

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Чорноморський національний університет імені Петра Могили
Факультет комп'ютерних наук
Кафедра інтелектуальних інформаційних систем

ДОПУЩЕНО ДО ЗАХИСТУ

Завідувач кафедри інтелектуальних
інформаційних систем

_____Юрій КОНДРАТЕНКО

« ____ » _____ 2024 р.

КВАЛІФІКАЦІЙНА РОБОТА
НА ЗДОБУТТЯ ОСВІТНЬОГО СТУПЕНЯ МАГІСТРА
ІНТЕЛЕКТУАЛЬНА СИСТЕМА ПРОГНОЗУВАННЯ
ЦІН КРИПТОВАЛЮТ НА ОСНОВІ РЕКУРЕНТНИХ
НЕЙРОННИХ МЕРЕЖ

Спеціальність 122 Комп'ютерні науки

Освітня програма «Інтелектуальні інформаційні системи»

Здобувач

_____ Владислав КИРИЛЕНКО

« ____ » _____ 2024 р.

Керівник д-р техн. наук, професор

_____ Олександр ГОЖИЙ

« ____ » _____ 2024 р.

Миколаїв – 2024

Чорноморський національний університет імені Петра Могили

(повне найменування закладу вищої освіти)

Факультет	Комп'ютерних наук
Кафедра	Інтелектуальних інформаційних систем
Рівень вищої освіти	Другий (магістерський)
Освітній ступень	Магістр
Спеціальність	122 Комп'ютерні науки
Освітня програма	Інтелектуальні інформаційні системи

ЗАТВЕРДЖУЮ

Завідувач кафедри інтелектуальних
інформаційних систем

_____ Юрій КОНДРАТЕНКО

« ____ » _____ 2024 р.

ЗАВДАННЯ на кваліфікаційну роботу здобувача

Кириленка Владислава Сергійовича

(прізвище, ім'я, по батькові здобувача)

1. Тема кваліфікаційної роботи: «Інтелектуальна система прогнозування цін криптовалют на основі рекурентних нейронних мереж».

Керівник роботи: Гожий Олександр Петрович, професор кафедри ІС, д-р техн. наук, професор.

Затверджена наказом ЧНУ ім. Петра Могили від «03» червня 2024 р. № 140/1.

2. Строк представлення кваліфікаційної роботи « ____ » _____ 2024 р.

3. Очікуваний результат роботи та початкові дані, якщо такі потрібні: готова нейронна мережа навчена за архітектурою LSTM, що здатна прогнозувати дельту зміни ціни на певний період часу у майбутньому; відсортовані за часом історичні дані про певні макроекономічні показники (ціни на золото, індекс долара, ціни на акції S&P 500, NASDAQ 100 тощо), а також про ціни (відкриття, закриття, найвища та найнижча за період) на певний криптоактив.

4. Перелік питань, що підлягають розробці: аналіз сучасного стану задачі прогнозування цін криптовалют із використанням методів машинного навчання; огляд існуючих архітектур рекурентних нейронних мереж (RNN), їх модифікацій та особливостей застосування для часових рядів; розробка та впровадження підходу до побудови інтелектуальної системи прогнозування цін криптовалют із використанням RNN; порівняльний аналіз результатів роботи розробленої системи з іншими методами прогнозування часових рядів.

5. Перелік графічних матеріалів: презентація.

Керівник роботи

(Особистий підпис)

Олександр ГОЖИЙ

(Власне ім'я ПРІЗВИЩЕ)

Здобувач

(Особистий підпис)

Владислав КИРИЛЕНКО

(Власне ім'я ПРІЗВИЩЕ)

Дата видачі завдання «07» червня 2024 р.

КАЛЕНДАРНИЙ ПЛАН

кваліфікаційної роботи

Тема: Інтелектуальна система прогнозування цін криптовалют на основі рекурентних нейронних мереж

№	Найменування роботи	Початок	Закінчення	Примітки
1	Отримання завдання на виконання КР	03.06.2024	07.06.2024	Виконано
2	Аналіз предметної області та постановка задачі	10.06.2024	20.06.2024	Виконано
3	Огляд літературних джерел за темою кваліфікаційної роботи, зокрема аналіз публікацій та аналогічних систем, щодо прогнозування криптоактивів за допомогою нейронних мереж різної архітектури	21.06.2024	01.07.2024	Виконано
4	Огляд існуючих архітектур штучних нейронних мереж для вирішення поставленої задачі	01.09.2024	25.10.2024	Виконано
5	Реалізація обраних технологій з аналізом отриманих результатів	26.10.2024	21.11.2024	Виконано
6	Перший попередній захист КР на засіданні комісії кафедри	22.11.2024	22.11.2024	Виконано
7	Корегування роботи за результатами попереднього захисту	23.11.2024	05.12.2024	Виконано
8	Другий попередній захист КР на засіданні комісії кафедри	06.12.2024	06.12.2024	Виконано
9	Доробка та остаточне оформлення КР	07.12.2024	10.12.2024	Виконано
10	Подання КР, її електронної копії та інших документів (відгуку, рецензії) до захисту	16.12.2024	17.12.2024	Виконано

Керівник роботи

(Особистий підпис)

Олександр ГОЖИЙ
(Власне ім'я ПРІЗВИЩЕ)

Здобувач

(Особистий підпис)

Владислав КИРИЛЕНКО
(Власне ім'я ПРІЗВИЩЕ)

Дата складання календарного плану
«19» червня 2024 р.

АНОТАЦІЯ

до кваліфікаційної роботи
здобувача групи 601м ЧНУ ім. Петра Могили

Кириленка Владислава Сергійовича

на тему: **“ІНТЕЛЕКТУАЛЬНА СИСТЕМА ПРОГНОЗУВАННЯ ЦІН
КРИПТОВАЛЮТ НА ОСНОВІ РЕКУРЕНТНИХ НЕЙРОННИХ МЕРЕЖ”**

Актуальність даного дослідження полягає у необхідності знайти правильний підхід до прогнозування цін криптоактивів за допомогою використання рекурентних нейронних мереж. Недостатня точність сучасних підходів підвищує актуальність до вирішення даного питання.

Об’єктом дослідження є процес прогнозування цін криптовалют з використанням рекурентних нейронних мереж.

Предметом дослідження є математичні моделі, методи машинного навчання та інформаційні технології, що використовуються для прогнозування.

Метою дослідження є розробка інтелектуальної системи прогнозування цін на основі рекурентних нейронних мереж, що забезпечить високу точність передбачення.

В результаті виконання роботи було досліджено кілька архітектур на базі рекурентних нейронних мереж (LSTM, GRU) та експериментальним шляхом порівняно результати ефективності всіх архітектур. Отримано навчену модель на основі архітектури LSTM, що враховує багатofакторний вплив різних показників ринку та має високу точність прогнозування.

Дана робота складається з трьох розділів. Кожен розділ відповідно присвячений: аналізу предметної області; інформаційним технологіям для прогнозування цін; проектуванню системи, її побудові, збору даних та аналізу результатів. Загальний обсяг роботи – 92 сторінок. Кваліфікаційна робота містить 4 додатки, 52 рисунки, 1 таблицю і 45 джерел посилання.

Ключові слова: Bitcoin, криптовалюти, прогнозування, рекурентні нейронні мережі, LSTM, машинне навчання, аналіз даних, макроекономічні індикатори.

ABSTRACT

to the qualification work by the student of the group 601m of Petro Mohyla Black Sea National University

Kyrylenko Vladyslav

“INTELLIGENT SYSTEM FOR FORECASTING CRYPTOCURRENCY PRICES BASED ON RECURRENT NEURAL NETWORKS”

The relevance of this research lies in the need to find the right approach to forecasting the prices of cryptoassets using recurrent neural networks. The lack of accuracy of modern approaches increases the urgency of solving this issue.

The object of the research is the process of forecasting the prices of cryptocurrencies using recurrent neural networks.

The subject of research is mathematical models, machine learning methods and information technologies used for forecasting.

The purpose of the research is to develop an intelligent price forecasting system based on recurrent neural networks, which will ensure high prediction accuracy.

Because of the work, several architectures based on recurrent neural networks (LSTM, GRU) were investigated and the results of the effectiveness of all architectures were experimentally compared. A trained model based on the LSTM architecture is obtained, which takes into account the multifactorial influence of various market indicators and has high forecasting accuracy.

This work consists of three sections. Each section is respectively devoted to: analysis of the subject area; information technologies for price forecasting; system design, its construction, data collection and analysis of results. The total volume of work is 91 pages. The qualification work contains 4 appendices, 52 pictures, 1 table and 45 reference sources.

Key words: Bitcoin, cryptocurrencies, forecasting, recurrent neural networks, LSTM, machine learning, data analysis, macroeconomic indicators.

ЗМІСТ

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ	3
ВСТУП.....	4
1 АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ ТА ПУБЛІКАЦІЙ.....	6
1.1 Історія та розвиток криптовалютних ринків. Особливості ринку криптовалют. Блокчейн	6
1.2 Огляд предметної сфери прогнозування цін криптовалют за допомогою нейронних мереж.....	14
1.3 Огляд наявних досліджень та публікацій. Проведення аналогій.....	16
Висновок до розділу 1	18
2 ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ДЛЯ ПРОГНОЗУВАННЯ ЦІН НА КРИПТОВАЛЮТИ	20
2.1 Огляд методів прогнозування.....	20
2.2 Інформаційні технології для збору та аналізу даних	23
2.3 Процес отримання даних.....	25
2.4 Аналіз отриманих даних.....	33
2.5 Методи оцінки ефективності моделей.....	41
Висновок до розділу 2	43
3 ПРОЄКТУВАННЯ СИСТЕМИ. ПОБУДОВА АРХІТЕКТУРИ НЕЙРОННОЇ МЕРЕЖІ. ЗБІР ТА ОБРОБКА ДАНИХ ДЛЯ НАВЧАННЯ.....	45
3.1 Вибір архітектури нейронної мережі та її побудова	45
3.2 Збір, обробка та підготовка даних до навчання нейронної моделі.....	49
3.3 Програмна середовище та програмна реалізація	65
3.4 Практичне застосування.....	68
Висновки до розділу 3	72
ВИСНОВКИ.....	74
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ.....	75
ДОДАТОК А Лістинг коду веб-скрапера для збору даних EFFF.....	80
ДОДАТОК Б Лістинг коду веб-скрапера новин про криптовалюту та оцінку настроїв новин	82
ДОДАТОК В Лістинг коду веб-скрапера для збору даних про макроекономічні показники	84
ДОДАТОК Г Лістинг коду скрипту для підготовки погодинних даних	86

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ

API – Application Programming Interface

ARIMA – AutoRegressive Integrated Moving Average

ATR – Average True Range

BTC – Bitcoin

EFFR – Effective Federal Funds Rate

GRU – Gated Recurrent Unit

IoT – Internet of Things

KNN – K-nearest neighbour method

LSTM – Long Short-Term Memory

MAE – Mean Absolute Error

PoS – Proof of Stake

PoW – Proof of Work

RMSE – Root Mean Square Error

RNN – Recurrent Neural Network

RSI – Relative Strength Index

SVM – Support Vector Machines

ВСТУП

Задача прогнозування зміни ціни на криптовалютні активи є доволі актуальною задачею на сьогоднішній день, враховуючи складність та волатильність сучасного фінансового ринку. Хоча й сфера фінансового аналізу містить великі досягнення, все ж достатньо завдань залишаються нерозв'язаними, враховуючи й невирішену задачу прогнозування криптоактивів, адже внаслідок непередбачуваних факторів (ринкові маніпуляції, регулятивні рішення, нові технологічні тренди тощо) не дають можливості остаточно знайти підхід до вирішення цього питання. Наявні результати прогнозування, а саме класичні методи машинного навчання, показують неідеальну ефективність. А успіх відомих та провідних компаній, як відомо, залежить від вміння застосовувати нестандартні та нові підходи, що можуть стати переломними у вирішенні будь-яких питань. Опанування технологій штучного інтелекту може дозволити досягти необхідних результатів, зокрема завдяки використанню рекурентних нейронних мереж.

Необхідність знаходження вирішення подібного питання постає в потребі знайти потужний інструмент, що дозволить прогнозувати ціни з високою точністю. Так як криптовалюти знаходяться наразі в тренді та мають вагомий вплив на глобальну економіку, створення інтелектуальної системи може стати величезним проривом у сфері інвестицій, що може значно знизити ризик втрати частини інвестиційного портфелю.

Дане дослідження базується на попередніх дослідженнях в галузі машинного навчання нейронних мереж, що були присвячені вирішенню питання прогнозування криптовалютних активів.

Для досягнення цієї мети необхідно вирішити наступні завдання:

- проаналізувати існуючі підходи до прогнозування криптовалют;
- розробити архітектуру системи на основі рекурентних нейронних мереж;
- навчити модель на основі обраної архітектури;

– дослідити ефективність використання запропонованої моделі для прогнозування ринкових змін.

Для вирішення поставленого завдання буде використано машинне навчання моделі, що базуватиметься на основі архітектури рекурентних нейронних мереж. Для побудови даної архітектури буде застосовано мову програмування Python та спеціалізовані для вирішення подібних питань бібліотеки TensorFlow і Keras.

1 АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ ТА ПУБЛІКАЦІЙ

1.1 Історія та розвиток криптовалютних ринків. Особливості ринку криптовалют. Блокчейн

Криптовалюти мають свої корені в концепції цифрових валют, що з'явилася на початку 1980-х років [1]. Одна з перших спроб створити цифрову валюту була реалізована за допомогою протоколу DigiCash [2], розробленого в 1990 році. DigiCash дозволяв користувачам здійснювати анонімні онлайн-транзакції, але його популярність швидко зникла через обмежене прийняття та відсутність інфраструктури для широкого використання.

Вплив на подальший розвиток криптовалют також мали концепції блокчейну і дистрибутивних обчислень. Блокчейн, як технологія, розпочав свою історію разом із появою Bitcoin [3]. У 2008 році особа або група осіб під псевдонімом Сатоші Накамото опублікувала документ, що описує Bitcoin, у якому було представлено новий підхід до дистрибутивної цифрової валюти. У січні 2009 року було видобуто перший блок Bitcoin, відомий як Genesis Block [4], що поклало початок новій епісі у світі фінансів.

Після створення Bitcoin ринок криптовалют почав розвиватися стрімкими темпами. У 2011 році з'явилися перші альтернативні криптовалюти, або альткоїни. Однією з перших була Litecoin [5], розроблена Чарлі Лі, яка впроваджувала нові алгоритми видобутку та скорочувала час генерування блоків. Цей період ознаменувався появою інших криптовалют, таких як Ripple, Peercoin, Namecoin, які пропонували різні рішення та функціональність. Згідно з даними Statista [6], до 2013 року кількість криптовалют перевищила 100, що свідчить про зростаючий інтерес до цього нового фінансового інструменту.

У 2013 році ринок криптовалют зазнав значного зростання, коли ціна Bitcoin перевищила 1000 доларів [7]. Цей етап привернув увагу медіа та інвесторів, що призвело до буму навколо криптовалют. Відповідно до статистики, у 2013 році ринкова капіталізація Bitcoin досягла 13 мільярдів доларів [7], а загальна

капіталізація всіх криптовалют перевищила 15 мільярдів доларів [7]. Разом із зростанням популярності Bitcoin почали з'являтися нові проекти, орієнтовані на децентралізовані фінанси (DeFi) [8], такі як Ethereum, що було запущено у 2015 році [9]. Ethereum, розроблений Віталієм Бутеріним, став основою для створення децентралізованих додатків (dApps) завдяки впровадженню смарт-контрактів, що розширило можливості використання технології блокчейн.

З ростом ринку криптовалют з'явилися й нові виклики. У 2014 році сталася одна з найбільших крадіжок в історії криптовалют — злам Mt. Gox [10], найбільшої на той час біржі Bitcoin, що призвело до втрати понад 850 тисяч BTC (близько 450 мільйонів доларів на той момент). Це подія привернула увагу регуляторів у багатьох країнах і змусила їх розглянути питання щодо регулювання криптовалютного ринку. Країни почали вводити закони та правила, спрямовані на захист інвесторів і запобігання відмиванню грошей. Такі регуляції різняться від країни до країни: деякі, як Японія, визнали Bitcoin як законний засіб платежу [11], в той час як інші, як Китай, заборонили торгівлю криптовалютами [12].

Сьогодні ринок криптовалют налічує тисячі активів, і його загальна капіталізація сягає трильйонів доларів. За даними CoinMarketCap [13], у 2021 році ринкова капіталізація всіх криптовалют перевищила 2 трильйони доларів. Криптовалюти стали частиною фінансового ландшафту, зокрема завдяки зростанню інституційних інвестицій. Великі компанії, такі як Tesla [14], MicroStrategy, і Square, почали інвестувати в Bitcoin, що суттєво підвищило інтерес до криптовалют серед традиційних інвесторів. Наприклад, у 2020 році MicroStrategy оголосила про покупку Bitcoin на суму 250 мільйонів доларів [15], ставши однією з перших публічних компаній, яка включила Bitcoin у свій баланс.

Крім того, зростання популярності DeFi та NFT [16] (незамінних токенів) представило нові можливості для користувачів. DeFi платформи, такі як Uniswap і Aave, дозволяють користувачам отримувати доходи, надаючи свої криптовалюти в кредит або торгуючи на децентралізованих біржах. За даними DeFi Pulse, станом на вересень 2021 року загальний обсяг заблокованих активів у DeFi перевищив 100

мільярдів доларів [17], що свідчить про зростаючий інтерес до децентралізованих фінансових рішень.

NFT, в свою чергу, дозволили цифровим художникам та креативним особам монетизувати свої роботи. У 2021 році аукціонний дім Christie's продав перший NFT-арт за рекордну суму 69 мільйонів доларів [18], що привернуло увагу до нової форми мистецтва і використання блокчейн-технології для підтвердження автентичності.

Історія розвитку криптовалют відображає еволюцію технологій, зміну фінансових парадигм та адаптацію ринків до нових викликів. Незважаючи на коливання цін, регуляторні обмеження та проблеми з безпекою, криптовалюти залишаються важливим інструментом у фінансовому світі, пропонуючи нові можливості для інвесторів, підприємців та споживачів. У подальшому розвитку ринку ключовими залишаються інновації в технологіях та механізмах регулювання, які вплинуть на майбутнє криптовалют.

Ринок криптовалют відрізняється від традиційних фінансових ринків своєю унікальною природою. Це децентралізована платформа, де активи обмінюються безпосередньо між користувачами за допомогою технології блокчейн. На відміну від акцій, облігацій або товарів, криптовалюти не мають фізичних активів за собою; їхня вартість визначається попитом і пропозицією на ринку. На відміну від традиційних валют, які підлягають контролю держави, криптовалюти не мають централізованого контролю, що робить їх привабливими для інвесторів, які прагнуть уникнути впливу традиційних фінансових інститутів.

Однією з найпомітніших характеристик ринку криптовалют є його висока волатильність. Ціни на криптовалюти можуть змінюватися на десятки відсотків за день. Наприклад, у 2021 році ціна Bitcoin сягнула рекордного рівня понад 60 000 доларів США, а потім різко впала до 30 000 доларів у червні того ж року [19]. Така волатильність може бути наслідком різних факторів, включаючи новини про регулювання, ринкові спекуляції та зміни в попиті на активи.

Волатильність криптовалют є результатом комбінації низького обсягу торгівлі, обмеженої ліквідності та великої кількості інвестиційних спекуляцій. Наприклад, в 2017 році ціна Bitcoin зросла з приблизно 1 000 до 20 000 доларів, після чого стала свідком різкого падіння, яке призвело до зниження вартості на 80% у 2018 році.

На ринку криптовалют представлено тисячі різних активів. Bitcoin, як перша криптовалюта, залишається найбільш впливовою, але існує безліч інших альткоїнів, таких як Ethereum, Ripple, Litecoin та багато інших. Кожен з цих активів має свої унікальні функції та характеристики. Наприклад, Ethereum дозволяє створювати смарт-контракти [20], що робить його особливим у порівнянні з Bitcoin, який в основному використовується як засіб збереження вартості. Тим не менш, протягом багатьох років і навіть до сьогодні, Bitcoin залишається головною валютою і навіть не дивлячись на швидкість появи нових криптоактивів, особливо популярних Altcoin, Bitcoin сильно домінує (див. рис. 1.1) і його частка використання наразі лише збільшується.



Рисунок 1.1 – Графік домінації Bitcoin над іншими криптоактивами протягом усього періоду існування криптовалют

Згідно з даними CoinMarketCap, станом на вересень 2021 року ринок налічував понад 11 000 криптовалют, з загальною капіталізацією понад 2 трильйони доларів США. Це свідчить про те, що інтерес до криптовалют зростає (див. рис. 1.2), і нові проекти постійно з'являються на ринку.

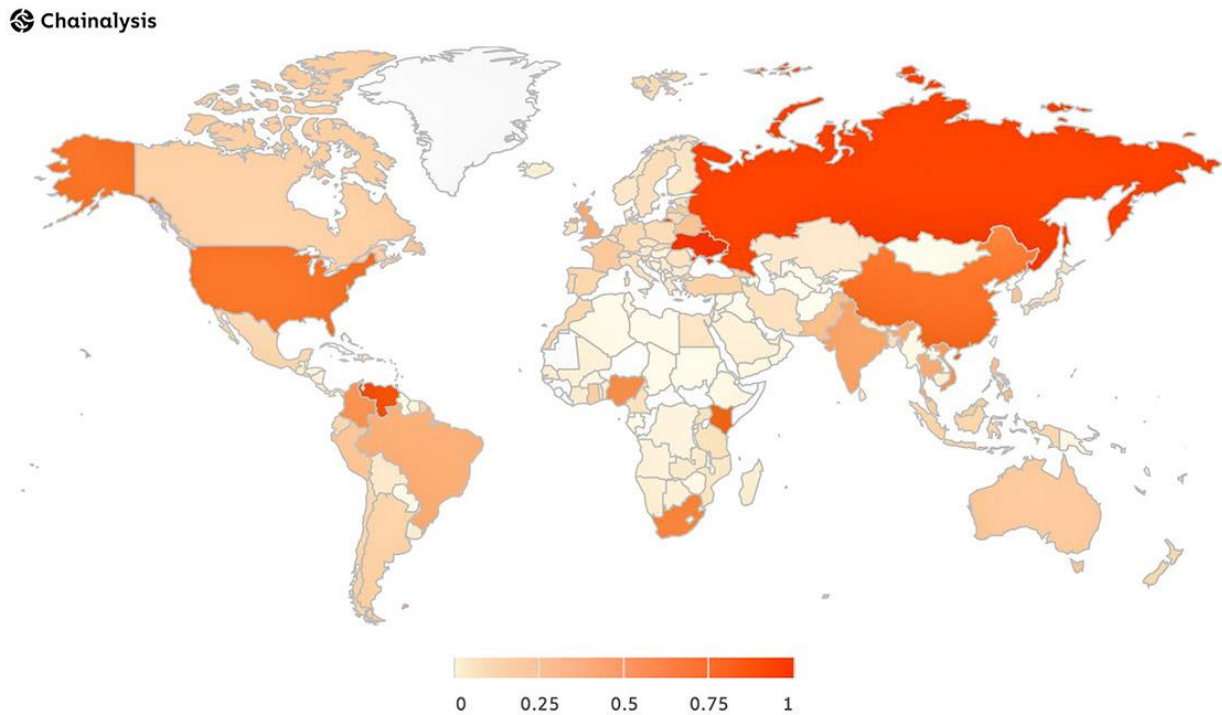


Рисунок 1.2 – Теплова мапа популярності використання криптовалют серед країн світу

В останні роки спостерігається бурхливий розвиток децентралізованих фінансів (DeFi), які надають нові можливості для інвесторів. DeFi платформи дозволяють користувачам брати кредити, надавати позики, торгувати активами та заробляти доходи без посередників. Наприклад, платформи, такі як Uniswap [21] та Aave [22], використовують смарт-контракти для автоматизації фінансових операцій, що робить процеси більш прозорими і ефективними.

Регуляція криптовалют стає актуальною темою, оскільки багато країн намагаються знайти баланс між захистом інвесторів та сприянням інноваціям. За даними, з 2020 року було зареєстровано понад 400 регуляторних актів [23], спрямованих на криптовалюту, що свідчить про зростаючу увагу до цього сектора.

Вкладення в криптовалюти пов'язані з певними ризиками. Крім волатильності, інвестори можуть стикатися з ризиками безпеки, такими як зломи бірж і шахрайство. За даними [10], у 2020 році було зафіксовано збитків на суму понад 1,8 мільярда доларів внаслідок крадіжок криптовалют.

Також існує ризик регуляторних змін, які можуть вплинути на вартість активів. Наприклад, оголошення про заборону криптовалют у Китаї в 2021 році [12] призвело до різкого падіння цін на активи.

Ринок криптовалют є складним і швидко змінюваним середовищем, яке відрізняється від традиційних фінансових ринків. Особливості, такі як висока волатильність, різноманітність активів, розвиток децентралізованих фінансів та регуляторні виклики, формують цей ринок. Незважаючи на ризики, пов'язані з інвестиціями в криптовалюти, інтерес до них продовжує зростати, що свідчить про їхнє потенційне майбутнє в глобальній економіці.

А щодо кращого розуміння, як працює ринок та можливість прогнозування, необхідно детальніше розібрати поняття блокчейну, з яким тісно пов'язана перша криптовалюта – Bitcoin.

