати ефективність інвестицій. У цілому, розвиток і застосування таких програмних модулів сприяє інтеграції криптовалют у традиційну фінансову систему, підвищуючи їхню привабливість та надійність як інвестиційного інструменту.

Список літератури

1. What Is Cryptocurrency? Crypto.com. URL: <https://crypto.com/university/what-is-cryptocurrency>
2. The future of money. Think Magazine. URL: <https://thinkmagazine.mt/the-future-of-money/>
3. Демківський Є.О., Демківська Т.І.. Прогнозування нестационарних фінансово-економічних процесів. IV-а Міжнародна науково-практична конференція «Мехатронні системи: інновації та інжиніринг»: Київ, КНУТД, 22 жовтня 2020 року, с. 39-40

Додаток Б

Код програмного модуля

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC
from xgboost import XGBClassifier
from sklearn import metrics
from sklearn.metrics import ConfusionMatrixDisplay

df_sorted = pd.read_csv('bitcoin.csv')

splitted = df_sorted['date'].str.split('-', expand=True)

df_sorted['year'] = splitted[0].astype('int')
df_sorted['month'] = splitted[1].astype('int')
df_sorted['day'] = splitted[2].astype('int')

df_sorted['open'] = pd.to_numeric(df_sorted['open'])
df_sorted['high'] = pd.to_numeric(df_sorted['high'])
df_sorted['low'] = pd.to_numeric(df_sorted['low'])
df_sorted['close'] = pd.to_numeric(df_sorted['close'])

df_sorted['open-close'] = df_sorted['open'] - df_sorted['close']
df_sorted['low-high'] = df_sorted['low'] - df_sorted['high']
df_sorted['target'] = np.where(df_sorted['close'].shift(-1) > df_sorted['close'], 1, 0)

numerical_df = df_sorted.select_dtypes(include=['float64', 'int64'])

corr_matrix = numerical_df.corr()
plt.figure(figsize=(10, 10))

sb.heatmap(corr_matrix > 0.9, annot=True, cbar=False)
plt.show()

features = df_sorted[['open-close', 'low-high', 'is_quarter_end']]
target = df_sorted['target']

scaler = StandardScaler()
features = scaler.fit_transform(features)

X_train, X_valid, Y_train, Y_valid = train_test_split(
    features, target, test_size=0.2, random_state=2022)
print(X_train.shape, X_valid.shape)
```

```

param = {
    'eta': 0.2,
    'max_depth': 7,
    'num_class': 2,
    'n_estimators': 20}

models.fit(X_train, Y_train)

print(f'{models} : ')
print('Training Accuracy : ', metrics.roc_auc_score(Y_train, models.predict_proba(X_train)[:,:1]))
print('Validation Accuracy : ', metrics.roc_auc_score(Y_valid, models.predict_proba(X_valid)[:,:1]))
print()
|

```