Блокчейн — це дистрибутивна, розподілена технологія реєстрації даних, яка забезпечує безпечне, прозоре та незмінне збереження інформації. Блокчейн складається з серії блоків, кожен з яких містить певну кількість транзакцій. Кожен блок пов'язаний із попереднім за допомогою криптографічних методів, утворюючи ланцюг, що ускладнює можливість зміни даних. Основна концепція полягає в децентралізації, що означає, що дані зберігаються на багатьох комп'ютерах (вузлах) по всьому світу, що робить систему менш вразливою до збоїв або атак.

Блокчейн працює за принципом консенсусу, що дозволяє учасникам мережі погоджуватися на правильність транзакцій. Найбільш поширеними алгоритмами консенсусу є Proof of Work (PoW) [24] та Proof of Stake (PoS) [25].

Proof of Work: У цьому механізмі учасники мережі (майнери) виконують складні обчислення для підтвердження транзакцій. Перший, хто вирішить задачу,

отримує винагороду у вигляді криптовалюти. Наприклад, Bitcoin використовує PoW, де майнери конкурують, щоб отримати право додати новий блок до ланцюга.

Proof of Stake: В цьому механізмі учасники, які володіють криптовалютою, мають право підтверджувати транзакції пропорційно до кількості монет, які вони утримують. Це зменшує потребу в енергії та обчислювальних ресурсах. Ethereum планує перейти на PoS, що має зменшити його вуглецевий слід.

Блокчейн є основною технологією, на якій побудовані криптовалюти. Він забезпечує безпеку, прозорість і децентралізацію, що робить криптовалюти привабливими для користувачів. Кожна транзакція в мережі блокується та записується в блокчейн, що ускладнює можливість підробки або фальсифікації даних.

Основні переваги блокчейну для криптовалют включають:

- безпека: використання криптографії забезпечує високий рівень захисту даних. Наприклад, дані, збережені у блокчейні Bitcoin, не можуть бути змінені або видалені без згоди більшості учасників мережі;
- прозорість: усі транзакції доступні для перегляду будь-якому учаснику мережі, що підвищує рівень довіри серед користувачів. Це також дозволяє легко відстежувати рух коштів;
- децентралізація: відсутність центрального контролю робить блокчейн стійким до зловживань та шахрайства. Наприклад, користувачі можуть безпосередньо взаємодіяти один з одним без посередників;
- швидкість і ефективність: традиційні банківські системи можуть займати кілька днів для обробки міжнародних транзакцій. У блокчейн-системах транзакції можуть бути підтвержені за хвилини або навіть секунди, що значно підвищує ефективність.

Незважаючи на численні переваги, блокчейн стикається з рядом викликів, які потребують вирішення. Основні проблеми включають:

– масштабованість: з ростом популярності криптовалют зростає й навантаження на мережі. Наприклад, у 2017 році під час буму Bitcoin, транзакційні збори сягали рекордних рівнів, що призвело до затримок у підтвердженні транзакцій [26].

– енергоефективність: алгоритм PoW вимагає значних енергетичних витрат. Дослідження показує, що енергія, споживана майнерами Bitcoin, дорівнює енергії, споживаній цілою країною, такою як Аргентина.

– регуляторні питання: відсутність чітких регуляцій може призвести до невизначеності та ризиків для інвесторів. Багато країн все ще визначають, як управляти криптовалютами та блокчейн-технологіями, що може вплинути на їхнє прийняття.

Попри виклики, технологія блокчейн має великий потенціал для зміни фінансової системи. Прогнози вказують на те, що блокчейн буде використовуватися не лише у фінансах, але й у інших сферах, таких як управління ланцюгами постачання, охорона здоров'я, право, та енергетика.

Ринок блокчейн-технологій може зрости до 67.4 мільярдів доларів до 2026 року, що свідчить про зростаючий інтерес до цієї технології [27]. Крім того, з розвитком нових алгоритмів консенсусу, таких як Delegated Proof of Stake (DPoS) [28], та інтеграції з іншими технологіями, такими як інтернет речей (IoT), блокчейн має потенціал стати основою для нового покоління дистрибутивних систем.

Активний розвиток технологій блокчейну дуже чітко відображає графік зміни загальної ринкової капіталізації криптовалют протягом останніх кількох років (див. рис. 1.3).



Рисунок 1.3 – Діаграма загальної ринкової капіталізації криптовалют за останні 4 роки

1.2 Огляд предметної сфери прогнозування цін криптовалют за допомогою нейронних мереж

Криптовалюти є цифровими активами, які використовують криптографію для забезпечення безпеки транзакцій та контролю за створенням нових одиниць. Найвідоміша криптовалюта — це Bitcoin [3], але на даний момент існує безліч інших, таких як Ethereum, Litecoin, Ripple, та інші, які мають свої унікальні характеристики та механізми роботи. Ці валюти працюють на децентралізованих мережах блокчейн [29], що робить їх незалежними від центральних банків і урядів.

Останніми роками криптовалюти стали популярним інструментом для інвестування завдяки їхній волатильності — змінності цін, що дозволяє отримувати значні прибутки. Але ця ж волатильність створює високий рівень ризику для інвесторів. Тому передбачення майбутніх цін криптовалют стало надзвичайно актуальною темою для дослідників, трейдерів і фінансових аналітиків.

У предметній сфері прогнозування цін криптовалют застосовуються різноманітні методи машинного та глибокого навчання. Розглянемо основні з них:

Традиційні методи машинного навчання.

До класичних методів машинного навчання, які застосовуються для прогнозування фінансових ринків, включаючи криптовалют, належать алгоритми [30], такі як:

- Random Forest (випадковий ліс)
- Support Vector Machines (SVM) (метод опорних векторів)
- XGBoost (бустинг на градієнтних деревах)

Хоча ці методи добре працюють для аналізу фондових ринків, вони не завжди можуть повністю враховувати часові залежності, характерні для цін криптовалют, що змінюються з плином часу. Проблема полягає в тому, що криптовалюты мають інші ринкові фактори [31] (наприклад, новини про регуляції, ринкову психологію, вплив ринкових китів), які важко передбачити з використанням цих моделей.

Глибоке навчання і часові ряди.

У зв'язку з недостатньою ефективністю традиційних методів, було запропоновано використовувати моделі глибокого навчання, зокрема **рекурентні нейронні мережі (RNN)**. Зокрема, моделі на основі LSTM [32] (довготривала короткострокова пам'ять) та GRU [33] (скорочена рекурентна одиниця) дозволяють краще враховувати часові залежності, що є ключовими для прогнозування на основі історичних даних.

LSTM і GRU усувають проблему зникнення градієнту, з якою стикаються звичайні RNN, дозволяючи моделі «пам'ятати» ключові зміни на ринку криптовалют у довготривалій перспективі. Дослідження показали, що ці моделі можуть надавати точніші прогнози порівняно з класичними методами машинного навчання, оскільки вони краще адаптовані до непостійності та складних залежностей у даних криптовалют.

Ринок криптовалют значно відрізняється від традиційних фінансових ринків. Криптовалютні ринки є децентралізованими, що означає відсутність єдиної точки контролю або регулятора. Вони працюють цілодобово, що створює безперервний

обіг активів. Деякі ключові особливості, які роблять прогнозування на цьому ринку складнішим:

- висока волатильність: ціни криптовалют можуть змінюватися на десятки відсотків за дуже короткий проміжок часу.
- ринкова психологія: новини, регулювання та навіть публічні висловлювання ключових фігур (наприклад, Ілона Маска [34]) можуть значно впливати на ціни.
- спекулятивний характер: більшість трейдерів використовують криптовалюту не як засіб обміну, а як інструмент для спекуляцій.

Хоча вже існує велика кількість досліджень, які спрямовані на прогнозування цін криптовалют, багато з них обмежуються використанням стандартних методів, які були розроблені для фондових ринків. Наприклад, часто ігноруються унікальні характеристики криптовалют, такі як транзакційні графи або блокчейн-дані (кількість транзакцій, хешрейт тощо).

Одним із найбільш перспективних підходів для вирішення проблеми прогнозування криптовалют є поєднання кількох моделей, таких як LSTM та вбудовані мережі (Embedding Networks [35]), які можуть враховувати взаємодії між різними криптовалютами та історичними даними. Наприклад, дослідження [36] пропонує модель, яка поєднує LSTM для захоплення часових залежностей та вбудовану мережу для врахування прихованих представлень зв'язаних криптовалют.

1.3 Огляд наявних досліджень та публікацій. Проведення аналогій

Алгоритм Random Forest є популярним завдяки своїй стійкості до перенавчання та здатності працювати з великими обсягами даних. У дослідженні [37] показано, що його точність передбачення криптовалют становить 87,73%. Це досить непоганий результат, оскільки алгоритм об'єднує кілька дерев прийняття рішень, що дозволяє зменшити ймовірність випадкових помилок. Проте цей показник є недостатнім для високоточних прогнозів, особливо у порівнянні з

нейронними мережами. Випадковий ліс також має певні обмеження у відношенні до динамічних змін на ринку криптовалют, де важливо враховувати не лише історичні ціни, але й інші показники, наприклад соціальні тренди та поведінку інвесторів.

У тому ж дослідженні [37] було проведено порівняння з алгоритмом К-найближчих сусідів (KNN). Точність цього методу становить 72,15%, що є нижчим результатом, ніж у випадкового лісу. Алгоритм KNN добре працює для невеликих вибірок, але стає менш ефективним для великих даних або складних багатовимірних систем, таких як криптовалютний ринок. Цей метод також чутливий до шумів у даних, що може знижувати точність прогнозів.

Методи Support Vector Machine (SVM), як показано в дослідженнях [38, 39], продемонстрували широкий діапазон точності – від 47,7% до 95,5%. Така мінливість результатів значною мірою залежить від розміру датасету та параметрів моделі. При невеликих наборах даних або недостатці важливих факторів, які враховуються в прогнозах, точність SVM суттєво знижується. З іншого боку, у випадках з більшими та якіснішими даними, результати цього методу можуть бути доволі точними, але все ще поступаються нейронним мережам у передбаченні складних патернів.

У дослідженнях [40, 41] алгоритм XGBoost демонструє результати, які є схожими на SVM, але стабільнішими. Значення метрик MAE та RMSE для XGBoost складають 46,22 та 59,95 відповідно, що свідчить про помірні похибки. Показник R^2 становить 0,90, що означає, що алгоритм добре відтворює варіації у даних. Проте, навіть із такою високою точністю, XGBoost не в змозі конкурувати з RNN у врахуванні динамічних змін на ринку, оскільки він здебільшого працює з історичними даними без інтеграції нових метрик або показників.

У поточній роботі використовуватиметься RNN, яка має здатність ефективно опрацьовувати послідовні дані та виявляти довготривалі залежності. Це дозволяє моделі краще передбачати цінові коливання криптовалют, враховуючи як короткострокові, так і довгострокові тренди. Рекурентні нейронні мережі здатні

опрацьовувати не лише історію цін, але й інші фактори, такі як обсяги торгів, поведінкові індикатори та соціальні тренди. Це робить їх значно ефективнішими в порівнянні з методами, що базуються виключно на історичних даних.

Хоча такі методи, як Random Forest, KNN, SVM, та XGBoost можуть показувати доволі високі результати в прогнозуванні криптовалют, вони мають обмеження через використання лише історичних даних. RNN, завдяки своїй здатності враховувати послідовності та додаткові метрики, демонструє вищу точність та здатність передбачати складні патерни на ринку криптовалют. Важливою перевагою RNN є можливість інтеграції зовнішніх факторів і здатність до адаптації в реальному часі.

Проаналізовуючи зі схожими публікаціями [42, 43], де в якості алгоритму використовувався RNN, їх об'єднує головний недолік – датасети, що базуються лише на основі історичних показників цін, що не враховують інші метрики, що і є головною задачею даної роботи – дослідити роботоспроможність моделі, навченої на основі алгоритму RNN, враховуючи різні складені показники та метрики.

Висновок до розділу 1

У першому розділі було проведено ґрунтовний аналіз стану розвитку криптовалютного ринку та основних його особливостей, включаючи технологію блокчейн. Огляд історії появи та розвитку криптовалют (в особливості популярних, таких як Bitcoin та Ethereum) дає відповідь на питання про актуальність та необхідність створити рішення прогнозу цін криптовалют. Огляд предметної сфери прогнозування цін криптовалют за допомогою нейронних мереж дозволив визначити ключові виклики та перспективи цього напрямку, а також показав, що використання рекурентних нейронних мереж (RNN) є одним із найпоширеніших підходів для вирішення задач прогнозування в умовах високої волатильності. Особливо для виконання даного завдання пасують дві поширені архітектури: LSTM та GRU.

Також було проаналізовано актуальні дослідження та публікації в цій галузі. Проведення аналогій із уже існуючими системами дало змогу виявити як їхні сильні сторони, так і недоліки, що дозволяє обґрунтувати необхідність розробки інтелектуальної системи прогнозування цін криптовалют, заснованої на використанні нейронних мереж. Серед недоліків (які практична частина даного дослідження повинна вирішити) є недостатня точність, нестабільність, що залежить від певного набору даних або складність навчання.

Зроблений аналіз підкреслив важливість вибору ефективних методів аналізу даних і підходів до побудови моделей, що враховують як ринкові, так і позаринкові фактори, такі як новинний фон та макроекономічні показники, а також і інші параметри, які включатимуться в навчання моделі.

2 ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ДЛЯ ПРОГНОЗУВАННЯ ЦІН НА КРИПТОВАЛЮТИ

2.1 Огляд методів прогнозування

Прогнозування в фінансовій сфері, зокрема на ринку криптовалют, є важливою задачею, яка передбачає використання різних методів і моделей для оцінки майбутніх цін активів. З урахуванням високої волатильності та непередбачуваності ринку криптовалют, правильний вибір методів прогнозування має критичне значення для інвесторів та трейдерів. У цьому розділі ми розглянемо основні методи прогнозування, що використовуються для аналізу цін криптовалют, їх переваги та недоліки, а також можливі комбінації цих методів.

Лінійна регресія є простим і широко використовуваним методом, який дозволяє моделювати взаємозв'язок між незалежними (пояснюючими) і залежними (відповідними) змінними. У випадку криптовалют, можна використовувати лінійну регресію для оцінки залежності цін від різних факторів, таких як обсяги торгівлі, новини та соціальні медіа. Основні переваги цього методу включають простоту реалізації та інтерпретації. Однак лінійна регресія (див. рис. 2.1) може бути недостатньо точною у випадках, коли існують нелінійні взаємозв'язки.

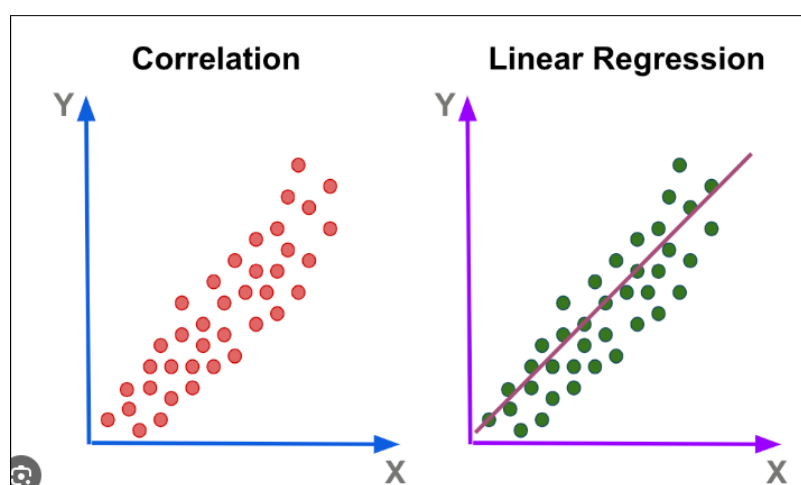


Рисунок 2.1 – Приклад лінійної регресії

ARIMA (AutoRegressive Integrated Moving Average) є популярним методом для прогнозування часових рядів, який враховує як авторегресивні (AR), так і середні (MA) компоненти. ARIMA (див. рис. 2.2) може бути дуже ефективною для прогнозування цін криптовалют, особливо у випадках, коли дані є стаціонарними. Основна перевага ARIMA полягає в її здатності моделювати сезонність та тренди. Однак для ефективного використання цього методу необхідно провести попередню обробку даних і перевірити їх на стаціонарність.

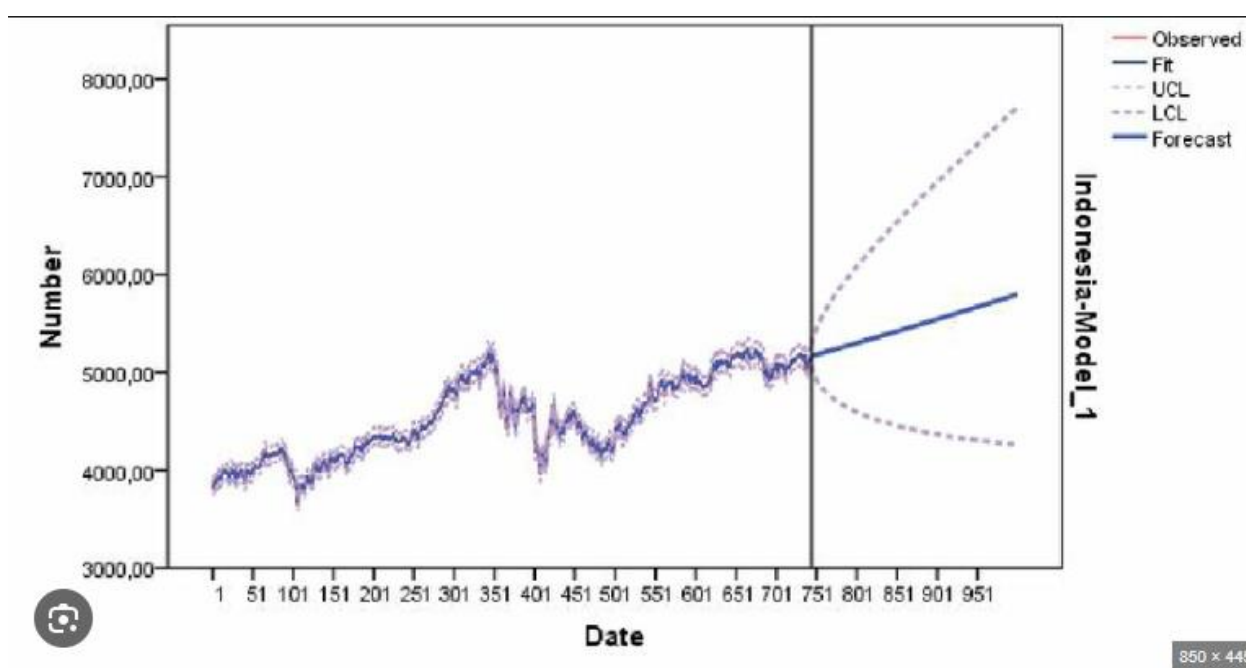


Рисунок 2.2 – Приклад графіку ARIMA

Дерев'яні моделі (Decision Trees (див. рис. 2.3), Random Forests) використовують структуру дерева для прийняття рішень на основі певних ознак. Random Forest, в свою чергу, є ансамблевим методом, який поєднує кілька дерев рішень, що робить його більш стабільним і менш вразливим до перенавчання. Ці методи добре справляються з великими наборами даних і можуть моделювати складні нелінійні залежності. Однак їхня інтерпретація може бути складною, особливо для Random Forest.

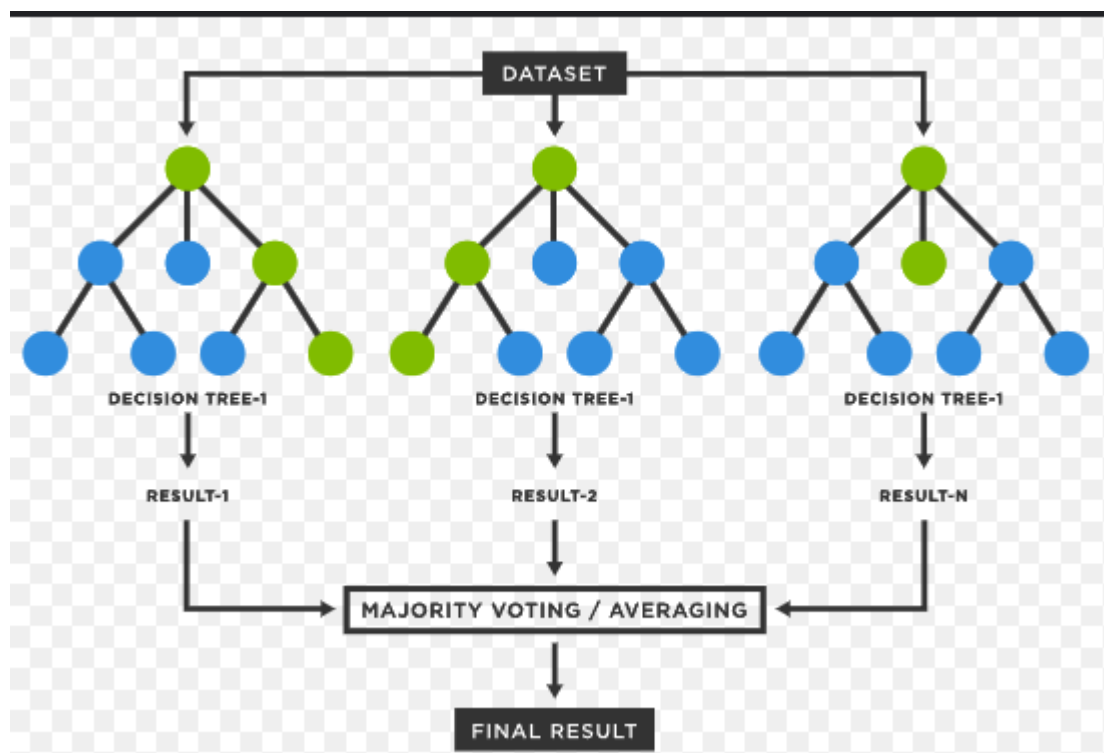


Рисунок 2.3 – Принцип роботи Decision Tree моделі

Нейронні мережі, і зокрема рекурентні нейронні мережі (RNN) (див. рис. 2.4) та їх варіанти, такі як LSTM (Long Short-Term Memory), здатні моделювати складні шаблони в даних. Вони особливо корисні для прогнозування часових рядів, оскільки можуть враховувати попередні значення для визначення майбутніх. Переваги нейронних мереж включають їхню гнучкість і здатність адаптуватися до нових даних, однак їхнє навчання може бути вимогливим до обчислювальних ресурсів і часу. Якщо казати про переваги LSTM, то це є модифікацією RNN алгоритму, але з покращеною пам'яттю, адже для RNN є відомим викликом саме проблема зі збереженням даних про дуже далеке минуле.

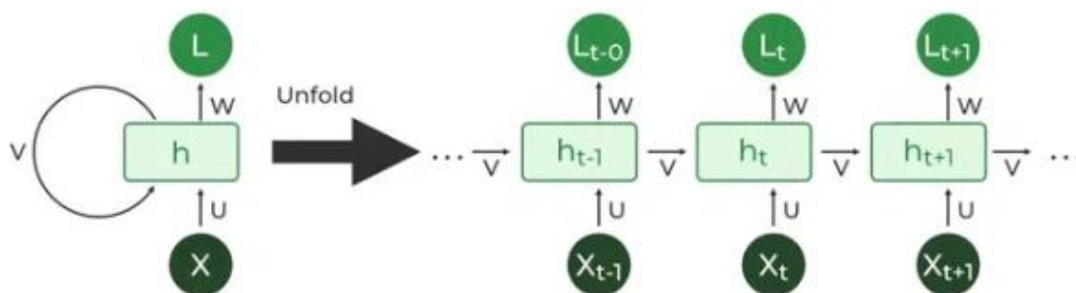


Рисунок 2.4 – Архітектура RNN алгоритму

Гібридні моделі поєднують кілька методів для досягнення кращих результатів. Наприклад, можна використовувати ARIMA для моделювання загальних трендів та RNN для виявлення короткострокових патернів. Такий підхід дозволяє зберегти переваги обох методів і знизити їх недоліки. У літературі є кілька досліджень, які підтверджують ефективність таких комбінацій для прогнозування цін криптовалют.

2.2 Інформаційні технології для збору та аналізу даних

Сучасні інформаційні технології забезпечують широкий спектр інструментів і методів для збору, обробки та аналізу даних, особливо у сфері фінансових ринків, таких як ринок криптовалют. Ефективний збір та аналіз даних є критично важливими для створення точних моделей прогнозування.

API (Application Programming Interface) є потужним інструментом для автоматизованого збору даних з різних джерел. Багато криптовалютних бірж, таких як Binance, Coinbase і Kraken, надають API для доступу до історичних і реальних даних про ціни, обсяги торгів та інші показники. Використання API дозволяє отримувати дані в реальному часі та інтегрувати їх у системи аналізу.

Веб-скрапінг — це процес автоматизованого збору даних з веб-сайтів. За допомогою спеціалізованих бібліотек, таких як BeautifulSoup або Scrapy, можна витягувати інформацію, таку як новини, аналітичні статті або ціни з різних сайтів.

Існують також публічно доступні датасети, які містять історичні дані про ціни криптовалют, обсяги торгів та інші метрики. Такі ресурси, як Kaggle або CryptoCompare, надають готові датасети для аналізу.

Для зберігання великих обсягів даних використовуються реляційні (SQL) та нереляційні (NoSQL) бази даних. Реляційні бази даних, такі як MySQL або PostgreSQL, добре підходять для структурованих даних, тоді як NoSQL рішення, такі як MongoDB, краще справляються з неструктурованими даними.

Для обробки та очищення даних можна використовувати мови програмування, такі як Python та R, а також бібліотеки, такі як Pandas, NumPy та Dplyr. Ці інструменти дозволяють виконувати різноманітні операції з даними: фільтрацію, агрегацію, трансформацію тощо.

Статистичні методи, такі як регресія, кореляційний аналіз і тестування гіпотез, використовуються для аналізу даних. Ці методи дозволяють виявити закономірності, тренди та залежності між змінними.

Для аналізу великих обсягів даних і створення прогнозних моделей використовуються алгоритми машинного навчання. Такі бібліотеки, як Scikit-learn, TensorFlow та Keras, надають інструменти для побудови, навчання та оцінювання моделей.

Для представлення результатів аналізу використовуються інструменти для візуалізації, такі як Matplotlib, Seaborn або Tableau. Візуалізація даних допомагає виявити тренди, аномалії та кореляції, що можуть бути неочевидні при простому перегляді таблиць.

Серед усього вищеперерахованого, було обрано такий набір технологій та засобів:

- API та web-сервіси для збору даних;
- Excel для зберігання невеликого об'єму даних та нескладної обробки даних;
- мова програмування Python, бібліотеки Pandas та NumPy для обробки деяких показників;

- бібліотека TensorFlow для машинного навчання;
- бібліотека Matplotlib для візуалізації даних.

2.3 Процес отримання даних

Отримання даних є необхідною складовою для будь-якого аналізу. В даному випадку, необхідно отримати дані про коливання ціни певної криптовалюти протягом певного проміжку часу.

Для отримання цих даних можна підійти кількома шляхами. Перший і логічний шлях – отримання даних з публічних API. Зазвичай, усі криптовалютні біржі пропонують свої API для користувачів, що здавалося б, є чудовим способом для витягування даних. Для цього було обрано доволі популярну біржу в Україні – Binance. Після пошуку документації щодо користування Binance API, наступним кроком було отримання унікального API-ключа. Але цей крок неможливо закінчити безкоштовним шляхом, адже правила (див. рис. 2.5) потребують верифікації особи та внесення певної суми на гаманець користувача.

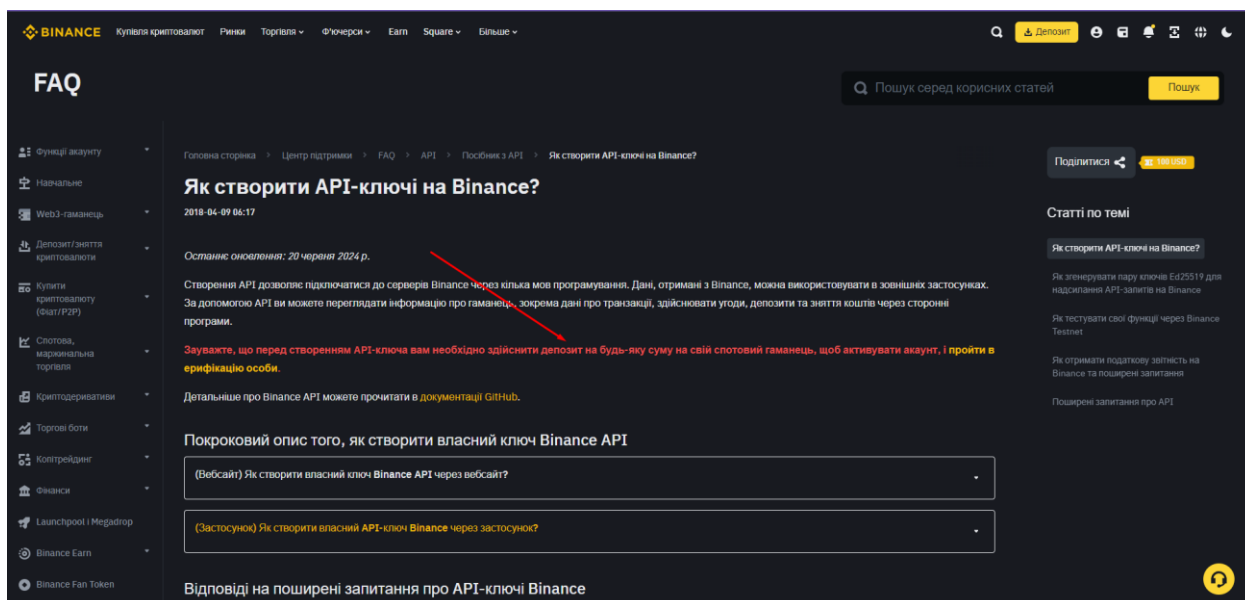
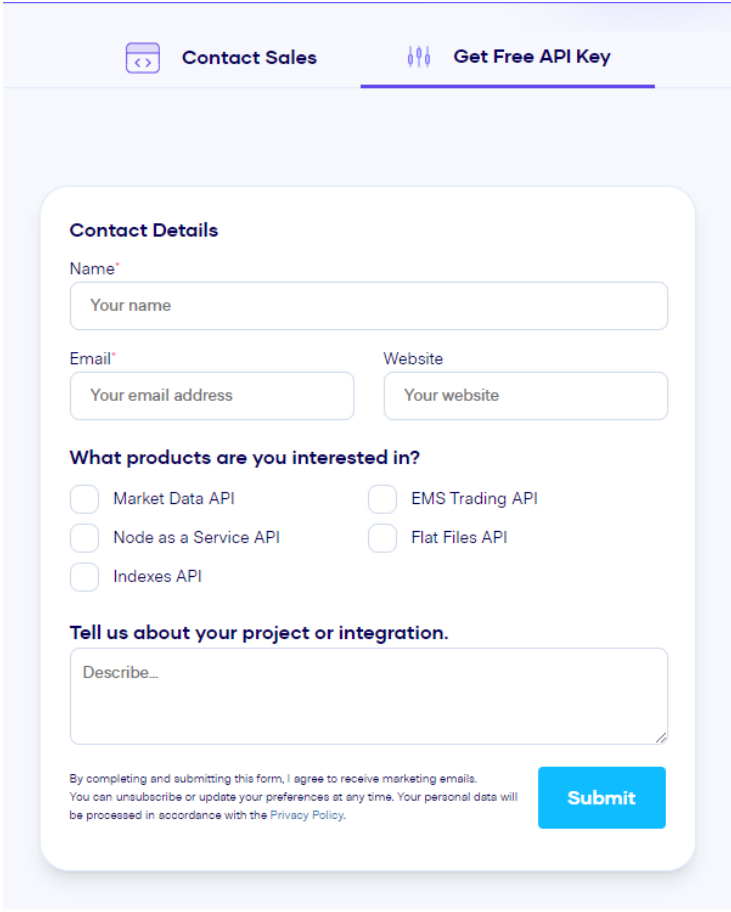


Рисунок 2.5 – Правила отримання API-ключа на біржі Binance

Отже, висновок можна зробити такий, що отримання даних щодо криптовалют з біржі може викликати певні труднощі. Тому наступним рішенням

було скористуватися “third-party” сервісами, метою яких і є збір даних з багаточисельних криптобірж і надання їх користувачам. Одним із таких було обрано CoinAPI: сервіс має безліч категорій даних щодо криптовалют, безліч криптобірж, з яких тягнуться дані та безліч криптовалют.

Отримати API-ключ було доволі не складно: після заповнення простої форми (рис. 2.6), ключ прийшов на пошту (рис. 2.7).



The image shows a web form titled "Get Free API Key" from CoinAPI. The form is divided into several sections:

- Contact Details:** Includes a "Name" field with the placeholder "Your name", an "Email" field with "Your email address", and a "Website" field with "Your website".
- What products are you interested in?:** A list of five checkboxes: "Market Data API", "EMS Trading API", "Node as a Service API", "Flat Files API", and "Indexes API".
- Tell us about your project or integration:** A large text area with the placeholder "Describe...".
- Footer:** A small disclaimer: "By completing and submitting this form, I agree to receive marketing emails. You can unsubscribe or update your preferences at any time. Your personal data will be processed in accordance with the Privacy Policy." and a blue "Submit" button.

Рисунок 2.6 – Заповнення форми для отримання API-ключа в сервісі CoinAPI

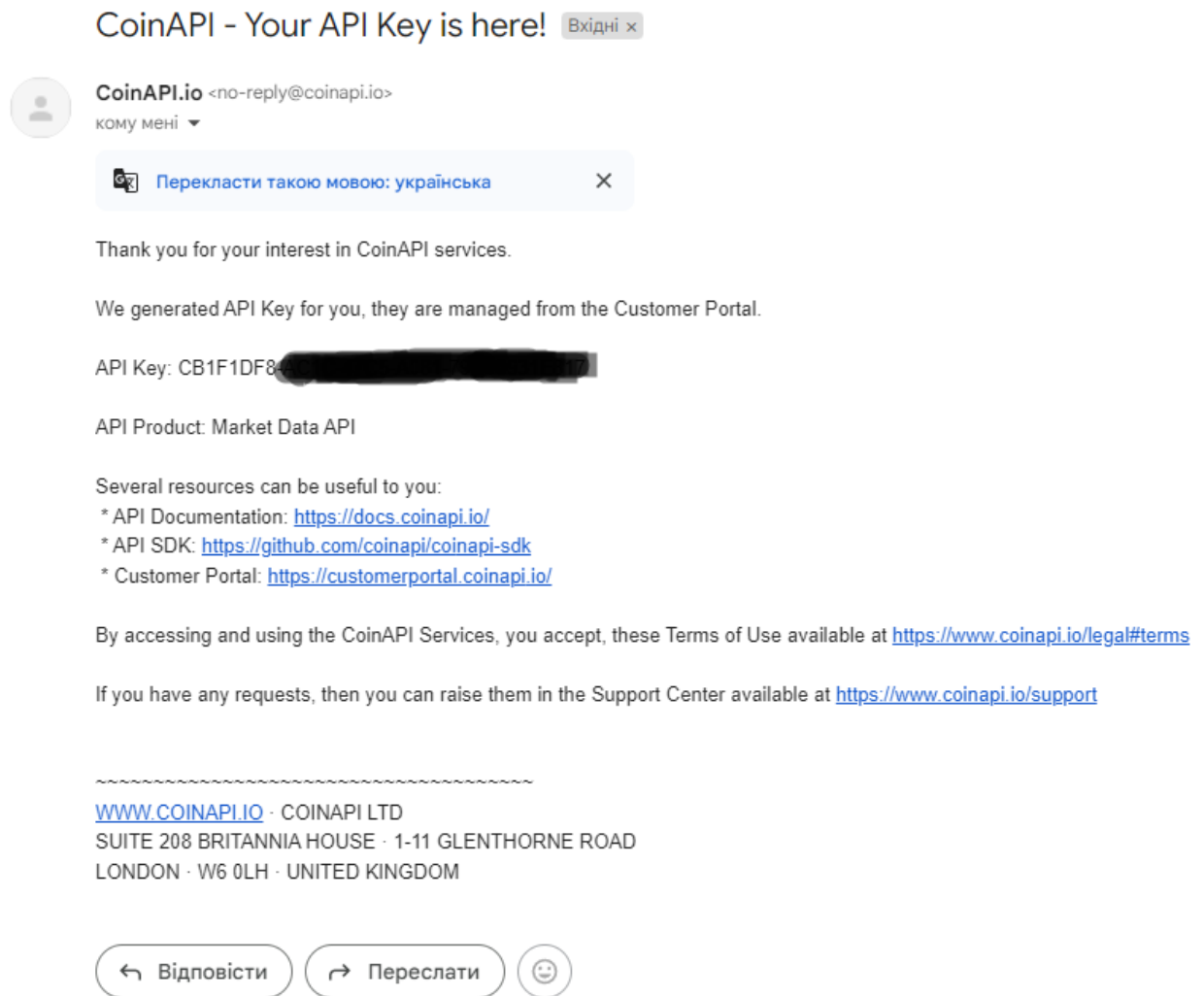


Рисунок 2.7 – Отриманий API-ключ

Даний сервіс пропонує нам кілька варіантів роботи з API, серед них є websockets, JSON RPC, FIX API тощо. Найпростіший є REST API, саме тому й було його обрано. Звертання до ендпоінтів застосовується за допомогою програми Postman.

Наступним кроком стає налаштування авторизації в Postman за допомогою API-ключа (рис. 2.8), після чого необхідно обрати необхідний запит (наприклад, отримання цінового відношення BTC/USD на біржі Binance протягом певного дня).

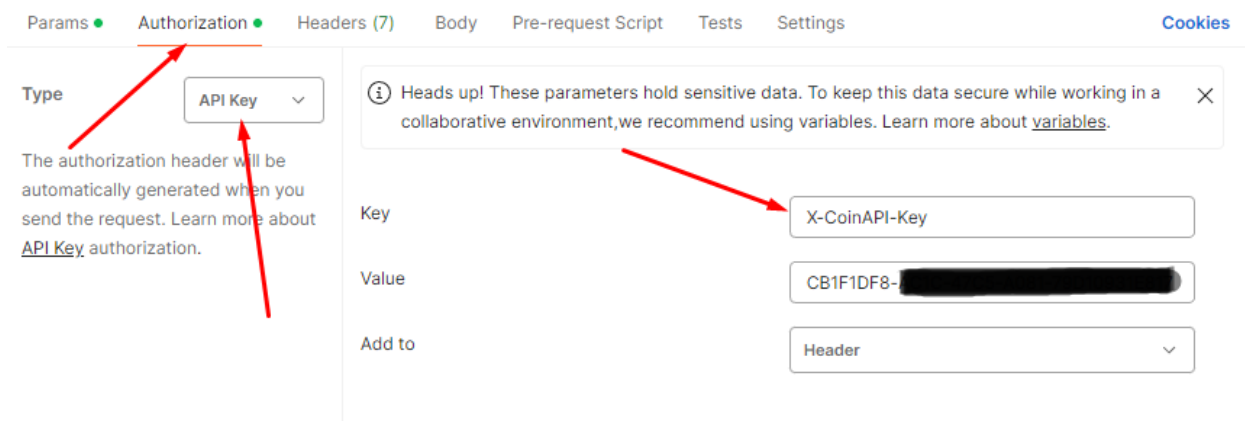


Рисунок 2.8 – Налаштування авторизації в Postman

Запит подібного характеру має назву OHLCV (open, high, low, close, volume) data – дані про ціну на початок періоду, на кінець періоду, найбільша та найменша ціна протягом періоду і об'єм валюти під час торгівлі протягом періоду.

Отже, запит сформований і готовий до відправки (див. рис. 2.9).

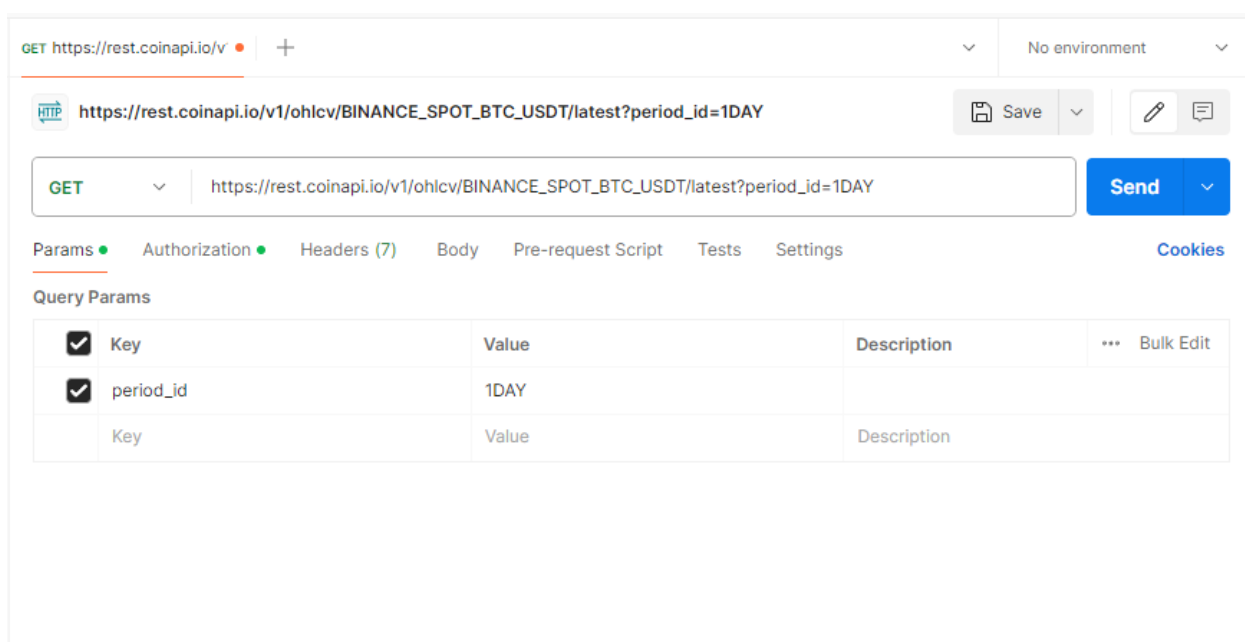


Рисунок 2.9 – Підготовка запиту до виконання

Отриманий результат представлений у форматі JSON (рис. 2.10), що є зручним для обробки за допомогою коду.

```

98
99     "time_period_start": "2024-09-02T00:00:00.0000000Z",
100    "time_period_end": "2024-09-03T00:00:00.0000000Z",
101    "time_open": "2024-09-02T00:00:00.0020000Z",
102    "time_close": "2024-09-02T23:59:59.7240000Z",
103    "price_open": 57301.77,
104    "price_high": 59425.69,
105    "price_low": 57128.27,
106    "price_close": 59132.13,
107    "volume_traded": 22472.92456999998,
108    "trades_count": 1931264
109

```

Рисунок 2.10 – Приклад відповіді

В документації вказано, що за замовчуванням у відповіді міститься до 100 результатів за обраним періодом (але в параметрах можна вказати більшу кількість запитів). Єдиною перепоною для використання цього варіанту для отримання даних є лімітованість – даний сервіс дозволяє обробити лише до 100 запитів на день (див. рис. 2.11) (кожен запит має власну вагу, через наявність «важких» запитів, сервіс обробить менше 100 запитів за день). Перевірити свою квоту можна за допомогою headers, що надходять разом з відповіддю сервісу:

Key	Value
content-type	application/json; charset=utf-8
date	Tue, 10 Sep 2024 16:31:15 GMT
content-encoding	br
vary	Accept-Encoding
x-ratelimit-quota-allocated	100
x-ratelimit-quota-used	0
x-ratelimit-quota-remaining	100
x-ratelimit-request-cost	1
transfer-encoding	chunked

Рисунок 2.11 – Квота використання запитів до сервісу на день

Allocated – квота, виділена користувачу на день за його тарифом, used – використана вага серед виділеної, remaining – залишок ваги, cost – вага запиту.

Тобто, у висновку можна сказати, що такий варіант підходить для створення власних рет-проектів та невеликих досліджень, що не потребують великої кількості даних та запитів.

Третім варіантом став пошук сервісу, який збирає дані і надає можливість просто викачати необхідні дані за необхідний період. Таким сервісом стає cryptodatadownload.com (див. рис. 2.12). Дані можна отримати з великої кількості бірж і навіть без реєстрації, що робить підготовку до аналізу даних максимально швидкою та простою:

SYMBOL LIST FOR GEMINI DAILY AND HOURLY TIMEFRAMES

Show entries Search:

Symbol	Timeframe	First Date Available	Last Date Available	File Link
IINCH/USD	day	2021-04-05 04:00:00	2024-02-04 00:00:00	Download CSV
IINCH/USD	hour	2022-02-04 04:00:00	2024-02-04 23:00:00	Download CSV
AAVE/USD	day	2021-04-05 04:00:00	2024-09-08 00:00:00	Download CSV
AAVE/USD	hour	2022-02-04 04:00:00	2024-09-08 23:00:00	Download CSV
ALCX/USD	day	2021-06-17 04:00:00	2024-02-04 00:00:00	Download CSV
ALCX/USD	hour	2022-02-04 04:00:00	2024-02-04 23:00:00	Download CSV
AMP/USD	day	2020-09-15 04:00:00	2024-09-08 00:00:00	Download CSV
AMP/USD	hour	2020-10-31 04:00:00	2024-09-08 23:00:00	Download CSV
ANKR/USD	day	2021-06-17 04:00:00	2024-09-08 00:00:00	Download CSV
ANKR/USD	hour	2022-02-04 04:00:00	2024-09-08 23:00:00	Download CSV

Showing 1 to 10 of 194 entries Previous 2 3 4 5 ... 20 Next

Рисунок 2.12 – Пошук даних про криптовалюту

Дані доступні щодобово та щогодинно, включаючи багато різних криптовалют та різних відношень і все у форматі .csv (див. рис. 2.13), що робить зручною обробку даних в Microsoft Excel.

	A	B	C	D	E	F	G	H	I	J	K	L
1	https://www.CryptoDataDownload.com											
2	unix,date,symbol,open,high,low,close,Volume BTC,Volume USD											
3	1725667200000,2024-09-07 00:00:00,BTC/USD,53951.3,54831.34,53743.65,54151.72,109.13587398,5909895.289720246											
4	1725580800000,2024-09-06 00:00:00,BTC/USD,56172.52,57000.0,52553.68,53951.3,1172.88047342,63278426.28562445											
5	1725494400000,2024-09-05 00:00:00,BTC/USD,57960.93,58303.4,55654.0,56172.52,716.9642867,40273690.73394148											
6	1725408000000,2024-09-04 00:00:00,BTC/USD,57514.16,58525.12,55591.06,57960.93,853.29733828,49457907.2932334											
7	1725321600000,2024-09-03 00:00:00,BTC/USD,59091.35,59817.86,57415.64,57514.16,587.25338622,33775385.21559888											
8	1725235200000,2024-09-02 00:00:00,BTC/USD,57321.57,59415.64,57120.71,59091.35,163.60767989,9667798.67506795											
9	1725148800000,2024-09-01 00:00:00,BTC/USD,58967.39,59054.15,57172.0,57321.57,105.58898413,6052526.345036685											
10	1725062400000,2024-08-31 00:00:00,BTC/USD,59115.12,59449.34,58756.86,58967.39,45.26103586,2668925.1533606052											
11	1724976000000,2024-08-30 00:00:00,BTC/USD,59389.91,59906.23,57699.76,59115.12,583.91655902,34518297.45645439											
12	1724889600000,2024-08-29 00:00:00,BTC/USD,59058.94,61188.11,58784.74,59389.91,539.26181603,32026710.720458258											
13	1724803200000,2024-08-28 00:00:00,BTC/USD,59446.47,60220.13,57874.85,59058.94,726.39603656,42900179.93943485											
14	1724716800000,2024-08-27 00:00:00,BTC/USD,62871.45,63218.22,58046.42,59446.47,765.64650075,45514981.73743985											
15	1724630400000,2024-08-26 00:00:00,BTC/USD,64273.19,64500.0,62822.77,62871.45,573.15153694,36034868.19714636											
16	1724544000000,2024-08-25 00:00:00,BTC/USD,64173.0,65000.0,63830.68,64273.19,127.39159729,8187864.337023656											
17	1724457600000,2024-08-24 00:00:00,BTC/USD,64092.32,64506.68,63625.51,64173.0,97.5155125,6257862.9836625											
18	1724371200000,2024-08-23 00:00:00,BTC/USD,60382.2,64986.78,60360.0,64092.32,786.22897133,50391238.823753186											
19	1724284800000,2024-08-22 00:00:00,BTC/USD,61154.5,61421.3,59758.56,60382.2,677.26626825,40894827.26272515											
20	1724198400000,2024-08-21 00:00:00,BTC/USD,59019.79,61838.68,58799.55,61154.5,828.34011235,50656725.40070808											
21	1724112000000,2024-08-20 00:00:00,BTC/USD,59478.67,61427.24,58581.28,59019.79,808.70173543,47729406.597714156											
22	1724025600000,2024-08-19 00:00:00,BTC/USD,58440.34,59626.68,57848.67,59478.67,479.2491949,28505104.711222783											
23	1723939200000,2024-08-18 00:00:00,BTC/USD,59518.27,60266.46,58400.0,58440.34,155.99572286,9116443.082484173											
24	1723852800000,2024-08-17 00:00:00,BTC/USD,58887.62,59700.0,58820.49,59518.27,86.19853704,5130387.801151721											
25	1723766400000,2024-08-16 00:00:00,BTC/USD,57547.19,59869.54,57099.03,58887.62,596.30852896,35115190.05615547											
26	1723680000000,2024-08-15 00:00:00,BTC/USD,58723.0,59851.19,56169.07,57547.19,796.37376391,45829072.30274391											
27	1723593600000,2024-08-14 00:00:00,BTC/USD,60632.72,61799.37,58480.58,58723.0,776.00539847,45569365.01435381											
28	1723507200000,2024-08-13 00:00:00,BTC/USD,59368.58,61609.11,58436.51,60632.72,632.09556852,38325673.61931398											
29	1723420800000,2024-08-12 00:00:00,BTC/USD,58712.65,60711.68,57673.0,59368.58,722.83438035,42913650.7365594											
30	1723334400000,2024-08-11 00:00:00,BTC/USD,60928.12,61850.36,58312.37,58712.65,259.64845354,15244648.775735281											
31	1723248000000,2024-08-10 00:00:00,BTC/USD,60856.63,61482.9,60287.13,60928.12,146.07097169,8899829.691644922											
32	1723161600000,2024-08-09 00:00:00,BTC/USD,61684.87,61758.63,59580.3,60856.63,637.25455802,38781164.85323667											
33	1723075200000,2024-08-08 00:00:00,BTC/USD,55152.38,62672.97,54761.73,61684.87,1084.55956737,66900915.92047469											
34	1722988800000,2024-08-07 00:00:00,BTC/USD,56066.45,57750.0,54565.88,55152.38,867.37589928,47837845.199932285											
35	1722902400000,2024-08-06 00:00:00,BTC/USD,54007.25,57122.0,53982.91,56066.45,907.71902738,50892583.462649405											
36	1722816000000,2024-08-05 00:00:00,BTC/USD,58132.77,58280.0,49259.0,54007.25,2430.12457811,131244345.6211313											
37	1722729600000,2024-08-04 00:00:00,BTC/USD,60654.71,61064.88,57177.0,58132.77,313.03457304,18197566.83658252											

Рисунок 2.13 – Приклад даних, отриманих з сервісу

Висновок: було розглянуто три варіанти отримання даних про криптовалюту для їхнього аналізу: отримання даних з API певної біржі, отримання даних з API “third-party” сервісів, отримання даних шляхом скачування .csv-файлу з сервісу, що зберігає дані про криптовалюту. Серед переваг та недоліків першого варіанту:

Переваги:

- розширений інструментарій для зареєстрованих на біржі користувачей;
- можливість не тільки отримувати дані з API, а й користуватися POST-запитами для розширення своїх можливостей.
- налагоджена безпека користування API;

- можливість отримувати дані в реальному часі;
- можливість отримувати дані автоматизовано.

Недоліки:

- зазвичай тривала і складна реєстрація та верифікація;
- іноді, користування API може бути платним;
- API-документація є складнішою через наявність більших інструментів;
- можливість виконати небажані дії через наявність POST-запитів.

Щодо другого варіанту:

Переваги:

- проста реєстрація;
- широкий спектр запитів та можливостей користування API-сервісом;
- користування може бути безкоштовним;
- можливість обробити відповідь за допомогою коду (наприклад, за допомогою Python);
- можливість отримувати дані в реальному часі;
- можливість отримувати дані автоматизовано.

Недоліки:

- ліміт безкоштовних запитів.

Щодо третього варіанту:

Переваги:

- реєстрація не є обов'язковою.
- великий вибір криптобірж та відношень між валютами
- доступ до даних є абсолютно безкоштовним у будь-якій кількості;
- можливість обробити відповідь за допомогою коду та Microsoft Excel.

Недоліки:

- обмежений спектр можливостей (лише дані про відношення між валютами);
- обмежений вибір періодичності (лише за подовово та погодинно);
- користування API є платним;
- даний вибір підходить лише для ручного отримання даних і лише для роботи з історичними даними.

2.4 Аналіз отриманих даних

Отже, після отримання даних про криптовалюту, наступним кроком стає аналіз. Загалом, криптовалютний аналіз можна розділити на кілька типів: трендовий, технічний, аналіз волатильності, аналіз кореляцій, сезонний тощо. За автоматизацією можна розподілити на ручне вирахування та отримання даних з інших веб-сервісів.

Трендовий аналіз

Розглянемо спочатку трендовий аналіз. Для цього знадобляться дані про ціни криптовалюти за певну дату. З отриманих даних (а саме дані OHLC) у вигляді .csv, вираховуємо середню ціну за формулою 2.1.

$$\text{Середня ціна за день} = \frac{x_{\text{поч}} + x_{\text{кін}} + x_{\text{вищ}} + x_{\text{нижч}}}{4}, \quad (2.1)$$

де $x_{\text{поч}}$ – ціна на початок дня;

$x_{\text{кін}}$ – ціна на кінець дня;

$x_{\text{вищ}}$ – найвища ціна;

$x_{\text{нижч}}$ – найнижча ціна.

Для обробки даних з файлу .csv можна скористатися будь-яким існуючим парсером або написаним власноруч за допомогою різноманітних мов програмування. В даному випадку, було написано парсер на мові С# (дод. А), який обрахував ціни, а потім було винесено в окремі 2 текстові файли ціни та відповідні

дати. Ці дані були завантажені в файл Microsoft Excel, де його засобами було побудовано тренд (див. рис. 2.14).

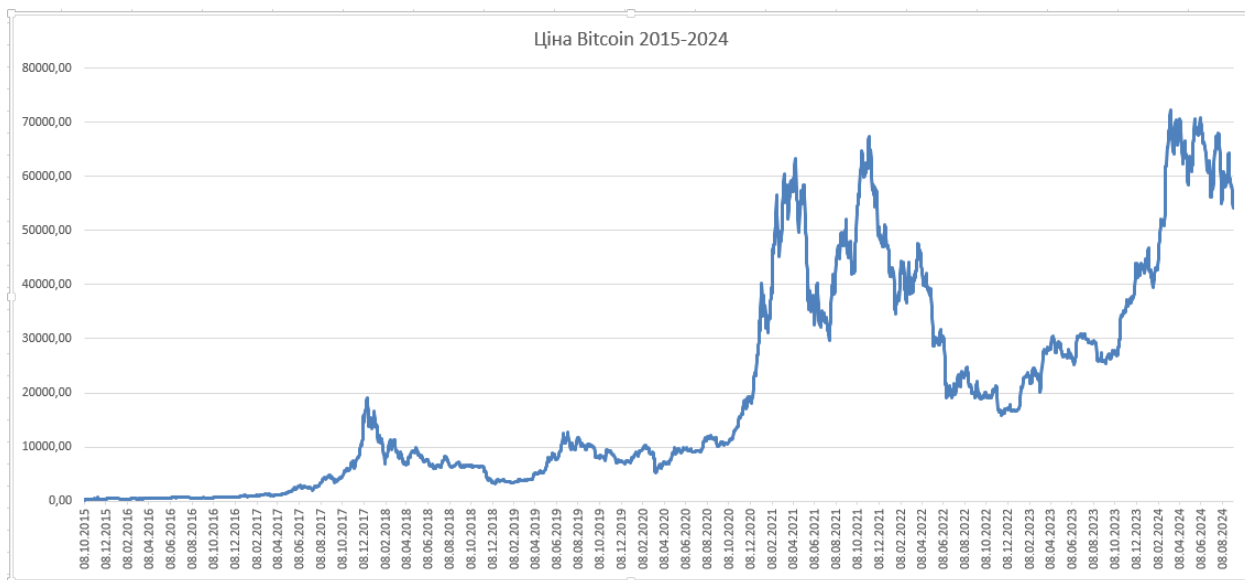


Рисунок 2.14 – Тренд ціни Bitcoin

Цей графік дає мало інформації, тому для практичності, варто будувати графік (див. рис. 2.15) за нетривалий і сучасний проміжок, наприклад за останні 30 днів (дані актуальні на 07.09.2024).



Рисунок 2.15 – Тренд ціни Bitcoin за останні 30 днів

За даним графіком можна сказати, що валюту **не можна назвати доволі стабільною** протягом останніх 30 днів, а сам тренд показує дещо **спадний** напрямок.

Для порівняння можна обрати другу за популярністю валюту – Ethereum. Ось її тренд за останні 30 днів (див. рис. 2.16).

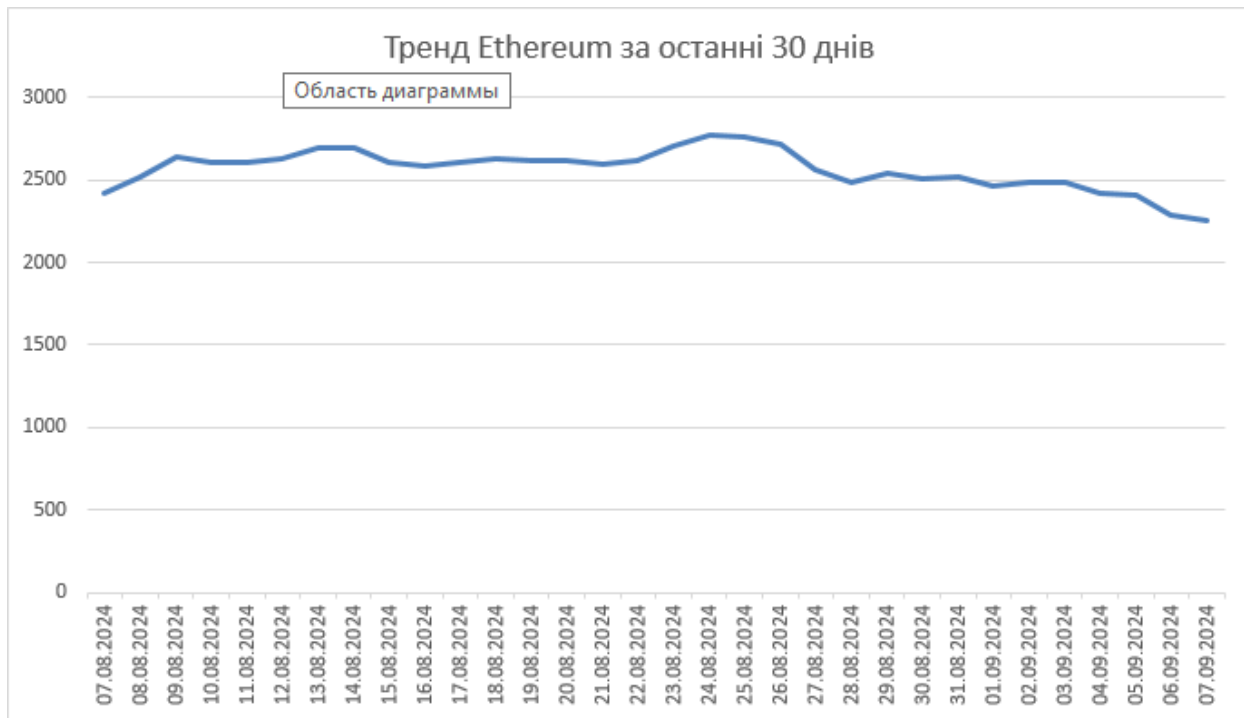


Рисунок 2.16 – Тренд Ethereum за останні 30 днів

У той же час, Ethereum показує себе **стабільно**, ціна майже не змінилась за останні 30 днів, що є **плюсом** для даної валюти.

Технічний аналіз

Зазвичай під технічним аналізом мають на увазі аналіз RSI (Relative Strength Index), що підказує, чи є актив перекупленим чи перепроданим і допомагає зрозуміти, коли варто купувати валюту, а коли продавати.

Для аналізу використовують зазвичай *close price* за останні 14 інтервалів (в даному випадку, за останні 14 днів). Розраховується це все за формулою 2.2.

$$RSI = 100 - \left(\frac{100}{1+RS} \right), \quad (2.2)$$

де RS – відношення середнього приросту до середнього падіння.

Цей показник можна легко порахувати засобами Excel. Спочатку беруться дані за 15 інтервалів, рахується щоденний приріст та падіння, далі береться середнє значення приросту та падіння, середнє значення приросту ділиться на падіння і отримуємо RS .

Результат відображає наступне:

- вище 70: актив перекуплений, варто продавати;
- між 30 та 70: актив торгується в нормальному діапазоні;
- нижче 30: актив вважається перепроданим, сигнал до покупки.

Результати підрахунків наведені на рис. 2.17 та рис. 2.18

День	Дата	Ціна	Приріст	Падіння		
1	24.08.2024	64173				Середній приріст 203,4379
2	25.08.2024	64273,19	100,19	0		Середнє падіння 919,2436
3	26.08.2024	62871,45	0	1401,74		RS 0,22131
4	27.08.2024	59446,47	0	3424,98		RSI 18,12071
5	28.08.2024	59058,94	0	387,53		
6	29.08.2024	59389,91	330,97	0		
7	30.08.2024	59115,12	0	274,79		
8	31.08.2024	58967,39	0	147,73		
9	01.09.2024	57321,57	0	1645,82		
10	02.09.2024	59091,35	1769,78	0		
11	03.09.2024	57514,16	0	1577,19		
12	04.09.2024	57960,93	446,77	0		
13	05.09.2024	56172,52	0	1788,41		
14	06.09.2024	53951,3	0	2221,22		
15	07.09.2024	54151,72	200,42	0		

Рисунок 2.17 – Результати підрахунку RSI Bitcoin за останні 14 днів

День	Дата	Ціна	Приріст	Падіння		
1	24.08.2024	2769,04				Середній приріст 18,36857
2	25.08.2024	2748,24	0	20,8		Середнє падіння 53,74857
3	26.08.2024	2680,53	0	67,71		RS 0,34175
4	27.08.2024	2458,39	0	222,14		RSI 25,47046
5	28.08.2024	2527,89	69,5	0		
6	29.08.2024	2528,54	0,65	0		
7	30.08.2024	2525,21	0	3,33		
8	31.08.2024	2513,58	0	11,63		
9	01.09.2024	2428,19	0	85,39		
10	02.09.2024	2537,9	109,71	0		
11	03.09.2024	2421,31	0	116,59		
12	04.09.2024	2448,58	27,27	0		
13	05.09.2024	2367,2	0	81,38		
14	06.09.2024	2223,69	0	143,51		
15	07.09.2024	2273,72	50,03	0		

Рисунок 2.18 – Результати підрахунку RSI Ethereum за останні 14 днів

Отже, обидві валюти мають низький RSI, що означає, що активи наразі є **перепроданими** і їх варто **купувати в короткостроковій** перспективі.

Аналіз волатильності

Для аналізу волатильності існує показник – ATR (Average True Range), який є дуже корисним інструментом для розуміння, коли варто входити у ринок і виходити з нього, особливо це стосується трейдерів, які займаються короткостроковим трейдингом.

Загалом, цей показник не враховує напрямку руху ціни, а сам принцип сили зміни ціни, тобто – абсолютна волатильність. Принцип розрахунку полягає у знаходженні максимального значення серед різниці найвищої та найнижчої цін за інтервал, різниці найвищої ціни і ціни під час закриття періоду, різниці найнижчої ціни і ціни під час закриття періоду. Зазвичай, розрахунок проводять за 5-10 днів, але можна підібрати і більш короткий інтервал, наприклад години. Результати підрахунку обох активів наведені на рис. 2.19 та рис. 2.20.

День	Дата	High	Low	Close	True Range	ATR
1	29.08.2024	61188,11	58784,74	59389,91		
2	30.08.2024	59906,23	57699,76	59115,12	2206,47	2206,47
3	31.08.2024	59449,34	58756,86	58967,39	692,48	1449,475
4	01.09.2024	59054,15	57172	57321,57	1882,15	1593,7
5	02.09.2024	59415,64	57120,71	59091,35	2294,93	1769,0075
6	03.09.2024	59817,86	57415,64	57514,16	2402,22	1895,65
7	04.09.2024	58525,12	55591,06	57960,93	2934,06	2068,718333
8	05.09.2024	58303,4	55654	56172,52	2649,4	2151,672857
9	06.09.2024	57000	52553,68	53951,3	4446,32	2438,50375
10	07.09.2024	54831,34	53743,65	54151,72	1087,69	2288,413333

Рисунок 2.19 – Результати розрахунку ATR Bitcoin за останні 10 днів

День	Дата	High	Low	Close	True Range	ATR
1	29.08.2024	2595,95	2506,39	2528,54		
2	30.08.2024	2550,86	2430,63	2525,21	120,23	120,23
3	31.08.2024	2530	2493,68	2513,58	36,32	78,275
4	01.09.2024	2514,57	2400	2428,19	114,57	90,37333333
5	02.09.2024	2563,22	2424,88	2537,9	138,34	102,365
6	03.09.2024	2554,04	2412,15	2421,31	141,89	110,27
7	04.09.2024	2489,26	2308,79	2448,58	180,47	121,97
8	05.09.2024	2465,8	2348,28	2367,2	117,52	121,3342857
9	06.09.2024	2406,54	2150,28	2223,69	256,26	138,2
10	07.09.2024	2310,13	2222,44	2273,72	87,69	132,5877778

Рисунок 2.20 – Результати розрахунку ATR Ethereum за останні 10 днів

Отже, згідно результатів, **обидві** валюти можна вважати приблизно **однаково** волатильними, адже показник ATR лише **зростає**, що підказує трейдерам, що ринок залишається **активним** для обидвох активів.

Аналіз кореляцій

Даний вид аналізу варто застосовувати для диверсифікації портфелю інвестора, адже якщо актив показує високу схожість до іншого активу, для кращої диверсифікації варто пошукати інший актив.

Для аналізу необхідні дані про ціну під часу закриття періоду. Для прикладу, буде проведено кореляцію активів за останні 30 днів. Результати наведені на рис. 2.21.

Дата	Ціна закриття BTC	Ціна закриття ETH	Зміна ціни BTC	Зміна ціни ETH		
07.08.2024	55152,38	2337,05				
08.08.2024	61684,87	2684,95	11,84443899	14,88628827	Кореляція	0,901774573
09.08.2024	60856,63	2600,09	-1,342695543	-3,160580272		
10.08.2024	60928,12	2611,02	0,117472821	0,420370064		
11.08.2024	58712,65	2555,19	-3,636202791	-2,138244824		
12.08.2024	59368,58	2723,23	1,117186841	6,576418975		
13.08.2024	60632,72	2703,54	2,129308129	-0,723038451		
14.08.2024	58723	2663,47	-3,149652531	-1,482130836		
15.08.2024	57547,19	2570,25	-2,002298929	-3,49994556		
16.08.2024	58887,62	2593,03	2,329270986	0,886295107		
17.08.2024	59518,27	2614,78	1,07093817	0,838787056		
18.08.2024	58440,34	2612,3	-1,811090947	-0,094845455		
19.08.2024	59478,67	2637,02	1,776735043	0,946292539		
20.08.2024	59019,79	2574	-0,771503465	-2,389818811		
21.08.2024	61154,5	2631,33	3,616939335	2,227272727		
22.08.2024	60382,2	2622,91	-1,262867001	-0,319990271		
23.08.2024	64092,32	2764	6,144393546	5,379139963		
24.08.2024	64173	2769,04	0,125880917	0,182344428		
25.08.2024	64273,19	2748,24	0,15612485	-0,751162858		
26.08.2024	62871,45	2680,53	-2,180909334	-2,463758624		
27.08.2024	59446,47	2458,39	-5,447591872	-8,28716709		
28.08.2024	59058,94	2527,89	-0,651897413	2,827053478		
29.08.2024	59389,91	2528,54	0,560406265	0,025713144		
30.08.2024	59115,12	2525,21	-0,462688022	-0,131696552		
31.08.2024	58967,39	2513,58	-0,249902225	-0,460555756		
01.09.2024	57321,57	2428,19	-2,79106808	-3,397146699		
02.09.2024	59091,35	2537,9	3,087459049	4,518180208		
03.09.2024	57514,16	2421,31	-2,669070854	-4,593955633		
04.09.2024	57960,93	2448,58	0,776800009	1,12624984		
05.09.2024	56172,52	2367,2	-3,085544004	-3,323558961		
06.09.2024	53951,3	2223,69	-3,954282272	-6,062436634		
07.09.2024	54151,72	2273,72	0,371483171	2,249863965		

Рисунок 2.21 – Аналіз кореляцій між активами Bitcoin та Ethereum

Отже, результат є дуже близьким до +1, що свідчить **про сильну позитивну кореляцію**, що означає, що загалом, разом із зростанням ціни на Bitcoin, **зростає** і ціна на Ethereum, аналогічно зі **спадом**.

Автоматичне отримання результатів аналізів

Для повного розуміння, як працює біржа та загалом коливання цін на криптовалюті дуже добре вміти аналізувати та збирати дані власноруч. Але, враховуючи сучасні зміни та потреби, існують різноманітні інструменти, які вже зібрали, акумулювали та підготували усі необхідні аналізи трендів та цін, що допоможе трейдерам зрозуміти, що робити на коротку та довгу перспективи.

Саме таким інструментом є TradingView, який має вигляд як на рис. 2.22.



Рисунок 2.22 – Інструмент аналізу криптовалют TradingView

На сайті є безліч інструментів для трейдингу: зібрано тисячі різноманітних графіків, аналізів і т.п., що оновлюються в реальному часі (див. рис. 2.23).

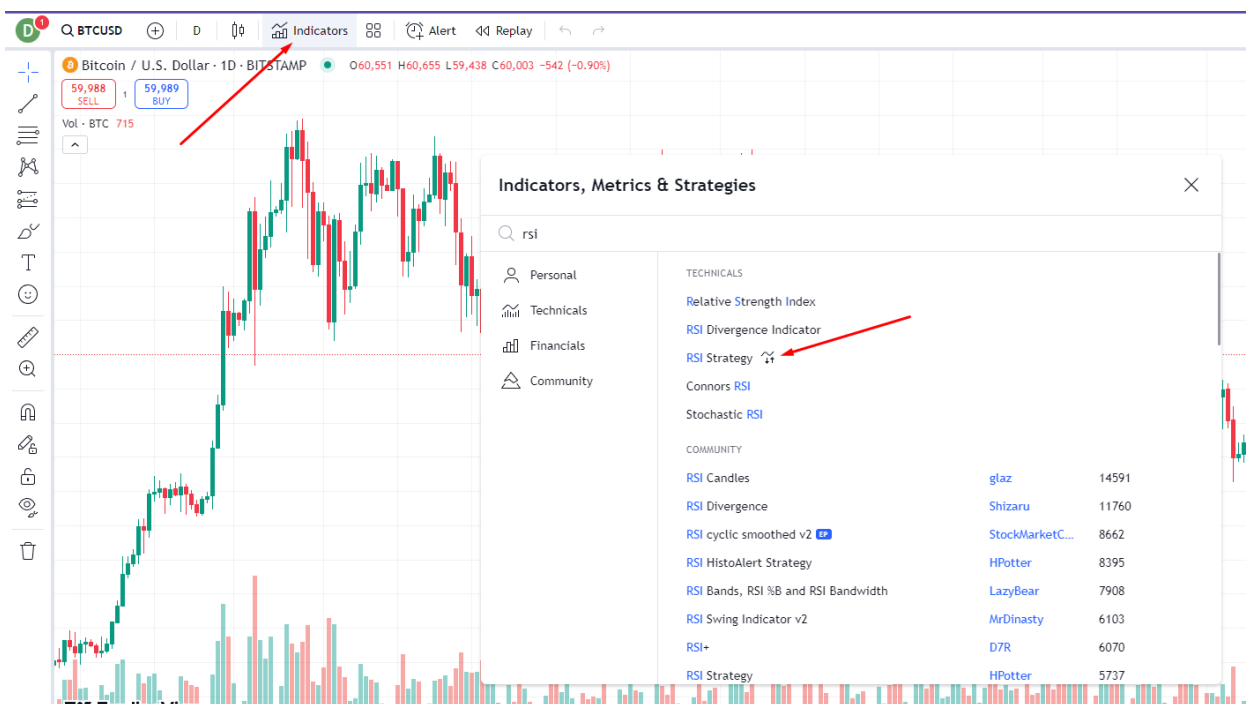


Рисунок 2.23 – Приклад пошуку RSI аналізу на сайті TradingView

Серед плюсів використання даного сервісу можна виділити **безкоштовність** (але лише умовна, за певні інструменти потребуються платна підписка), **наочність**, адже наявні графіки, які можна співставляти та порівнювати, якісна **масштабованість**, від років до секунд. Але серед мінусів – немає **ввідного «гайду»** для новачків, що значно ускладнює орієнтованість в подібній системі.

2.5 Методи оцінки ефективності моделей

Оцінка ефективності моделей машинного навчання є ключовим етапом у розробці прогнозних систем, особливо у випадку фінансових ринків, таких як ринок криптовалют. Існує багато методів для вимірювання продуктивності моделей, кожен з яких має свої переваги й обмеження. Давайте розглянемо основні підходи до оцінки ефективності моделей у прогнозуванні криптовалют.

MAE (Mean Absolute Error — Середня абсолютна помилка) є простим і зрозумілим показником, що відображає середню абсолютну різницю між прогнозованими значеннями та реальними. Він дозволяє оцінити, наскільки далеко

в середньому прогнози моделі відхиляються від фактичних значень. MAE (див. рис. 2.26) легко інтерпретується, оскільки виражається в тих самих одиницях, що й дані.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (2.3)$$

де y_i – фактичне значення;

\hat{y}_i – прогнозоване значення;

n – кількість точок даних.

Тобто, якщо MAE для моделі прогнозування цін криптовалют становить 50, це означає, що середня абсолютна похибка між реальними та прогнозованими значеннями ціни складає 50 доларів.

MAPE (Mean Absolute Percentage Error — Середня абсолютна відносна помилка) виражає помилку прогнозування у відсотках і є зручним показником, оскільки дозволяє оцінити відносну величину помилки незалежно від одиниць вимірювання. Однак MAPE може бути менш точним, коли реальні значення близькі до нуля.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| * 100, \quad (2.4)$$

де y_i – фактичне значення;

\hat{y}_i – прогнозоване значення;

n – кількість точок даних.

Тобто, MAPE в 5% означає, що середня відносна помилка моделі становить 5% від реальних значень.

RMSE (Root Mean Squared Error — Корінь середньоквадратичної помилки) також є поширеним показником для оцінки ефективності моделі, який підкреслює великі похибки через квадратичну форму. RMSE (див. рис. 2.28) чутливіший до великих відхилень між прогнозом і реальністю, тому він краще підходить, коли важливо мінімізувати великі помилки.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (2.5)$$

R^2 (Коефіцієнт детермінації) — це метрика, що відображає, наскільки добре модель пояснює варіацію залежної змінної. Він приймає значення від 0 до 1, де значення, близьке до 1, означає, що модель дуже точно описує залежність між прогнозованими і реальними значеннями (див. рис. 2.29).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (2.6)$$

де y_i — фактичне значення;

\hat{y}_i — прогнозоване значення;

n — кількість точок даних;

\bar{y} — середнє значення реальних даних.

Якщо $R^2 = 0.99$, це означає, що модель пояснює 99% варіації залежної змінної, що вказує на високу точність моделі.

Висновок до розділу 2

У другому розділі було детально розглянуто інформаційні технології та методи, які використовуються для прогнозування цін на криптовалюту. На основі проведеного огляду методів прогнозування встановлено, що рекурентні нейронні мережі (зокрема LSTM) є одними з найефективніших інструментів для роботи з часовими рядами завдяки їх здатності враховувати довгострокові залежності в даних.

Розглянуто сучасні підходи до збору та аналізу даних, які включають використання API, парсингу новинних ресурсів та агрегаторів даних. Було описано процес отримання даних, що включає вибір релевантних джерел, фільтрацію й попередню обробку для усунення пропусків та забезпечення високої якості підготовлених даних.

Аналіз отриманих даних показав, що ключовими факторами, які впливають на прогнозування, є торгові обсяги, новинні індекси, хешрейт, макроекономічні показники та інші параметри. Було зазначено, що дані потребують нормалізації та оптимізації для покращення якості навчання моделі. Оптимально, нормалізація потребує складного та неоднозначного підходу, адже дані мають різну природу та здатність до різких стрибків, що не піддається лінійній інтерполяції.

Методи оцінки ефективності моделей прогнозування дозволили визначити основні метрики, зокрема середньоквадратичну похибку (MSE), середню абсолютну похибку (MAE), корінь середньоквадратичної помилки (RMSE) та коефіцієнт детермінації (R^2), які будуть використані для порівняння та вибору оптимальної моделі.

3 ПРОЄКТУВАННЯ СИСТЕМИ. ПОБУДОВА АРХІТЕКТУРИ НЕЙРОННОЇ МЕРЕЖІ. ЗБІР ТА ОБРОБКА ДАНИХ ДЛЯ НАВЧАННЯ

3.1 Вибір архітектури нейронної мережі та її побудова

Процес вибору архітектури нейронної мережі для розв'язання задачі прогнозування цін криптовалют базувався на аналізі специфіки вихідних даних і обраної методології. Основними вимогами до архітектури були здатність моделі враховувати послідовність часових рядів, адаптація до багатовимірних вхідних даних і мінімізація помилок прогнозування.

Зважаючи на особливості часових рядів, було вирішено використовувати рекурентні нейронні мережі (RNN), а саме їх модифікацію — блоки Gated Recurrent Unit (GRU). GRU є оптимальним вибором через меншу складність у порівнянні з LSTM (Long Short-Term Memory) і здатність ефективно запам'ятовувати довгострокові залежності. Архітектура GRU забезпечує обробку часових послідовностей завдяки наявності вентилів оновлення та обнулення, які дозволяють контролювати потік інформації.

а) Основні етапи побудови моделі:

Попередня обробка даних: Вхідні дані склалися з історичних цін на криптовалюту, обсягів торгів, хешрейту, макроекономічних показників. Для забезпечення коректного навчання модель отримала масштабовані значення вхідних даних. Для цього застосовувався StandardScaler, який стандартизував вхідні дані, щоб уникнути проблем через різні масштаби параметрів. Дані були перетворені у формат, сумісний із LSTM/GRU, тобто у тривимірний масив розмірності (кількість зразків, кількість кроків у часі, кількість ознак).

б) Архітектура моделі: Модель містить такі шари:

1) LSTM/GRU-шар: Основний шар із 100 нейронів, з активаційною функцією tan. Dropout у розмірі 0.2 застосовується для випадкового обнулення деяких нейронів під час навчання, що також сприяє узагальненню моделі.

2) Вихідний шар (Dense): Останній шар, що генерує один вихідний параметр – прогнозоване значення ціни.

в) Функція втрат: Для навчання моделі обрано функцію втрат Mean Absolute Percentage Error (MAPE), яка зручно інтерпретується у вигляді середньої відносної помилки.

г) Оптимізатор: Для мінімізації функції втрат використовувався оптимізатор Adam із початковим коефіцієнтом навчання 0.001. Adam є адаптивним методом, який забезпечує швидшу збіжність навіть для великих і складних моделей.

д) Навчання моделі: Модель навчалася протягом 50 епох із розміром батчу 32. Експериментально застосовувались різні кількості епох (від 10 (що показало низьку ефективність через високу складність задачі) до 200 з застосуванням ранньої зупинки у випадку зупинки покращення моделі (що показало не дуже гарну ефективність внаслідок перенавчання моделі)), різні кількості батчу (але загалом це не дуже впливало на ефективність).

е) Оцінка якості моделі: Для оцінки точності передбачень використовували метрики:

- 1) Середня абсолютна помилка (MAE);
- 2) Квадратична середня помилка (RMSE);
- 3) Коефіцієнт детермінації (R^2).

ж) Тестування: На тестовому наборі модель демонструвала невеликі значення помилки, що вказує на її ефективність для прогнозування цін криптовалют. У деяких випадках, залежно від конфігурації даних та архітектури, було отримано значення метрик, які наближаються до нуля, що свідчить про високу точність передбачень.

з) Можливі вдосконалення: Для підвищення точності прогнозування планується додати нові параметри, зокрема ціни інших криптовалют, які можуть корелювати з обраною криптовалютою. Також можливим є експериментування з іншими архітектурами RNN або комбінація GRU з механізмом уваги.

Отже, для вирішення поставленої задачі було обрано дві архітектури: LSTM та GRU з доволі схожими вхідними шарами та методами навчання. Візуально вони виглядають як на рис. 3.1 та рис. 3.2.

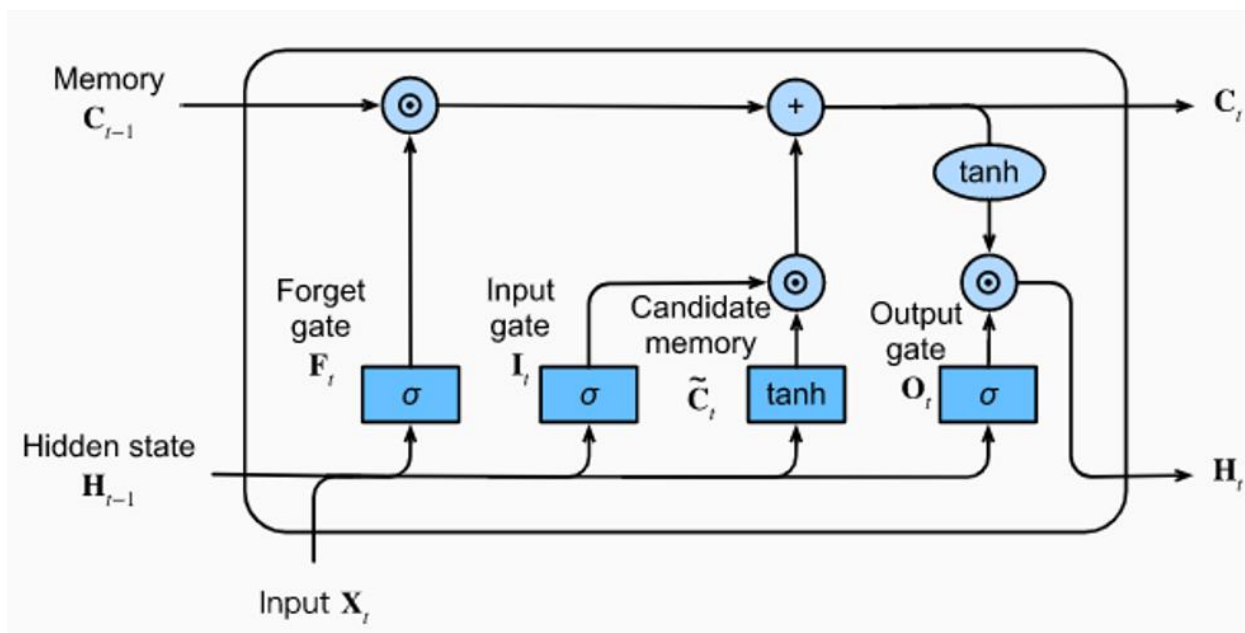


Рисунок 3.1 – Архітектура LSTM

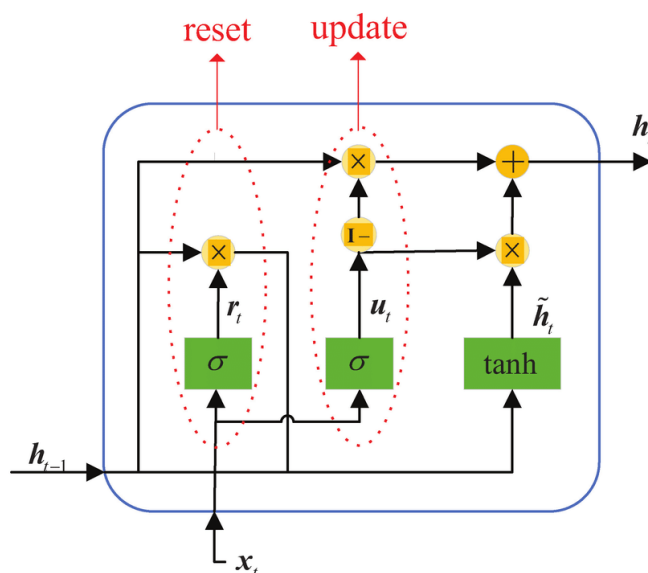


Рисунок 3.2 – Архітектура GRU

Архітектури LSTM (Long Short-Term Memory) та GRU (Gated Recurrent Unit) є оптимальними виборами для вирішення задачі прогнозування цін криптовалют

через їх здатність ефективно обробляти послідовні дані. У задачах прогнозування важливо враховувати часові залежності між значеннями параметрів, а рекурентні нейронні мережі (RNN) з архітектурами LSTM та GRU спеціалізуються на цьому завданні завдяки вбудованим механізмам роботи з довготривалою пам'яттю. Нижче детально розглянуто обґрунтування вибору цих архітектур та нюанси налаштування навчання.

Такі архітектури було обрано, бо звичайні RNN не можуть ефективно обробляти довгі часові ряди через проблему затухаючого градієнта. LSTM та GRU вирішують цю проблему завдяки спеціальним воротам, які дозволяють зберігати або забувати інформацію.

У фінансових часових рядах, таких як ціни криптовалют, часто спостерігаються складні залежності, що включають як короткотермінові, так і довготермінові тренди. LSTM та GRU добре підходять для ідентифікації таких залежностей.

У порівнянні з LSTM, GRU має простішу структуру (менше параметрів), що дозволяє швидше навчатися та зменшує ризик перенавчання на невеликих наборах даних. У нашій задачі це може бути корисним для швидкого експериментального тестування.

Щодо кількості епох навчання (від 10 до 200) була обрана з урахуванням експериментальних результатів. У задачах прогнозування криптовалют важливо враховувати складність функції втрат. Навчання протягом 50-100 епох дозволяє досягти стабільної збіжності без перенавчання. Також використання ранньої зупинки з моніторингом метрики на валідаційному наборі запобігає перенавчанню. Зазвичай, у складних моделях, як GRU/LSTM, оптимум досягається у перших 50-100 епохах.

Розмір батчу 32 є типовим вибором, що забезпечує баланс між швидкою обробкою даних і точністю оновлення ваг мережі. Для великих обсягів даних розмір батчу 32 дозволяє ефективно використовувати пам'ять GPU, бо малий батч

(<32) може бути нестабільним через високий рівень шуму, а великий батч (>64) може призводити до надто малих оновлень, що уповільнює навчання.

Було додано вихідний шар Dropout у розмірі 0.2 для випадкового обнулення частини нейронів під час навчання. Це сприяє узагальненню та покращує точність моделі на тестових даних.

Для покращення ефективності роботи моделі було додано оптимізатор Adam, який є адаптивним методом, який коригує коефіцієнт навчання для кожного параметра окремо. Це забезпечує швидшу збіжність у порівнянні зі стандартним SGD (Stochastic Gradient Descent).

Значення коефіцієнту навчання 0.001 є стандартним вибором для більшості задач, зокрема, для складних моделей, таких як GRU/LSTM. Це дозволяє швидко зближуватись до мінімуму функції втрат без ризику "перескоку" оптимального значення.

Також для оцінки помилок моделі було обрано середню абсолютну помилку (MAE) та квадратичну середню помилку (RMSE), що забезпечують різні перспективи оцінки якості прогнозу: MAE акцентує на абсолютних відхиленнях, а RMSE більше чутливий до великих відхилень. У той же час R^2 дозволяє оцінити, наскільки модель пояснює варіативність даних. У поставленій задачі високий R^2 свідчить про ефективність моделі у відображенні трендів.

3.2 Збір, обробка та підготовка даних до навчання нейронної моделі

Згідно поставленого завдання, для навчання нейронної мережі необхідно було зібрати дані про ціни криптоактивів та макроекономічні показники. Спочатку, було прийнято рішення про збір історичних добових даних та потім їх загальне об'єднання (merge) за датою. В розділі 2 було описано, як вже було отримано дані про ціни криптоактивів. Тому наступним кроком було отримання макроекономічних показників. Шляхом пошуку на спеціальних сервісах (наприклад Kaggle) не було знайдено схожих датасетів, отже збір даних проводився власноруч.

Почався пошук з даних про хешрейт (дані про необхідну хешову швидкість для знаходження блоку за 1 годину). Ці дані можуть корелювати з цінами на криптоактив, адже показують залученість майнерів до цього активу – чим більше необхідно розв'язувати хеші, тим важче здобувається блок і відповідно тим дорожче він стає.

Дані було знайдено за допомогою сервісу BitcoinVisuals. Дані було подано графіком із можливістю скачувати їх у вигляді CSV-файлу. Дані було зібрано за період 2008-2024 рік (див. рис. 3.3).

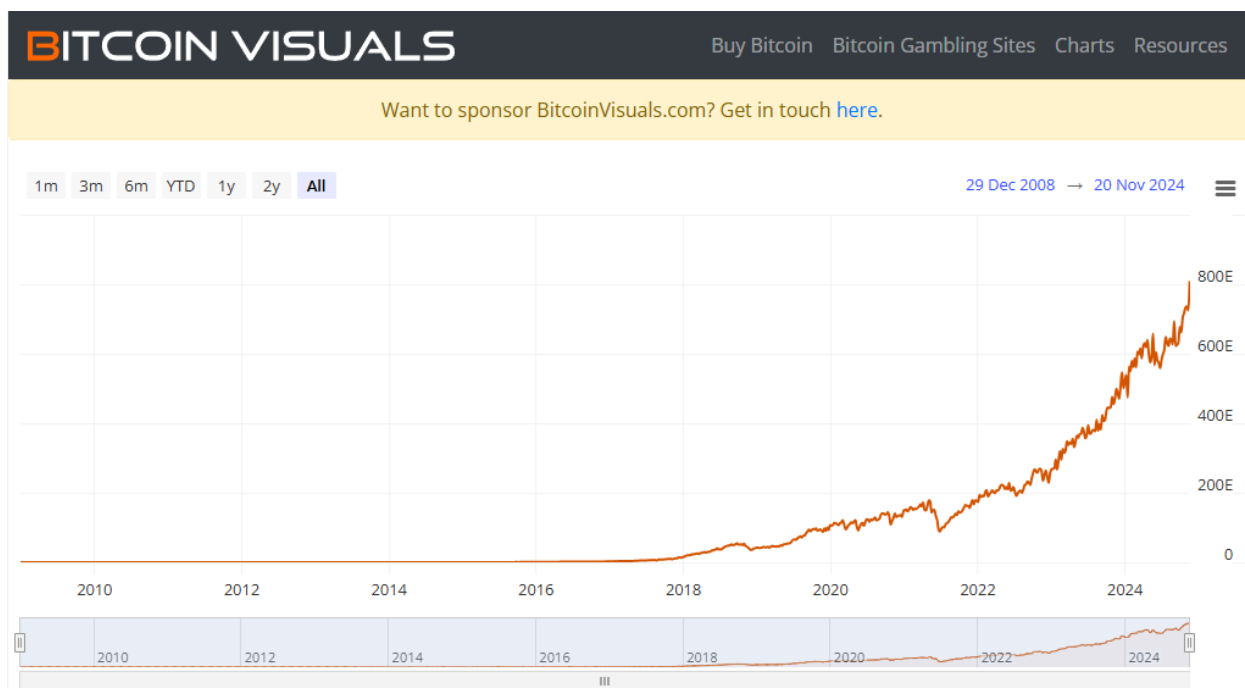


Рисунок 3.3 – Графік хешрейту Bitcoin

Наступним показником стала EFFR (Effective Federal Funds Rate, ефективна ставка федеральних фондів). Зміни цього показника прямо впливають на стан економіки США, стабільність та інфляцію Долара США і відповідно на настрої інвесторів, їхній рівень ризику та залученості в інвестиції.

Дані такого типу є важкодоступними, а їхнє масове збирання за допомогою API є платним або сильно обмеженим. Найкращим варіантом в такому випадку було створення власного парсеру для сайту зі статичною версткою (або, принаймні, статичним відображенням необхідних даних).

Сайт Федерального банку Нью-Йорку (див. рис. 3.5) мав відкритий доступ до цих даних і мав статичне відображення, які були зібрані за допомогою веб-скрапера (дод. А). Дані мали наступний вигляд (див. рис. 3.4):

date,effr,Interpolated effr	
2019-09-14,,2.25	
2019-09-15,,2.25	
2019-09-16,2.25,2.25	
2019-09-17,2.3,2.3	
2019-09-18,2.25,2.25	
2019-09-19,1.9,1.9	
2019-09-20,1.9,1.9	
2019-09-21,,1.9	
2019-09-22,,1.9	
2019-09-23,1.9,1.9	
2019-09-24,1.9,1.9	
2019-09-25,1.9,1.9	
2019-09-26,1.85,1.85	
2019-09-27,1.83,1.83	

Рисунок 3.4 – Добові дані EFR

FEDERAL FUNDS DATA

[Chart](#) | [Historical Search](#)

DATE	RATE (%)	1ST PERCENTILE (%)	25TH PERCENTILE (%)	75TH PERCENTILE (%)	99TH PERCENTILE (%)	VOLUME (\$Billions)	TARGET RATE/RANGE (%)
12/06	4.58	4.56	4.57	4.59	4.60	107	4.50 - 4.75
12/05	4.58	4.56	4.57	4.59	4.60	109	4.50 - 4.75
12/04	4.58	4.56	4.58	4.59	4.60	99	4.50 - 4.75
12/03	4.58	4.56	4.57	4.59	4.60	94	4.50 - 4.75
12/02	4.58	4.56	4.58	4.59	4.64	92	4.50 - 4.75
11/29	4.58	4.55	4.58	4.59	4.60	95	4.50 - 4.75
11/27	4.58	4.56	4.57	4.59	4.60	103	4.50 - 4.75
11/26	4.58	4.56	4.57	4.59	4.60	101	4.50 - 4.75
11/25	4.58	4.56	4.58	4.59	4.60	103	4.50 - 4.75
11/22	4.58	4.56	4.57	4.59	4.70	100	4.50 - 4.75
11/21	4.58	4.56	4.57	4.59	4.60	101	4.50 - 4.75
11/20	4.58	4.56	4.57	4.59	4.60	102	4.50 - 4.75
11/19	4.58	4.56	4.57	4.59	4.60	99	4.50 - 4.75
11/18	4.58	4.56	4.57	4.59	4.60	103	4.50 - 4.75
11/15	4.58	4.56	4.57	4.59	4.65	104	4.50 - 4.75
11/14	4.58	4.56	4.57	4.59	4.64	107	4.50 - 4.75
11/13	4.58	4.56	4.57	4.59	4.65	105	4.50 - 4.75
11/12	4.58	4.56	4.57	4.59	4.60	101	4.50 - 4.75
11/08	4.58	4.56	4.57	4.59	4.60	103	4.50 - 4.75
11/07	4.83	4.81	4.82	4.84	4.90	100	4.75 - 5.00
11/06	4.83	4.81	4.82	4.84	4.90	102	4.75 - 5.00
11/05	4.83	4.81	4.82	4.84	4.90	104	4.75 - 5.00
11/04	4.83	4.81	4.83	4.84	4.90	98	4.75 - 5.00

Рисунок 3.5 – Добові дані EFFR на сайті Федерального банку Нью-Йорку США

Так як дані за деякі дні були відсутні, було застосовано лінійну інтерполяцію для заповнення пропущених даних, адже це важливо для збереження часових рядів.

Наступним кроком була спроба зібрати незвичайні та неоднозначні для сприйняття моделлю дані – новини про криптовалюту та їхній настрій. Адже існує купа історичних прикладів, як певні новини дуже сильно впливали на ціну Bitcoin. Наприклад, в травні 2022 року ціна біткоіна дуже сильно зросла після звичайного твіту Ілона Маска про відношення компанії Tesla до Bitcoin [44], або як перемога Трампа на Президентських виборах у США змусила Bitcoin «взлетіти» [45] в ціні на більше ніж пару десятків тисяч доларів США за 1 монету. Отже, тут видно, як

новини та їхній настрої прямо корелюють з ціною на криптоактиви і можуть сильно впливати на них.

Для цієї задачі було застосовано інший веб-скрапер (дод. Б), який одночасно збирав усі новини (див. рис. 3.6) за певними HTML-атрибутами.

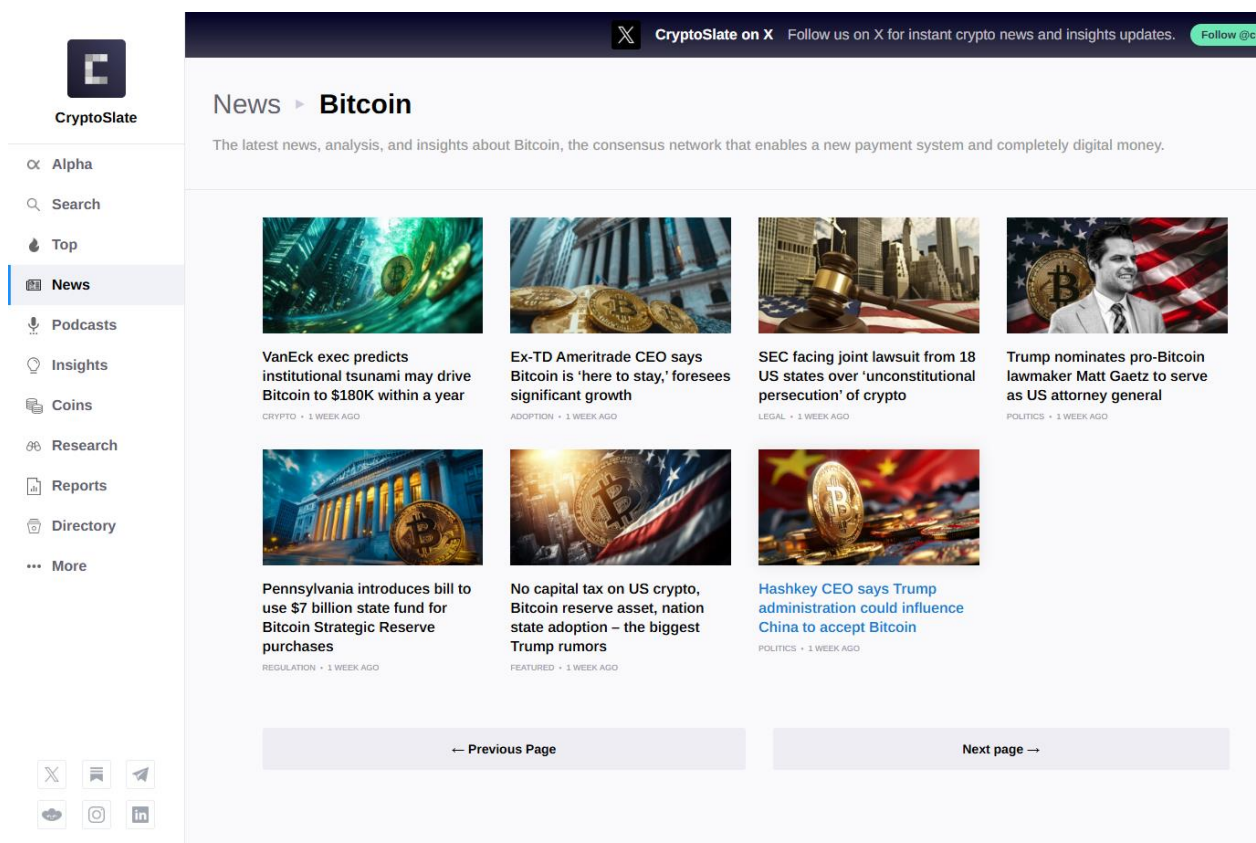


Рисунок 3.6 – Сайт для збору новин, пов'язані з криптовалютами

У той же час за допомогою бібліотеки SentimentIntensityAnalyzer було проведено оцінку настроїв новин. Такі дані було представлено у числовому вигляді в діапазоні від 0 до 1 (див. рис. 3.7). Два показники – негативна складова та позитивна складова. Дані були представлені наступним чином:

title	positive_sentiment	negative_sentiment	date	link
How DePIN is disrupting Google, Amazon, and Nvidia	0.195	0.0	2024-10-22	https://cryptoslate.com/how-d
Northern Data embraces AI boom, mulls ditching Bitcoin mining roots	0.0	0.0	2024-10-21	https://cryptosla
US election hopes spark three-month high inflow of over \$2 billion in crypto investments	0.281	0.0	2024-1	
Global regulators discussing ways to 'eliminate' Bitcoin highlights cracks in fiat system	0.0	0.0	2024-10-21	https://cryptosla
US Department of Defense places Bitcoin Softwar thesis under security review, rockets price to \$300	0.231			
Post halving, Bitcoin miners are choosing between hodling BTC and upgrading to AI	0.0	0.0	2024-10-19	https://cryptosla
Institutional demand and rising ETP flows signal Bitcoin breakout – VanEck	0.0	0.13	2024-10-19	https://cry
Analyzing Bitcoin's supply trends as surge in long-term Bitcoin holdings points to investor confidence	0.20			
Crypto liquidation breakdown – 77,548 traders hit by \$307M liquidation, Bitcoin and Ethereum lead losses				
Nasdaq files proposal seeking SEC rule change to allow spot Ethereum ETF	0.147	0.0	2023-11-14	https://cry
Coinbase expects spot Bitcoin ETFs to unlock new era of crypto investment options	0.0	0.0	2023-11-14	https://cry
Cboe Digital to offer margin futures trading for Bitcoin and Ethereum in 2024	0.0	0.0	2023-11-13	https://cry
Google parent company Alphabet dumps last of its Robinhood shares	0.171	0.209	2023-11-13	https://crypt
Bitcoin fund holdings hit record high as ETF anticipation stirs optimism	0.353	0.0	2023-11-13	https://crypto
Are you exposed? How Chainalysis cracked the Wasabi Bitcoin privacy wallet	0.0	0.115	2022-03-15	https://
Bitcoin's global carbon footprint accounts for 0.19%, is green energy the solution?	0.306	0.0	2022-03-15	https://
EU votes against a de-facto ban on Bitcoin's POW consensus method in MiCA draft	0.0	0.217	2022-03-14	https://

Рисунок 3.7 – Дані про настрої криптовалютних новин

Отже, в CSV файл було завантажено заголовок, позитивну та негативну оцінку, дату публікації новини та посилання на неї.

Наступними даними для завантаження були ціни на різні активи, які мають вплив (або залежать від впливу зміни ціни чи стабільності долара, що можна теоретично віднести до кореляції із зміною ціни криптоактиву) на зміну ціни криптоактиву. Тобто, це ціни на золото, ціна активу S&P 500, NASDAQ 100 та індекс долара США (DXY).

Для виконання цієї задачі було написано веб-скрапер (дод. В). На вхід скрипт приймав HTML-структуру таблиці з історичними даними (див. рис. 3.8).

Кафедра інтелектуальних інформаційних систем
Інтелектуальна система прогнозування цін криптовалют на основі рекурентних нейронних мереж

Date	Price	Open	High	Low	Vol	Change %
Nov 22, 2024	5,962.56	5,947.80	5,972.01	5,944.36	0.00K	+0.23%
Nov 21, 2024	5,948.71	5,940.58	5,963.32	5,887.26	0.00K	+0.53%
Nov 20, 2024	5,917.11	5,914.34	5,920.67	5,860.56	0.00K	0.00%
Nov 19, 2024	5,916.98	5,870.05	5,923.51	5,855.29	0.00K	+0.40%
Nov 18, 2024	5,893.62	5,874.17	5,908.12	5,865.95	0.00K	+0.39%
Nov 15, 2024	5,870.62	5,912.79	5,915.32	5,853.01		-1.32%
Nov 14, 2024	5,949.17	5,989.68	5,993.88	5,942.28		-0.60%
Nov 13, 2024	5,985.38	5,985.75	6,008.19	5,965.91		+0.02%
Nov 12, 2024	5,983.99	6,003.60	6,009.92	5,960.08		-0.29%
Nov 11, 2024	6,001.35	6,008.86	6,017.31	5,986.69		+0.10%
Nov 08, 2024	5,995.54	5,976.76	6,012.45	5,976.76		+0.38%
Nov 07, 2024	5,973.10	5,947.21	5,983.84	5,947.21		+0.74%
Nov 06, 2024	5,929.04	5,864.89	5,936.14	5,864.89		+2.53%
Nov 05, 2024	5,782.76	5,722.43	5,783.44	5,722.10		+1.23%
Nov 04, 2024	5,712.69	5,725.15	5,741.43	5,696.51		-0.28%

Рисунок 3.8 – Представлення даних про макроекономічні показники

Отже, засобами DevTools, що знаходяться в браузері Google Chrome, можна було знайти структуру таблиці і вивантажити у вигляді HTML (див. рис. 3.9).

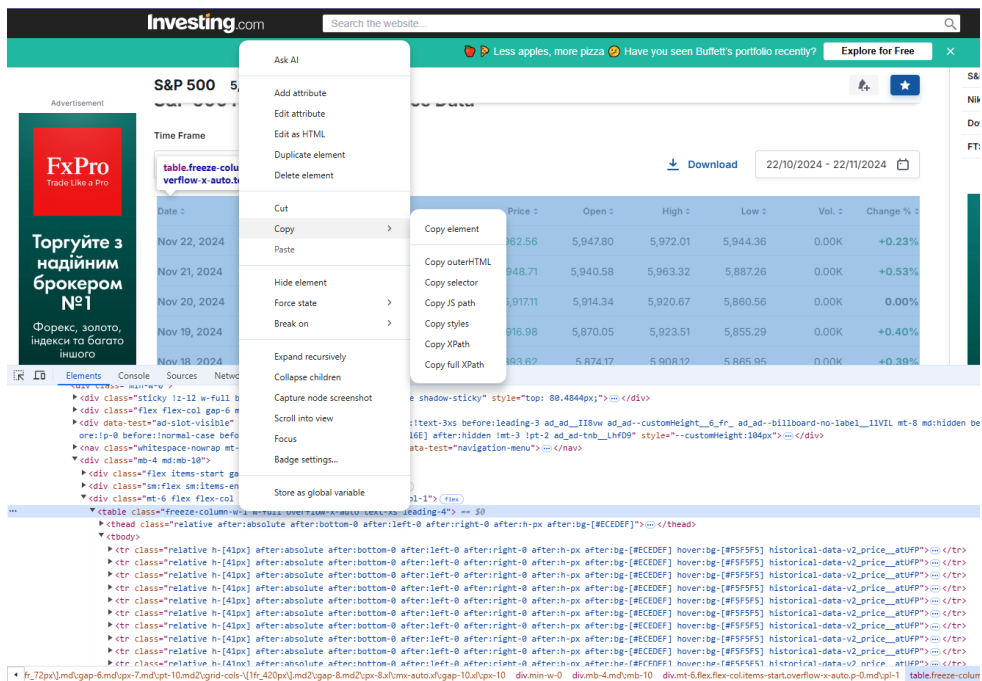


Рисунок 3.9 – Вивантаження таблиці з даними у вигляді HTML

Дані були вивантажені у CSV-файл (див. рис. 3.10).

date,price
2008-01-01,
2008-01-02,2049.71
2008-01-03,2051.76
2008-01-04,1963.52
2008-01-05,
2008-01-06,
2008-01-07,1957.44
2008-01-08,1910.33
2008-01-09,1949.2
2008-01-10,1953.64
2008-01-11,1912.81
2008-01-12,
2008-01-13,
2008-01-14,1949.15
2008-01-15,1894.09
2008-01-16,1872.29
2008-01-17,1842.1
2008-01-18,1844.09

Рисунок 3.10 – Дані про макроекономічні показники до інтерполяції

Так як дані містили пропущені дні, було застосовано лінійну інтерполяцію для заповнення пропущених значень. Дані мають вигляд як на рис. 3.11.

date,price,Interpolated price
2019-09-14,,1497.9
2019-09-15,,1497.9
2019-09-16,1497.9,1497.9
2019-09-17,1500.85,1500.85
2019-09-18,1491.2,1491.2
2019-09-19,1499.8,1499.8
2019-09-20,1517.1,1517.1
2019-09-21,,1518.3766666666666
2019-09-22,,1519.6533333333334
2019-09-23,1520.93,1520.93
2019-09-24,1531.8,1531.8
2019-09-25,1506.52,1506.52
2019-09-26,1506.85,1506.85
2019-09-27,1496.93,1496.93
2019-09-28,,1489.1933333333334
2019-09-29,,1481.4566666666667
2019-09-30,1473.72,1473.72
2019-10-01,1481.82,1481.82
2019-10-02,1500.11,1500.11
2019-10-03,1508.4,1508.4

Рисунок 3.11 – Дані про макроекономічні показники після інтерполяції

Також було отримано дані про інфляцію долара США. Але проблема постала в неможливості використовувати такі дані через те, що вони можуть бути в місячному або річному інтервалі, а потрібно використовувати погодинні дані (див. рис. 3.12).

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2024	3.1	3.2	3.5	3.4	3.3	3.0	2.9	2.5	2.4	2.6		
2023	6.4	6.0	5.0	4.9	4.0	3.0	3.2	3.7	3.7	3.2	3.1	3.4
2022	7.5	7.9	8.5	8.3	8.6	9.1	8.5	8.3	8.2	7.7	7.1	6.5
2021	1.4	1.7	2.6	4.2	5.0	5.4	5.4	5.3	5.4	6.2	6.8	7.0
2020	2.5	2.3	1.5	0.3	0.1	0.6	1.0	1.3	1.4	1.2	1.2	1.4
2019	1.6	1.5	1.9	2.0	1.8	1.6	1.8	1.7	1.7	1.8	2.1	2.3
2018	2.1	2.2	2.4	2.5	2.8	2.9	2.9	2.7	2.3	2.5	2.2	1.9
2017	2.5	2.7	2.4	2.2	1.9	1.6	1.7	1.9	2.2	2.0	2.2	2.1
2016	1.4	1.0	0.9	1.1	1.0	1.0	0.8	1.1	1.5	1.6	1.7	2.1
2015	-0.1	0.0	-0.1	-0.2	0.0	0.1	0.2	0.2	0.0	0.2	0.5	0.7
2014	1.6	1.1	1.5	2.0	2.1	2.1	2.0	1.7	1.7	1.7	1.3	0.8
2013	1.6	2.0	1.5	1.1	1.4	1.8	2.0	1.5	1.2	1.0	1.2	1.5
2012	2.9	2.9	2.7	2.3	1.7	1.7	1.4	1.7	2.0	2.2	1.8	1.7
2011	1.6	2.1	2.7	3.2	3.6	3.6	3.6	3.8	3.9	3.5	3.4	3.0
2010	2.6	2.1	2.3	2.2	2.0	1.1	1.2	1.1	1.1	1.2	1.1	1.5
2009	0.0	0.2	-0.4	-0.7	-1.3	-1.4	-2.1	-1.5	-1.3	-0.2	1.8	2.7
2008	4.3	4.0	4.0	3.9	4.2	5.0	5.6	5.4	4.9	3.7	1.1	0.1
2007	2.1	2.4	2.8	2.6	2.7	2.7	2.4	2.0	2.8	3.5	4.3	4.1
2006	4.0	3.6	3.4	3.5	4.2	4.3	4.1	3.8	2.1	1.3	2.0	2.5
2005	3.0	3.0	3.1	3.5	2.8	2.5	3.2	3.6	4.7	4.3	3.5	3.4
2004	1.9	1.7	1.7	2.3	3.1	3.3	3.0	2.7	2.5	3.2	3.5	3.3
2003	2.6	3.0	3.0	2.2	2.1	2.1	2.2	2.3	2.0	1.8	1.9	
2002	1.1	1.1	1.5	1.6	1.2	1.1	1.5	1.8	1.5	2.0	2.2	2.4
2001	3.7	3.5	2.9	3.3	3.6	3.2	2.7	2.7	2.6	2.1	1.9	1.6
2000	2.7	3.2	3.8	3.1	3.2	3.7	3.7	3.4	3.5	3.4	3.4	3.4

Рисунок 3.12 – Дані про інфляцію долара США

Після збору необхідних даних, в ході аналізу виникла проблема у вигляді нестачі готових зразків для навчання, адже загалом їхня кількість становила приблизно 2 тис., що було замало для навчання моделі такої складності. Отже, процес збору даних почався знову, але використовувались цього разу вже ширший діапазон даних (спочатку були дані за останні 6 років), який розширили на ще 4 роки (2015-2024), а також цього разу використовувались погодинні дані, що дуже

збільшувало кількість даних до 80 тис. зразків, що повинно бути достатнім для навчання.

Якщо зі збором погодинних даних про ціни криптоактиву не виникло проблем, то історичних даних про макроекономічні показники просто були недоступні у відкритих джерелах. Отже, для отримання погодинних даних було використано масштабування і кубічна (для показників хешрейту, адже вони мали властивість різко змінюватись) або лінійна інтерполяція (для послідовних і пологих даних, які не схильні до різких змін, тобто макроекономічні показники в даному випадку). Для цього було написано окремий скрипт (дод. Г), що допоміг підготувати погодинні дані (див. рис. 3.12).

24.09.2018 8:00	1,9351
24.09.2018 9:00	1,9357
24.09.2018 10:00	1,9362
24.09.2018 11:00	1,9366
24.09.2018 12:00	1,937
24.09.2018 13:00	1,9373
24.09.2018 14:00	1,9374
24.09.2018 15:00	1,9374
24.09.2018 16:00	1,9373
24.09.2018 17:00	1,9371
24.09.2018 18:00	1,9366
24.09.2018 19:00	1,9361
24.09.2018 20:00	1,9353
24.09.2018 21:00	1,9343
24.09.2018 22:00	1,9331
24.09.2018 23:00	1,9317
25.09.2018 0:00	1,93
25.09.2018 1:00	1,9281
25.09.2018 2:00	1,926
25.09.2018 3:00	1,9237

Рисунок 3.12 – Підготовлені погодинні дані

Наступним етапом було об'єднання даних в один файл для підготовки моделі до навчання. Після об'єднання дані мали вигляд як на рис. 3.12.

Кафедра інтелектуальних інформаційних систем
Інтелектуальна система прогнозування цін криптовалют на основі рекурентних нейронних мереж

datetime	open	high	low	close	Volume B	Volume U	target_ab	BitcoinVis	newyorkfi	uk_value	uk_value	uk_value	usagold_v
08.10.2015 13:00	0	245	0	245	0,6067	148,6303	0	0	0,13	94,9409	4363,819	2014,251	1154,077
08.10.2015 14:00	245	245	244,5	245	4,4536	1091,144	0,08	0	0,13	94,9228	4364,558	2014,314	1154,378
08.10.2015 15:00	245	245	244,92	244,92	3,0169	738,9055	0,67	0	0,13	94,9062	4365,259	2014,376	1154,652
08.10.2015 16:00	244,92	244,92	244,25	244,25	3,8953	951,4154	0,74	0	0,13	94,8908	4365,922	2014,438	1154,901
08.10.2015 17:00	244,25	244,99	244,02	244,99	3,9206	960,5156	0,99	0	0,13	94,8768	4366,548	2014,499	1155,126
08.10.2015 18:00	244,99	244,99	244	244	3,6905	900,4752	0	0	0,13	94,8639	4367,139	2014,559	1155,328
08.10.2015 19:00	244	244	244	244	1,5315	373,6752	0	0	0,13	94,8523	4367,697	2014,618	1155,508
08.10.2015 20:00	244	244	243,6	244	0,177	43,2	0,05	0	0,13	94,8417	4368,221	2014,676	1155,666
08.10.2015 21:00	244	244	243,95	243,95	0,0413	10,0806	0	0	0,13	94,8323	4368,713	2014,734	1155,805
08.10.2015 22:00	243,95	244	243,8	243,95	4,713	1149,744	0	0	0,13	94,8239	4369,174	2014,79	1155,924
08.10.2015 23:00	243,95	243,95	243,72	243,95	5,7083	1392,53	0,35	0	0,13	94,8165	4369,606	2014,846	1156,026
09.10.2015 0:00	243,95	243,95	243,6	243,6	3	730,8	0	0	0,13	94,81	4370,01	2014,9	1156,11
09.10.2015 1:00	243,6	243,6	243,6	243,6	0	0	0	0,0055	0,13	94,8044	4370,386	2014,953	1156,178
09.10.2015 2:00	243,6	243,6	243,6	243,6	0	0	0	0,0112	0,13	94,7997	4370,737	2015,006	1156,232
09.10.2015 3:00	243,6	243,6	243,6	243,6	0	0	0	0,0171	0,13	94,7958	4371,062	2015,057	1156,271
09.10.2015 4:00	243,6	243,6	243,6	243,6	0	0	0	0,0233	0,13	94,7927	4371,364	2015,106	1156,297
09.10.2015 5:00	243,6	243,6	243,6	243,6	0	0	0	0,0296	0,13	94,7903	4371,643	2015,155	1156,311
09.10.2015 6:00	243,6	243,6	243,6	243,6	0	0	0	0,0361	0,13	94,7886	4371,901	2015,202	1156,314

Рисунок 3.12 – Об'єднані дані

Серед цих параметрів з'явився новий – target_abs, що дорівнював модулю різниці ціни закриття між поточним і наступним зразком. Цей параметр виступав Y-характеристикою, у той час як всі інші параметри відносились до X-характеристики.

Дані було поділено на 3 датасети: тренувальний (70%), валідаційний (15%) та тестувальний (15%). Тренувальні дані містили більш старі дані, у той час як тестувальний містив найновіші.

Спочатку була спроба навчання моделі на основі архітектури LSTM. Для початку використовувався лише один шар із 50 нейронів, з оптимізатором Adam. Але така структура виявилась занадто простою для вирішення подібної задачі, адже за показниками втрат (loss) модель показувала незадовільні результати (див. рис. 3.13).

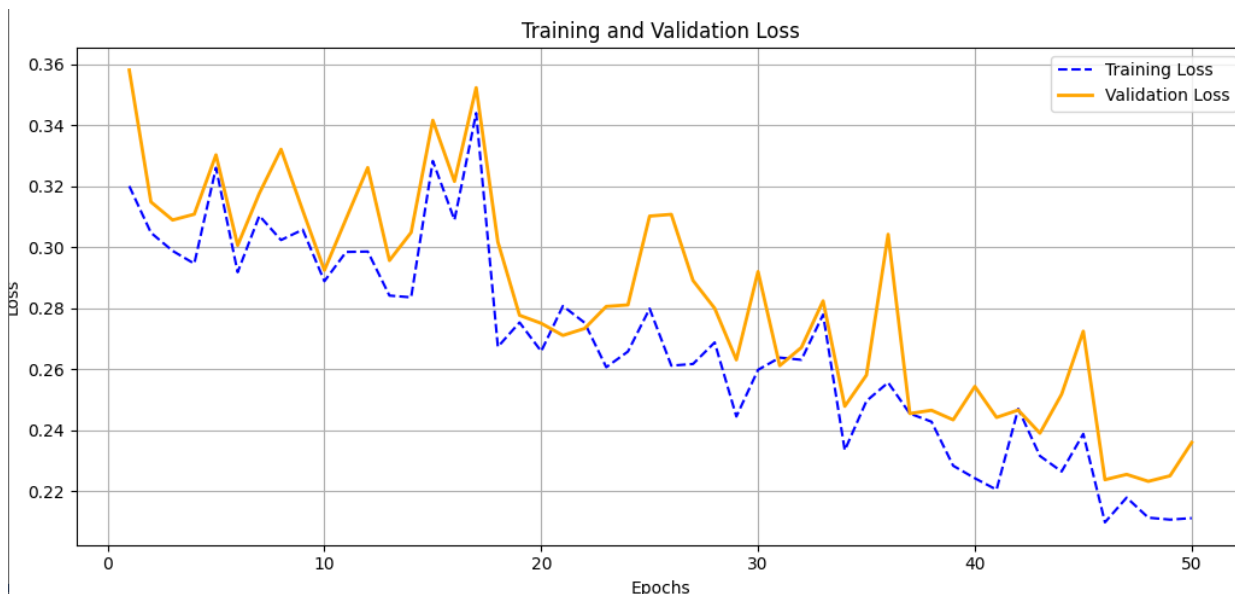


Рисунок 3.13 – Тренувальна та валідаційна похибки простої структури

Відповідна картина спостерігалась і під час аналізу порівняння результатів прогнозованих та реальних цін (див. рис. 3.14).

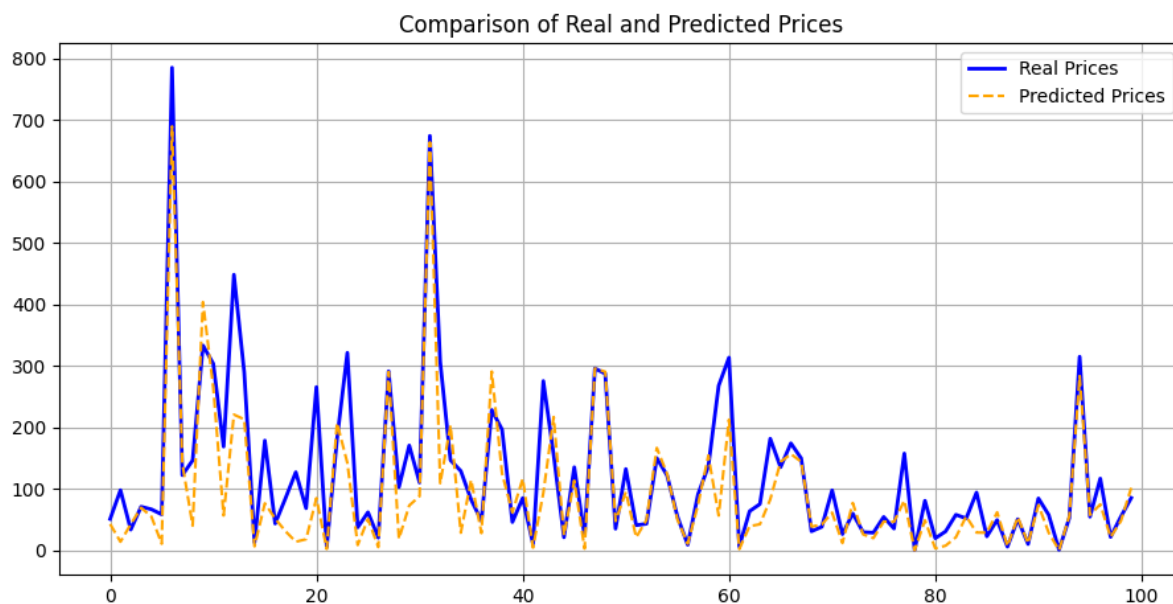


Рисунок 3.14 – Порівняння прогнозованих та реальних цін

З графіків стає зрозумілим, що структура моделі є занадто простою, щоб передбачати подібні стрибки, адже прогнозовані ціни, згідно графіку, є більш пологими, ніж реальні ціни з датасету, а показники R^2 , MAE та RMSE підтвердили

незадовільні результати. Вони становили 0.44, 95.66 та 228.77 відповідно, що показує значні відхилення і невисоку точність.

Відштовхуючись від цих результатів, наступна спроба була націлена зробити модель більш складною. Було додано 2 вхідних шари, в кожному з яких по 50 нейронів з активуючою функцією \tanh , кількість епох становила 50.

На цей раз, модель показала краще результати, але досі недостатньо хорошими. Графіки втрат явно зменшились (див. рис. 3.15).

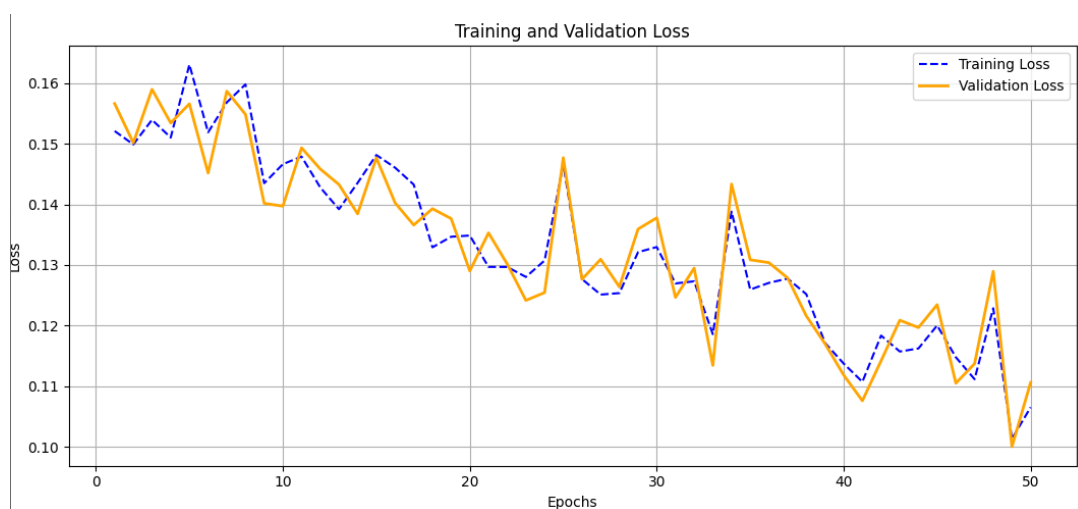


Рисунок 3.15 - Тренувальна та валідаційна похибки складнішої структури

Кращі результати спостерігаються також і в порівнянні прогнозованих цін із реальними (див. рис. 3.16).

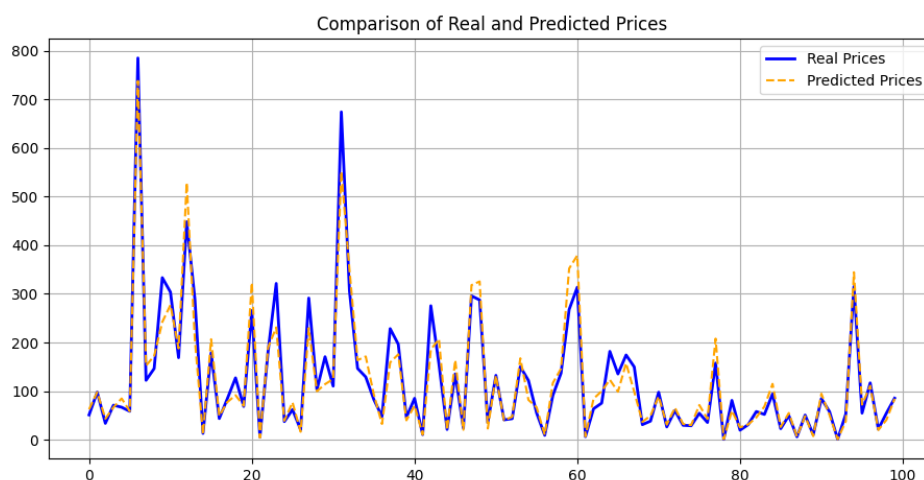


Рисунок 3.16 - Порівняння прогнозованих та реальних цін

Найголовніша метрика R^2 показувала вже значно кращі результати – 0.62, що сигналізує про прогрес і рух у правильному напрямку для досягнення кращої точності моделі.

Отже, модель потребувала ще більшого ускладнення. Тому цього разу модель складалась з 3 шарів з активуючою функцією \tanh , кожен з яких містив 100 нейронів. Також було додано шар dropout зі значенням 0.2 в кожен з вхідних шарів для підвищення ефективності навчання моделі. Було додано вихідний шар Dense. Модель була оптимізована за допомогою Adam. Кількість епох було збільшено з 50 до 100, але з можливістю early stopping (рання зупинка) з очікуванням в 10 епох, яка протягом вказаної кількості епох робить перевірку на позитивні зрушення показників валідаційних втрат. В разі, якщо позитивних змін нема, навчання призупиняється, що сигналізує про ймовірне перенавчання моделі.

Отже, цього разу результати вже були задовільними, це можна прослідкувати на графіках (див. рис. 3.17).

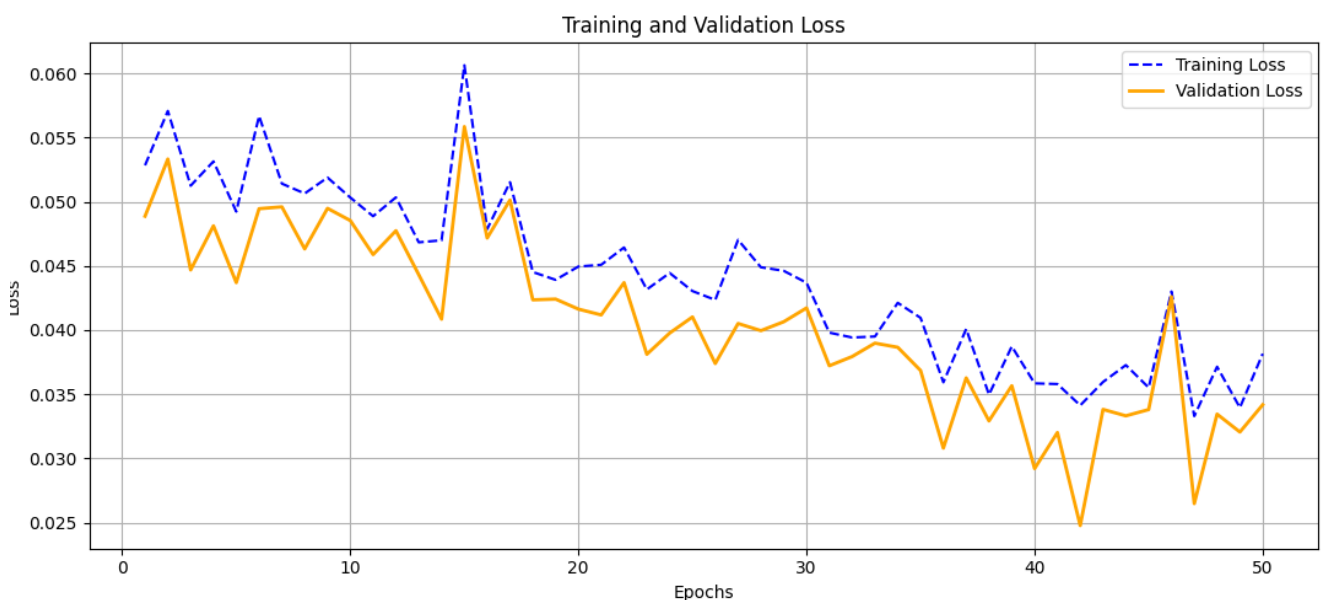


Рисунок 3.17 – Тренувальна та валідаційна похибки остаточного варіанту

Результати порівняння прогнозованих та реальних цін також показують гарну перспективу (див. рис. 3.18).

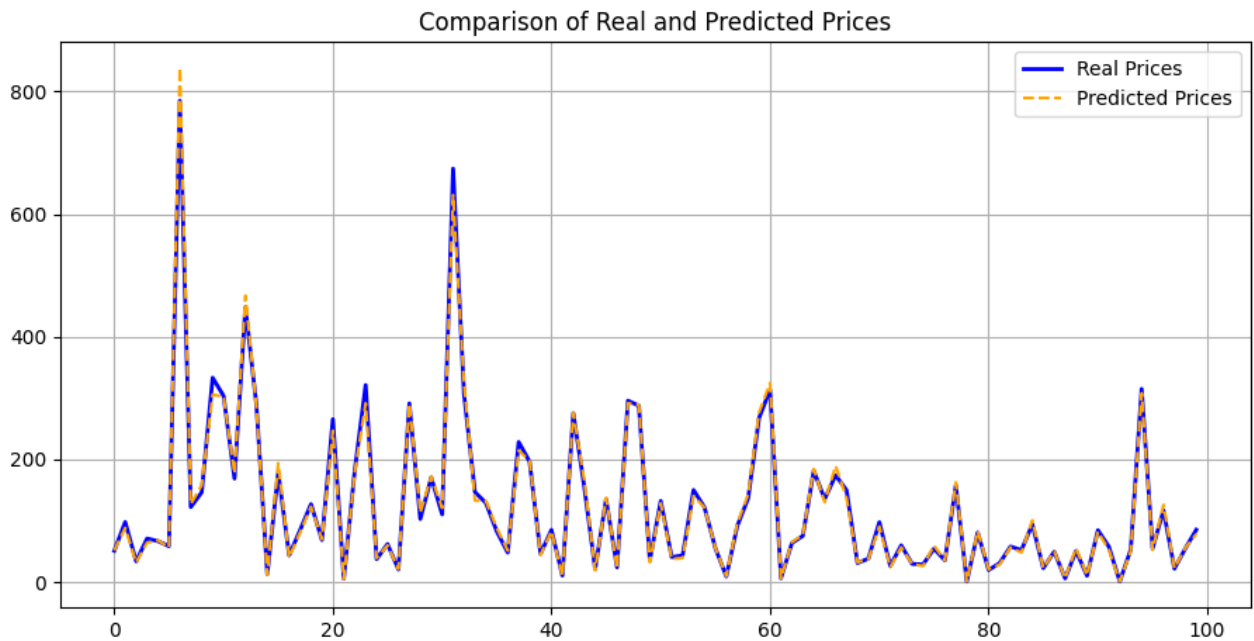


Рисунок 3.18 – Порівняння прогнозу моделлю

Супровідні метрики також сигналізують про непогані результати, адже R^2 показав 0.91, MAE – 0.098, RMSE – 0.144. Модель зупинила своє навчання на 50 епосі, адже далі результати вже були неефективними або навіть негативними, що могло призвести до перенавчання.

Після отримання наступних результатів необхідно було порівняти як працює модель з GRU архітектурою і чи здатна вона показати такі ж результати або навіть кращі. Так як вже було відомо, наскільки модель повинна бути складною, навчання почалося із такими ж налаштуваннями, але замість LSTM шару було взято GRU шар. Результати зображено на рис. 3.19.

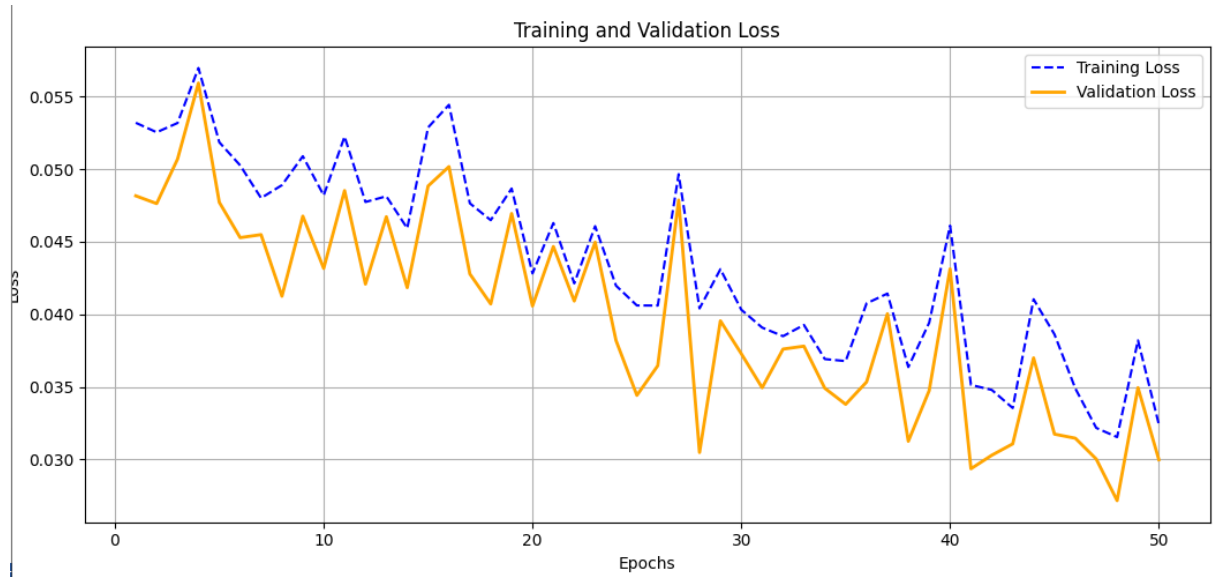


Рисунок 3.19 – Результати втрат на моделі GRU

Результати втрат показують дещо гірші результати, ніж до цього показувала себе модель LSTM. Показники R^2 становили 0.72, MAE – 0.177, RMSE – 0.215, хоч це і краще ніж початкова модель, все ж ці результати не є задовільними. Також результати прогнозів моделі вказують на певні проблеми із точністю (див. рис. 3.20).

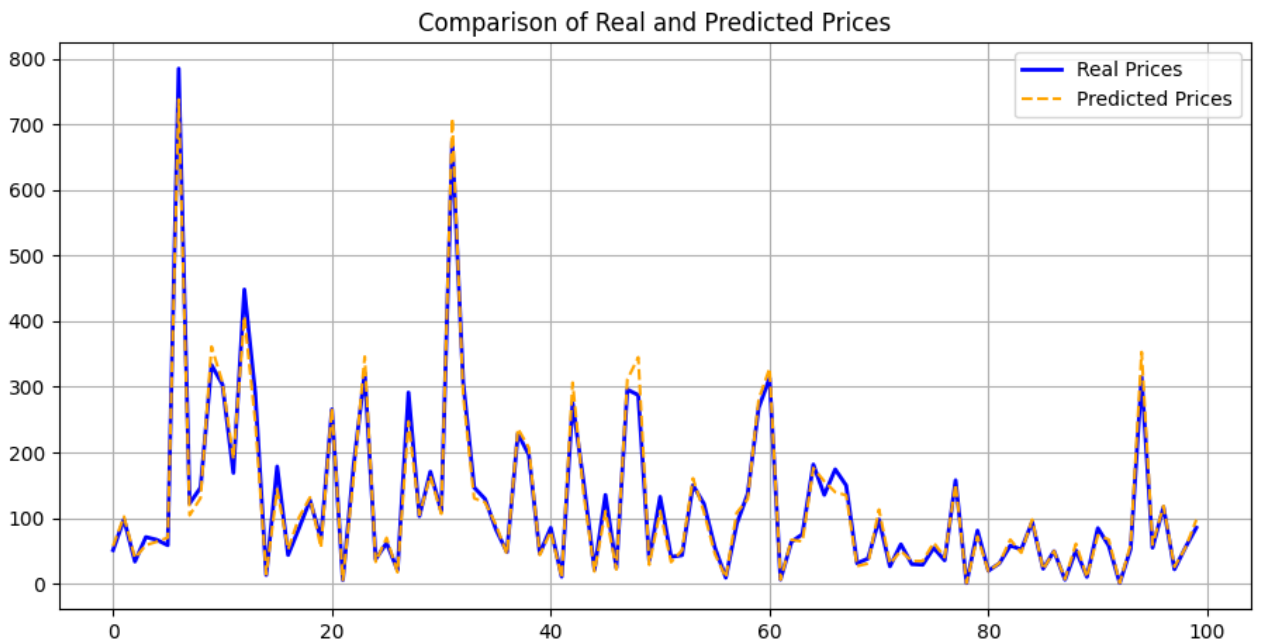


Рисунок 3.20 – Прогнозування моделі GRU

Також були наступні спроби зробити модель ще складнішою, адже місце для вдосконалення ще залишалось, але збільшення кількості нейронів не дало відчутних результатів, тому найкращим варіантом виявивалась саме модель на основі архітектури LSTM.

Отже, для підведення підсумків порівняння двох архітектур, було знайдено відмінності, зображені на табл. 3.1.

Таблиця 3.1 – Порівняння архітектур

Параметр	LSTM	GRU
Архітектура	Має три шлюзи: вхідний, вихідний та забуття (input, output, forget gates).	Має два шлюзи: оновлення та скидання (update, reset gates).
Параметри моделі	Більше параметрів (важчий обчислювально).	Менше параметрів (швидший у навчанні та прогнозуванні).
Гнучкість	Краще працює з довгими часовими рядами через складну обробку залежностей.	Менш обчислювально важкий, краще підходить для задач, де важливі короткі залежності.
Тренування	Може вимагати більше часу на тренування.	Швидше тренується завдяки простішій архітектурі.
Якість прогнозу	Часто забезпечує кращу точність для складних задач із довгими залежностями.	Може давати подібну точність, але з меншою обчислювальною складністю для простих задач.

Саме тому для обраної задачі найкраще підходить саме архітектура LSTM, хоч і навчається вона довше, найголовніше – результати метрик є значно кращими, ніж у моделі з архітектурою GRU.

3.3 Програмна середовище та програмна реалізація

Для написання коду для навчання нейронної моделі було обрано мову програмування Python, адже він має велику кількість бібліотек для виконання

подібних завдань (що дуже спрощує виконання завдання, адже не потрібно писати кожного разу якусь реалізацію того, що вже існує і створено іншими розробниками) і є доволі простою мовою для опанування, маючи нескладний синтаксис.

Для машинного навчання було обрано бібліотеки TensorFlow та Keras. Для створення моделей на основі LSTM-архітектури ці бібліотеки показали високу ефективність і зручність, адже TensorFlow має оптимізацію обчислень та підтримку GPU для прискорення обчислень, а Keras дозволяє створювати зрозумілі та складені модульні моделі.

Для роботи з даними використовувалась бібліотека Pandas. Це дуже потужний та зручний інструмент для обробки великих об'ємів даних, особливо у вигляді CSV або JSON. За допомогою цієї бібліотеки об'єднувались таблиці з різними колонками (merge-операція), вивантажувались та компонувались дані з різних джерел тощо. За допомогою бібліотеки NumPy забезпечувались швидкі математичні обчислення різного роду, зокрема операції з багатовимірними масивами. Щоб розподілити датасет на різні види, була використана бібліотека Scikit-learn. В її реалізацію входять такі інструменти як розподілення даних, перемішування, сортування тощо.

Для роботи з текстовими даними і настройками була використана бібліотека VADER (Valence Aware Dictionary and sEntiment Reasoner). Дана бібліотека спеціально налаштована для невеличких текстів, такі як заголовки новин або статей. За її допомогою було оброблено заголовки статей, що відносились до теми криптовалют і визначено їхні настрої. Хоча й ця бібліотека має можливість для вдосконалення для більш вузького використання, стандартна версія може бути використаною і має простий та інтуїтивно зрозумілий інструментарій.

Для інтеграцій та роботи з API використовувалась бібліотека Requests. Вона забезпечує зручний спосіб роботи з різними HTTP-запитами, що дозволяють отримати дані із зовнішніх джерел (в даному випадку для отримання даних з API криптологічних бірж та різних сервісів). Бібліотека підтримує асинхронні запити,

що суттєво прискорює роботу. Відповіді формувалися у зручному JSON-форматі, який переводився у потрібний CSV-формат за допомогою вищезгаданих бібліотек.

Щодо візуалізації, в нагоді стали бібліотеки Matplotlib та Seaborn. Matplotlib використовувався для побудови графіків, здебільшого підходить для відображення (див. рис. 3.21) лінійних графіків, показів історичних даних та трендів. Має розширений інструментарій для зручного перегляду результатів, такі як масштабування, збереження результатів у форматі PNG, збереження налаштувань для майбутніх досліджень (кастомізація). У свою чергу, Seaborn доповнює Matplotlib іншими функціями, що допомагають створювати більш інформативні графіки, можливість показу теплових карт та діаграм розсіювання.

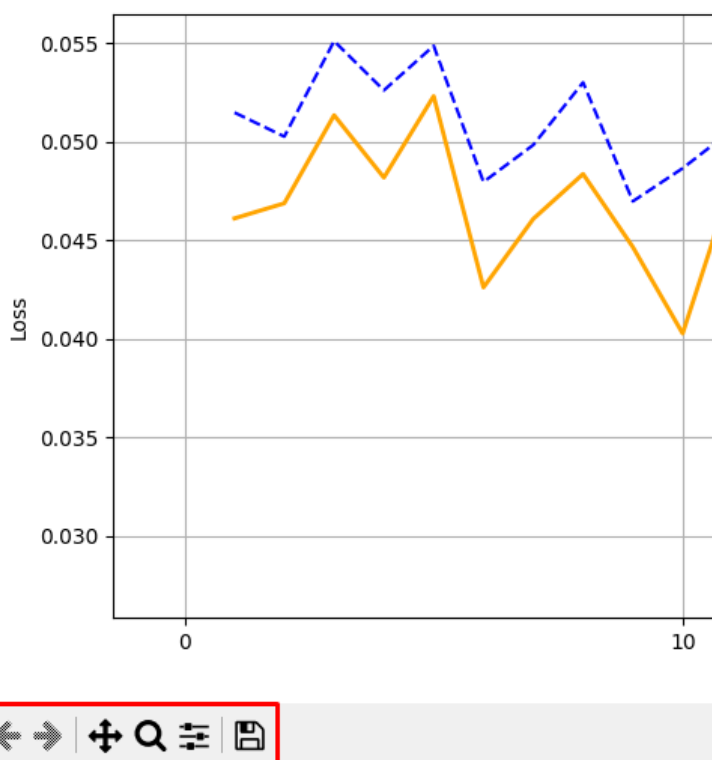


Рис. 3.21 – Інтерфейс відображення графіків за допомогою Matplotlib

В ролі середовища розробки виступав інтерфейс (рис. 3.22), розроблений за допомогою бібліотеки Jupyter Notebook. Ця бібліотека дозволяє зручно писати код, хоча й не має IntelliSense, але це компенсується швидким налаштуванням середовища і його розміром. Також дана бібліотека дозволяє тестувати та

документувати код. Jupyter підтримує кілька інших мов програмування, окрім Python, має інтегровані всередину бібліотеки Matplotlib, Seaborn та Plotly, що дозволяє створювати графіки прямо в середовищі. Також за допомогою інтеграції з Git та Google Colab можна робити сумісне редагування та розробку проєкту.

В середовищі можна розробляти проєкт навіть без локальної установки за допомогою хмарних сервісів. Також є можливість застосовувати безкоштовні GPU-потужності для машинного навчання моделей.

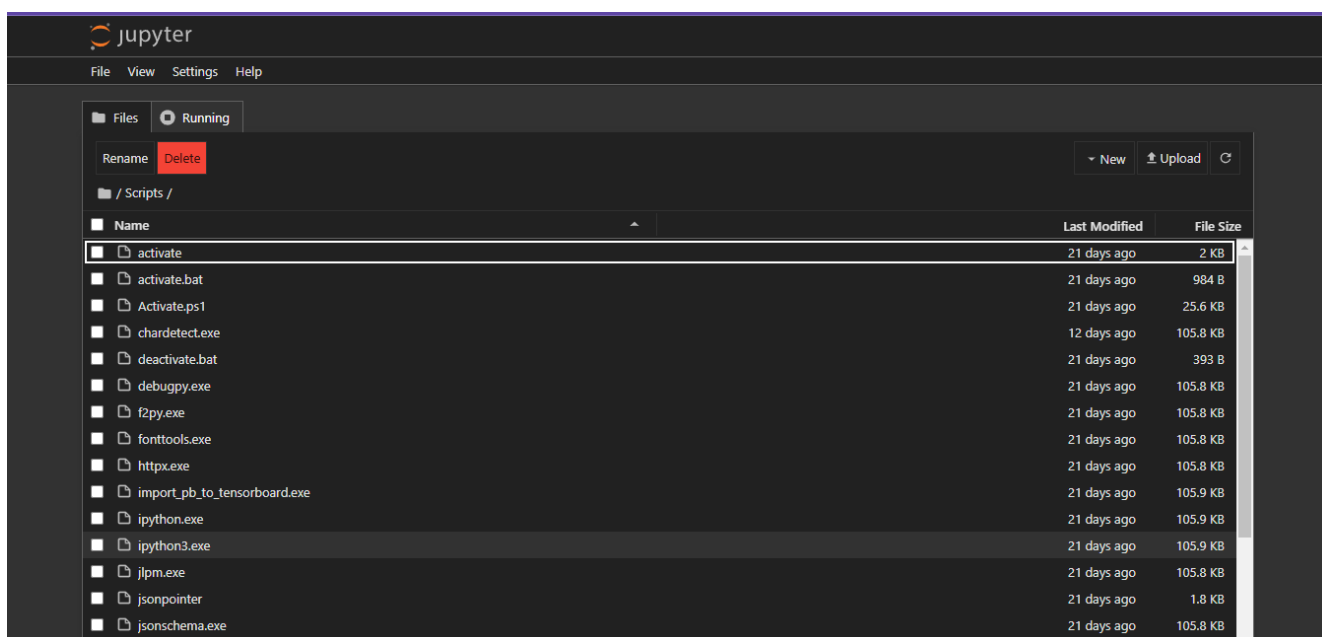


Рис. 3.22 – Інтерфейс Jupyter Notebook

Головним завданням програмної реалізації було машинне навчання моделі для прогнозування цін. Таке завдання не потребує певного інтерфейсу, тому було обрано невеликий стек технологій та бібліотек, що дозволило системі швидко та ефективно навчити модель, а також швидко налаштувати середовище до розробки проєкту. Необхідна реалізація наведена у додатках А-Г.

3.4 Практичне застосування

Однією із найбільших і найважливіших інтеграцій, які можна реалізувати, враховуючи дане дослідження – криптотрейдинг. Враховуючи популярність криптовалют, даний інструмент зібрав би неабияке охоплення та увагу серед

молоді і зацікавлених в нових технологіях трейдерів. Адже передбачення за допомогою нейронної моделі може допомогти інвестувати свій портфель більш якісно і зробити цю справу більш прибутковою.

Системи прогнозування допомагають трейдерам аналізувати ринок і приймати рішення щодо купівлі або продажу активів, використовуючи реальні дані та прогнозні індикатори.

Припустимо, система виявляє наступні фактори:

- зростання обсягу торгів на 30% за останні 24 години;
- позитивні новини про інтеграцію Bitcoin у платіжні системи;
- зниження волатильності.

На основі цих даних прогноз показує 80% ймовірність зростання ціни Bitcoin у короткостроковій перспективі. Трейдер може скористатися цією інформацією для купівлі активу до зростання його ціни.

Також даний інструмент може містити певні додаткові можливості, які можна реалізувати і таким чином покращити реалізацію:

- сигнали на основі моделей нейронних мереж (наприклад, сигнали купівлі/продажу);
- виявлення патернів на графіках (наприклад, "трикутник", "подвійне дно").

До переваг можна віднести:

- швидке реагування на ринкові зміни;
- зниження впливу емоційного фактору.

Другим варіантом реалізації може стати управління портефелем в інвестиційному фонді. Інвестиційні фонди використовують прогнози для оптимізації своїх криптовалютних портфелів, балансуючи між ризиками та потенційним прибутком. Наприклад: система прогнозує падіння ціни Ethereum на 15% через очікування високих витрат на газ у мережі. У відповідь фонд:

- скорочує частку Ethereum у портфелі;

- збільшує інвестиції в стейблкоїни, наприклад USDT або USDC, для зниження ризиків;
- розглядає інвестування у менш волатильні активи, такі як Bitcoin.

Додатковими стратегіями можуть виступити:

- диверсифікація портфеля на основі прогнозів кореляції між активами;
- інвестиції у нові трендові криптовалюти, такі як токени, пов'язані з DeFi чи NFT.

Оптимізація розподілу активів та мінімізація втрат під час ринкових спадів можуть стати перевагами інтеграції нейронної моделі прогнозування цін криптовалют в інвестиційний фонд.

Наступним варіантом для розробки інтеграції може стати розробка стратегій хеджування. Інституційні інвестори можуть використовувати прогнози для створення стратегій захисту від ризиків (хеджування) за допомогою похідних фінансових інструментів. Тобто, якщо система прогнозує падіння ціни Bitcoin, фонд може:

- купувати ф'ючерси на зниження ціни Bitcoin (шортові позиції);
- інвестувати в опціони типу "пут", які дозволяють продати актив за фіксованою ціною.
- хеджування ризиків, пов'язаних із волатильністю (наприклад, через контракти на волатильність);
- використання алгоритмів, що адаптуються до змін ринку в реальному часі.

Такий варіант реалізації інтеграції може стати потужним інструментом для зменшення інвестиційних ризиків і має наступні переваги:

- зменшення втрат у разі несприятливих змін;
- можливість заробити на падінні ринку.

Ще одним варіантом інтеграції може стати також регуляція ринку за допомогою виявлення маніпуляцій. Прогнозні системи можуть бути використані

регуляторами для моніторингу ринку та виявлення підозрілих дій, які можуть свідчити про маніпуляції або шахрайство. Наприклад, система прогнозує аномальне зростання ціни певної криптовалюти, не підкріплене жодними ринковими чи новинними факторами. Це може свідчити про:

- памп і дамп (Pump & Dump): штучне підвищення ціни активу, за яким слідує різкий продаж;
- інсайдерську торгівлю: використання закритої інформації для особистої вигоди.

Як наслідок, рішенням даної проблеми є комплексна інтеграція, що включає в себе:

- створення автоматизованих індикаторів для сповіщення про можливі ризики;
- аналіз динаміки ринку перед впровадженням нових регуляцій.

Це забезпечує наступні переваги:

- забезпечення стабільності ринку;
- захист інвесторів від шахрайських схем.

Серед подальших реалізацій можна додатково розглянути мобільний застосунок, що може збільшити охоплення користуванням сервісом, наприклад як на рис. 3.23.



Рисунок 3.23 – Приклад реалізації мобільного застосунку

Висновки до розділу 3

У третьому розділі було здійснено проектування системи для прогнозування цін на криптовалюти, що включає вибір архітектури нейронної мережі, обробку даних для її навчання та реалізацію в обраному програмному середовищі.

На основі аналізу вимог було обрано архітектуру рекурентної нейронної мережі LSTM, яка найбільш підходить для роботи з часовими рядами завдяки її

здатності ефективно обробляти залежності між попередніми та майбутніми значеннями. Було побудовано структуру моделі з оптимальними параметрами для навчання та тестування на основі зібраних даних.

Процес збору, обробки та підготовки даних включав інтеграцію даних з різних джерел, зокрема даних про торгові обсяги, макроекономічних показників, хешрейту та новинних індексів. Проведено очищення, нормалізацію та трансформацію даних, що забезпечило їх придатність для використання в навчанні нейронної мережі. Відбулась також інтерполяція даних лінійним та кубічним шляхом в залежності від схильності певних даних до різких стрибків в показниках.

Для реалізації системи було обрано програмне середовище Python з використанням бібліотек TensorFlow, Keras та багатьох інших, які дозволяють ефективно реалізувати і тестувати моделі глибокого навчання. Розроблено програмний код, що включає модулі для збору, обробки даних, побудови та навчання моделі, а також оцінки її ефективності за допомогою візуалізації у вигляді графіків.

Результати цього розділу забезпечують основу для проведення експериментів та аналізу ефективності побудованої системи, що дозволяє оцінити її практичну застосовність у задачах прогнозування цін на криптовалюту.

Практична частина цього дослідження показує важливість реалізації та можливе практичне застосування у наступних варіантах:

- трейдинг;
- управління портфелями в інвестиційних фондах;
- розробка стратегій хеджування (зменшення ризиків);
- регулятори ринку (виявлення маніпуляцій).

Розширений вибір реалізацій показує широкий спектр можливостей для зменшення ризиків під час процесу інвестування та зменшує витрачений час на обробку даних людиною власноруч, покладаючись на результат нейронної мережі, враховуючи її високу точність.

ВИСНОВКИ

У цій кваліфікаційній роботі було розроблено інтелектуальну систему для прогнозування цін криптовалют із використанням нейронних мереж, зокрема рекурентних нейронних мереж, що використовує архітектуру LSTM. В рамках дослідження було здійснено детальний аналіз існуючих підходів до прогнозування фінансових ринків, таких як Support Vector Machines, Random Forest, XGBoost та інші алгоритми машинного навчання. Визначено основні недоліки цих методів, зокрема обмеженість в обробці тимчасових рядів і недостатня точність при роботі з нестабільними даними, характерними для ринку криптовалют.

Розроблена система на основі LSTM показала високу точність прогнозування, зокрема за рахунок можливості моделювання залежностей у часових рядах. Метрики оцінки моделі, такі як MAE, RMSE та R^2 , продемонстрували покращені результати у порівнянні з іншими методами, що підтверджує доцільність використання LSTM для цієї задачі.

Інформаційні технології для збору та аналізу даних, а також застосовані математичні моделі та методи забезпечили гнучкість системи, її здатність швидко адаптуватися до змін на ринку та працювати з великими обсягами даних. Окрім того, було враховано можливість інтеграції додаткових змінних, таких як соціальні та економічні фактори, для подальшого підвищення точності прогнозів.

Загалом, дослідження показало, що використання рекурентних нейронних мереж у поєднанні з сучасними інформаційними технологіями має високий потенціал для вирішення складних задач фінансового прогнозування, зокрема у сфері криптовалют.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. "Were There Cryptocurrencies Before Bitcoin?" Investopedia : вебсайт. URL: <https://www.investopedia.com/tech/were-there-cryptocurrencies-bitcoin/> (last accessed 10.10.2024).
2. "CEF Digital – Newsroom." Європейська Комісія : вебсайт. URL: <https://ec.europa.eu/newsroom/cef/items/658303> (дата звернення 11.10.2024).
3. "Про нас." Bitcoin.org : вебсайт. URL: <https://bitcoin.org/uk/about-us> (дата звернення 11.10.2024).
4. "Genesis Block." Binance Academy : вебсайт. URL: <https://academy.binance.com/en/glossary/genesis-block> (last accessed 11.10.2024).
5. "What is Litecoin?" Investopedia : вебсайт. URL: <https://www.investopedia.com/terms/l/litecoin.asp> (last accessed 11.10.2024).
6. "Number of Cryptocurrencies in Circulation." Statista : вебсайт. URL: <https://www.statista.com/statistics/863917/number-crypto-coins-tokens/> (last accessed 11.10.2024).
7. "Bitcoin." CoinMarketCap : вебсайт. URL: <https://coinmarketcap.com/currencies/bitcoin/> (last accessed 11.10.2024).
8. "What is Decentralized Finance?" Hedera : вебсайт. URL: <https://hedera.com/learning/decentralized-finance/what-is-decentralized-finance> (last accessed 12.10.2024).
9. "A Very Short History of Ethereum Everyone Should Read." Bernard Marr : вебсайт. URL: <https://bernardmarr.com/blockchain-a-very-short-history-of-ethereum-everyone-should-read/> (last accessed 12.10.2024).
10. "What is Mt. Gox?" Trust Wallet : вебсайт. URL: <https://trustwallet.com/uk/blog/mt-gox-explained> (дата звернення 12.10.2024).
11. "Cryptocurrency Regulations in Japan." Sanction Scanner : вебсайт. URL: <https://www.sanctionscanner.com/blog/cryptocurrency-regulations-in-japan-492> (last accessed 12.10.2024).

12. "Price of Cryptocurrencies Totally Dependent on China." Investopedia : вебсайт. URL: <https://www.investopedia.com/news/price-cryptocurrencies-totally-dependent-china/> (last accessed 12.10.2024).
13. "CoinMarketCap." CoinMarketCap : вебсайт. URL: <https://coinmarketcap.com/> (last accessed 12.10.2024).
14. "Tesla Bitcoin Holdings Reclaim \$1 Billion Threshold." Investing : вебсайт. URL: <https://in.investing.com/news/cryptocurrency-news/tesla-bitcoin-holdings-reclaim-1-billion-threshold-4521679> (last accessed 14.10.2024).
15. "MicroStrategy Buys \$2.03 Billion Worth of Bitcoin." Yahoo Finance : вебсайт. URL: <https://finance.yahoo.com/news/microstrategy-buys-2-03-billion-133103841.html> (last accessed 14.10.2024).
16. "The History of NFTs." LCX : вебсайт. URL: <https://www.lcx.com/the-history-of-nfts/> (last accessed 14.10.2024).
17. "Around the Block: Issue 11." Coinbase : вебсайт. URL: <https://www.coinbase.com/learn/market-updates/around-the-block-issue-11> (last accessed 14.10.2024).
18. "Christie's Auction Highlights 2021." Christie's : вебсайт. URL: <https://www.christies.com/en/stories/christies-auction-highlights-2021-b74452f502fe4587b29119b5cd537f25> (last accessed 14.10.2024).
19. "Crypto Market Today." CNBC : вебсайт. URL: <https://www.cnbc.com/2024/03/04/crypto-market-today.html> (last accessed 17.10.2024).
20. "Smart Contracts." Ethereum : вебсайт. URL: <https://ethereum.org/en/developers/docs/smart-contracts/> (last accessed 17.10.2024).
21. "Uniswap V2 Smart Contracts." Uniswap : вебсайт. URL: <https://docs.uniswap.org/contracts/v2/concepts/protocol-overview/smart-contracts> (last accessed 17.10.2024).
22. "Smart Contracts." Aave : вебсайт. URL: <https://aave.com/docs/developers/smart-contracts> (last accessed 18.10.2024).

23. "How the US and Europe Are Regulating Crypto in 2020." CoinTelegraph : вебсайт. URL: <https://cointelegraph.com/news/how-the-us-and-europe-are-regulating-crypto-in-2020> (last accessed 18.10.2024).
24. "Proof of Work." Investopedia : вебсайт. URL: <https://www.investopedia.com/terms/p/proof-work.asp> (last accessed 15.10.2024).
25. "Proof of Stake Explained." Binance Academy : вебсайт. URL: <https://academy.binance.com/da/articles/proof-of-stake-explained> (last accessed 15.10.2024).
26. "Bitcoin Rising Fees, Confirmation Queues See Users, Investors Switching to Altcoins." CoinTelegraph : вебсайт. URL: <https://cointelegraph.com/news/bitcoin-rising-fees-confirmation-queues-see-users-investors-switching-to-altcoins> (last accessed 11.10.2024).
27. "Growth of Blockchain Market Will Reach \$67.4B by 2026." MarketsandMarkets : вебсайт. URL: <https://blockchain.news/news/growth-of-blockchain-market-will-reach-67.4b-by-2026-marketsandmarket-report> (last accessed 20.10.2024).
28. "Delegated Proof of Stake (DPoS)." Tangem : вебсайт. URL: <https://tangem.com/en/blog/post/delegated-proof-of-stake-dpos/> (last accessed 21.10.2024).
29. "What Is Blockchain and How Does It Work?" Binance Academy : вебсайт. URL: <https://academy.binance.com/uk/articles/what-is-blockchain-and-how-does-it-work> (last accessed 21.10.2024).
30. "Crypto Predict Using RNN and CNN." DIVA Portal : вебсайт. URL: <https://www.diva-portal.org/smash/get/diva2:1778251/FULLTEXT03.pdf> (last accessed 10.11.2024).
31. Kim, S. "Factors Influencing the Trend of Cryptocurrency." ResearchGate : вебсайт. URL: https://www.researchgate.net/publication/357982093_Factors_Influencing_the_Trend_of_Cryptocurrency (last accessed 10.11.2024).

32. "Bitcoin Price Prediction Using Machine Learning Techniques." IEEE Xplore : вебсайт. URL: <https://ieeexplore.ieee.org/abstract/document/7966019> (last accessed 11.11.2024).
33. Villaruz, C. "A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM, and bi-LSTM Machine Learning Algorithms." ResearchGate : вебсайт. URL: https://www.researchgate.net/publication/355202227_A_Novel_Cryptocurrency_Price_Prediction_Model_Using_GRU_LSTM_and_bi-LSTM_Machine_Learning_Algorithms (last accessed 11.11.2024).
34. "Blockchain-based Cryptocurrency Management." ScienceDirect : вебсайт. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0040162522006333> (last accessed 11.11.2024).
35. "A Review of Deep Learning Applications for Cryptocurrency." IEEE Xplore : вебсайт. URL: <https://ieeexplore.ieee.org/document/10113599> (last accessed 11.11.2024).
36. "Cryptocurrency Forecasting Models." MDPI : вебсайт. URL: <https://www.mdpi.com/2079-9292/11/15/2349> (last accessed 11.11.2024).
37. "Improved Cryptocurrency Forecasting with Random Forest." IEEE Xplore : вебсайт. URL: <https://ieeexplore.ieee.org/document/10689789> (last accessed 11.11.2024).
38. "Predicting the Price of Cryptocurrency Using Support Vector Regression Methods." ResearchGate : вебсайт. URL: https://www.researchgate.net/publication/340926463_Predicting_the_Price_of_Cryptocurrency_using_Support_Vector_Regression_Methods (last accessed 11.11.2024).
39. "Cryptocurrency Price Forecasting Models Using Machine Learning." ScienceDirect : вебсайт. URL: <https://www.sciencedirect.com/science/article/pii/S1877050919321647> (last accessed 12.11.2024).
40. "XGBoost Cryptocurrency Prediction Model." IEEE Xplore : вебсайт. URL: <https://ieeexplore.ieee.org/document/9935871> (last accessed 13.11.2024).

41. "A Comparative Study of Cryptocurrency Forecasting Models." IEEE Xplore : вебсайт. URL: <https://ieeexplore.ieee.org/document/9758430> (last accessed 11.11.2024).
42. "Cryptocurrency Market Predictions with Machine Learning." ScienceDirect : вебсайт. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0169023X22000234> (last accessed 15.11.2024).
43. Villaruz, C. "Crypto Predict Using RNN and CNN." Kaggle : вебсайт. URL: <https://www.kaggle.com/code/xtianvillaruz/crypto-predict-using-rnn-and-cnn> (last accessed 14.11.2024).
44. "10 Elon Musk Tweets That Created Waves in Crypto World." Outlook Business : вебсайт. URL: <https://www.outlookbusiness.com/news/10-elon-musk-tweets-that-created-waves-in-crypto-world-news-233190> (last accessed 01.12.2024).
45. "China's Digital Currency Test: A Look at the World's First Central Bank Digital Currency." BBC : вебсайт. URL: <https://www.bbc.com/news/articles/c6246e3w935o> (last accessed 01.12.2024).

ДОДАТОК А

Лістинг коду веб-скрапера для збору даних EFFF

```
import requests
import time
import logging
import json
import pandas as pd
from datetime import datetime, timedelta

# Налаштування логуювання
logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s - %(message)s')
logger = logging.getLogger()

# URL бази та початкові параметри
base_url = "https://markets.newyorkfed.org/read"
params = {
    'startDt': '2008-01-01',
    'endDt': '2024-11-19',
    'eventCodes': '500',
    'productCode': '50',
    'sort': 'postDt:-1,eventCode:1',
    'limit': 100,
    'startPosition': 0
}

# Кількість запитів (будемо робити пагінацію)
total_records = 4243 # Всього записів
records_per_page = 100 # Кількість записів на одну сторінку
total_requests = total_records // records_per_page # Загальна кількість запитів

# Збираємо всі дані в список
all_data = []

# Пагінація для всіх даних
for i in range(total_requests):
    params['startPosition'] = i * records_per_page
    logger.info(f"Збір даних для позиції: {params['startPosition']}")

    # Відправка запиту
    try:
        response = requests.get(base_url, params=params)
        response.raise_for_status() # Перевірка на помилки запиту
        data = response.json()

        # Перевірка, чи є дані в відповіді
        if "data" in data:
            all_data.extend(data["data"])
            logger.info(f"Отримано {len(data['data'])} записів для позиції {params['startPosition']}")
        else:
            logger.warning(f"Немає даних для позиції {params['startPosition']}")
    except requests.exceptions.RequestException as e:
        logger.error(f"Помилка при отриманні даних для позиції {params['startPosition']}: {e}")
        break

# Затримка 1-2 секунди перед наступним запитом
time.sleep(1 + i % 2) # Затримка між запитом
```

```
# Перевірка чи є зібрані дані
if all_data:
    logger.info(f"Зібрано {len(all_data)} записів.")
else:
    logger.warning("Не вдалося отримати жодних даних.")

# Генерація всіх дат в діапазоні
start_date = datetime(2008, 1, 1)
end_date = datetime.today()
all_dates = pd.date_range(start=start_date, end=end_date).date

# Створюємо словник для збереження ставок
effr_data = {date: None for date in all_dates}

# Обробка даних
for entry in all_data:
    try:
        post_date = entry['postDt']
        daily_rate = json.loads(entry['data'])['dailyRate']
        effr_data[datetime.strptime(post_date, '%Y-%m-%d').date()] = daily_rate
    except Exception as e:
        logger.error(f"Помилка обробки даних: {e}")

# Перевірка зібраних даних
logger.info(f"Оброблено {len(effr_data)} дат.")

# Перетворюємо на DataFrame
df = pd.DataFrame(list(effr_data.items()), columns=['Date', 'EFFR'])

# Перевірка на пропуски
if df['EFFR'].isnull().sum() > 0:
    logger.warning(f"Є пропущені ставки для {df['EFFR'].isnull().sum()} дат.")

# Збереження в CSV
df.to_csv('effr_data_full.csv', index=False)
logger.info("Дані збережено у файл effr_data_full.csv")
```

ДОДАТОК Б

Лістинг коду веб-скрапера новин про криптовалюти та оцінку настроїв

НОВИН

```
import asyncio
import aiohttp
from bs4 import BeautifulSoup
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
import pandas as pd
from tqdm import tqdm # Для прогрес бару
import random # Для випадкових затримок

# Функція для збору даних зі сторінки
async def get_article_data(session, page_url):
    headers = {
        "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko)
        Chrome/58.0.3029.110 Safari/537.3"
    }

    # Додаємо випадкову затримку між запитами (від 1 до 3 секунд)
    delay = random.uniform(1, 3)
    await asyncio.sleep(delay)

    async with session.get(page_url, headers=headers) as response:
        soup = BeautifulSoup(await response.text(), "html.parser")

    # Збір заголовків новин, посилань та дат тільки з <section class="list-feed grid">
    section = soup.find("section", class_="list-feed grid")
    if not section:
        return [] # Якщо секція не знайдена, повертаємо порожній список

    articles = section.find_all("div", class_="list-post")
    article_data = []

    for article in articles:
        title_tag = article.find("a", {"title": True})
        if title_tag:
            title = title_tag["title"]
            link = title_tag["href"]

            # Переходимо по посиланню на саму статтю, щоб знайти точну дату публікації
            async with session.get(link, headers=headers) as article_response:
                article_soup = BeautifulSoup(await article_response.text(), "html.parser")

            # Перевірка на наявність першого варіанту для дати
            date_tag = article_soup.find("div", class_="post-date")
            if not date_tag: # Якщо не знайдено, шукаємо другий варіант
                date_tag = article_soup.find("span", class_="post-date col")

            # Якщо знайдено дату
            if date_tag:
                date = date_tag.get_text(strip=True)
            else:
                date = "Date not found"
```

```

# Аналіз настрою заголовка
analyzer = SentimentIntensityAnalyzer()
sentiment = analyzer.polarity_scores(title)
positive_score = sentiment["pos"]
negative_score = sentiment["neg"]

# Збираємо дані
article_data.append({
    "title": title,
    "positive_sentiment": positive_score,
    "negative_sentiment": negative_score,
    "date": date,
    "link": link
})

return article_data

# Збір новин з кількох сторінок
async def scrape_news(base_url, num_pages):
    all_articles = []

    async with aiohttp.ClientSession() as session:
        tasks = []

        # Використовуємо tqdm для відслідковування прогресу
        with tqdm(total=num_pages, desc="Scraping pages", ncols=100) as pbar:
            for page_num in range(1, num_pages + 1):
                page_url = f"{base_url}/page/{page_num}/"
                task = asyncio.ensure_future(get_article_data(session, page_url))
                tasks.append(task)

        # Очікуємо виконання всіх завдань
        for future in asyncio.as_completed(tasks):
            result = await future
            all_articles.extend(result)
            pbar.update(1) # Оновлення прогресу після кожної сторінки

    return all_articles

# Основна частина скрипту
base_url = "https://cryptoslate.com/news/bitcoin"
num_pages = 560 # Збір даних з УСІХ сторінок сайту

async def main():
    articles = await scrape_news(base_url, num_pages)

    # Збереження даних у CSV файл
    df = pd.DataFrame(articles)
    df.to_csv("bitcoin_news_sentiment.csv", index=False, encoding='utf-16')

    print("Дані збережено у файл 'bitcoin_news_sentiment.csv'")

# Запуск асинхронного скрипту
asyncio.run(main())

```


ДОДАТОК В

Лістинг коду веб-скрапера для збору даних про макроекономічні показники

```
from bs4 import BeautifulSoup
import csv
from datetime import datetime, timedelta

# Шлях до HTML-файлу
html_file = "data.html"
output_file = "dxy_data.csv"

# Завантаження HTML-файлу
with open(html_file, "r", encoding="utf-8") as file:
    soup = BeautifulSoup(file, "html.parser")

# Знайти всі рядки з даними
rows = soup.select("tr.historical-data-v2_price__atUfP")

# Парсинг доступних даних
data = {}
for row in rows:
    # Витягти дату
    date_tag = row.select_one("td time")
    raw_date = date_tag["datetime"] if date_tag else None

    # Форматувати дату в стандарт ISO (%Y-%m-%d)
    if raw_date:
        date = datetime.strptime(raw_date, "%b %d, %Y").strftime("%Y-%m-%d")

    # Витягти ціну
    price_tag = row.select_one("td.datatable_cell--align-end__qgxDQ")
    price = price_tag.text.strip().replace(",", "") if price_tag else None

    if raw_date and price:
        data[date] = float(price)

# Заповнення дат за повний період
start_date = datetime(2008, 1, 1) # Початкова дата
end_date = datetime(2024, 11, 19) # Кінцева дата
filled_data = []

current_date = start_date
while current_date <= end_date:
    date_str = current_date.strftime("%Y-%m-%d")
    price = data.get(date_str) # Отримати ціну, якщо є
    filled_data.append({"Date": date_str, "Price": price})
    current_date += timedelta(days=1)

# Збереження в CSV
with open(output_file, "w", encoding="utf-8", newline="") as csvfile:
    fieldnames = ["Date", "Price"]
    writer = csv.DictWriter(csvfile, fieldnames=fieldnames)

    writer.writeheader()
    writer.writerows(filled_data)

print(f'Дані заповнено і збережено в {output_file}')
```


ДОДАТОК Г

Лістинг коду скрипту для підготовки погодинних даних

```
import csv
import numpy as np
from scipy.interpolate import interp1d
import os

# Шлях до директорії з файлами
input_directory = "E:/labs/myenv/data/btc/final_with_values"
output_directory = input_directory # Збережемо результати в ту ж директорію

# Створення директорії для результатів, якщо вона не існує
os.makedirs(output_directory, exist_ok=True)

# Обробка всіх файлів у директорії
for input_file in os.listdir(input_directory):
    if input_file.endswith("_final_with_values.csv"): # Перевіряємо на файли з потрібним суфіксом
        input_path = os.path.join(input_directory, input_file)
        output_file = input_file.replace("_final_with_values.csv", "_final_interpolated.csv")
        output_path = os.path.join(output_directory, output_file)

        # Читання файлу
        with open(input_path, 'r') as file:
            reader = csv.reader(file, delimiter=',')
            lines = list(reader)

        # Ініціалізація даних
        timestamps = []
        values = []

        # Збір значень з другого стовпця
        for i, line in enumerate(lines):
            if i == 0: # Пропускаємо заголовок
                continue
            timestamp = line[0]
            value = line[1]
            timestamps.append(timestamp)
            if value == "a": # Позначаємо незаповнені значення як NaN
                values.append(np.nan)
            else:
                values.append(float(value.replace(',', '.'))) # Заміна коми на крапку для чисел

        # Знаходимо індекси відомих значень
        known_indices = ~np.isnan(values)
        unknown_indices = np.isnan(values)

        # Підготовка до інтерполяції
        x_known = np.arange(len(values))[known_indices]
        y_known = np.array(values)[known_indices]
        x_unknown = np.arange(len(values))[unknown_indices]

        # Кубічна інтерполяція
        interp_function = interp1d(x_known, y_known, kind='cubic', fill_value="extrapolate")
        interpolated_values = interp_function(x_unknown)

        # Заповнення значень у масиві
        for i in range(len(values)):
```

```
if np.isnan(values[i]):
    interpolated_value = interpolated_values[np.where(x_unknown == i)[0][0]]
    # Якщо значення менше 0, замінюємо його на 0
    values[i] = max(0, interpolated_value)

# Форматування значень
values = [f"{v:.4f}".replace('.', ',') for v in values]

# Запис у новий файл
with open(output_path, 'w', newline='') as file:
    writer = csv.writer(file, delimiter=',')
    for i, line in enumerate(lines):
        if i == 0: # Пишемо заголовок без змін
            writer.writerow(line)
        else: # Записуємо інтерпольовані значення
            line[1] = values[i - 1]
            writer.writerow(line)

print(f"Інтерпольовані дані для файлу {input_file} збережено у новому файлі: {output_path}")
```