

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАДИМА ГЕТЬМАНА

Навчально-науковий інститут
«Інститут інформаційних технологій в економіці»
Кафедра математичного моделювання та статистики

Освітньо-професійна програма	Економічна кібернетика і Дата Сайнс
Галузь знань	05 Соціальні та поведінкові науки
Спеціальність	051 Економіка

Форма навчання: очна (денна)

КВАЛІФІКАЦІЙНА МАГІСТЕРСЬКА РОБОТА
на тему «Аналіз та прогнозування динаміки цін на криптовалютному ринку
методами машинного та глибокого навчання»

здобувача **Каїки Олексія Павловича**

Науковий керівник: доктор фізико-математичних наук,
професор Притоманова О.М.

Робота допущена до захисту перед екзаменаційною комісією
з атестації здобувачів вищої освіти (ЕК)

Завідувач кафедри **кандидат фізико-математичних наук,**
професор Великоіваненко Г.І.

Київ 2023

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАДИМА ГЕТЬМАНА**

Інститут інформаційних технологій в економіці

Кафедра математичного моделювання та статистики

**ОСВІТНЬО-ПРОФЕСІЙНА ПРОГРАМА
«Економічна кібернетика і Дата Сайнс»**

Галузь знань 05 Соціальні та поведінкові науки

Спеціальність 051 Економіка

ПОГОДЖЕНО

Керівник проєктної групи (гарант)
освітньо-професійної програми

ЗАТВЕРДЖУЮ

Завідувач кафедри

(підпис) О.В. Піскунова
20__ р. (ініціали, прізвище)

(підпис) Г.І. Великоіваненко
20__ р. (ініціали, прізвище)

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ

здобувача вищої освіти

Каїка Олексій Павлович
(прізвище, ім'я, по батькові)

_____очної (денної)_____ **форми навчання**
на підготовку кваліфікаційної магістерської роботи

на тему

«Аналіз та прогнозування динаміки цін на криптовалютному ринку методами машинного та глибокого навчання»

Тему затверджено наказом ректора Університету від «__» _____ 20__ р. № _____
Кваліфікаційна магістерська робота виконується на матеріалах офіційних сайтів криптовалютних бірж.

План кваліфікаційної магістерської роботи

Розділ 1	Теоретичні та методичні основи прогнозування динаміки цін на криптовалютному ринку
Розділ 2	Застосування методів машинного та глибокого навчання для аналізу цінових коливань криптовалюти
Розділ 3	Застосування методів машинного та глибокого навчання для прогнозування динаміки цін криптовалют

Об'єкт дослідження:	ціни криптовалют на криптовалютних біржах
Предмет дослідження:	методи машинного та глибокого навчання для аналізу та прогнозування динаміки цін на криптовалютних біржах
Мета виконання кваліфікаційної магістерської роботи:	вибір ефективних моделей машинного та глибокого навчання та їх застосування для аналізу та прогнозування динаміки цін на криптовалютному ринку.

Конкретні завдання, які здобувач повинен виконати для досягнення поставленої мети:

У розділі 1:

- 1) огляд криптовалют та їх бірж для гривні;
- 2) огляд джерел з прогнозами цін криптовалют;
- 3) порівняльний аналіз криптовалют та їх відбір для подальшого моделювання динаміки їх цін.

У розділі 2:

- 1) інформаційна база для моделювання; графічний аналіз динаміки цін обраних криптовалют та факторів, що можуть впливати на ціни криптовалют;
- 2) методи машинного та глибокого навчання для аналізу прибутковості та ризиковості криптовалют.

У розділі 3:

- 1) методи машинного навчання для прогнозування цін криптовалют;
- 2) методи глибокого навчання для прогнозування цін криптовалют
- 3) порівняння застосованих методів та отриманих прогнозів.

Завдання підготував
науковий керівник

(підпис)

О.М. Притоманова
(ініціали, прізвище)
«__» _____ 2023 р.

Завдання одержав
здобувач

(підпис)

О.П. Каїка
(ініціали, прізвище)
«__» _____ 2023 р.

Реферат

Кваліфікаційна магістерська робота містить 77 сторінок, 48 рисунків, 1 таблицю, перелік використаних джерел з 53 найменувань.

«Аналіз та прогнозування динаміки цін на криптовалютному ринку методами машинного та глибокого навчання»

Об'єктом дослідження кваліфікаційної магістерської роботи є ціни криптовалют на криптовалютних біржах.

Предметом дослідження є методи машинного та глибокого навчання, які застосовуються для аналізу та прогнозування динаміки цін на криптовалютних біржах.

Мета і завдання дослідження. Основною метою кваліфікаційної магістерської роботи є вибір ефективних моделей машинного та глибокого навчання та їх застосування для аналізу та прогнозування динаміки цін на криптовалютному ринку.

Відповідно до поставленої мети визначені такі *завдання*:

- опрацювати джерела з аналізу та прогнозування цін на криптовалютному ринку;
- провести огляд криптовалютних бірж, інструментів та засобів для аналізу та прогнозування динаміки цін на криптовалютному ринку, відзначити правові аспекти для українських користувачів для обґрунтування актуальності обраної теми дослідження;
- охарактеризувати інформаційну базу для моделювання та провести графічний аналіз динаміки ціни обраної криптовалюти;
- провести аналіз факторів та їх впливу на ціну криптовалют;
- розробити та налаштувати моделі машинного та глибокого навчання для прогнозування цінових коливань криптовалют;
- оцінити ефективність та порівняти прогностичну здатність розроблених моделей.

Теоретична, методична та практична значущість отриманих результатів. Отримані результати мають теоретичну, методичну та практичну важливість через актуальність досліджуваної тематики щодо розвитку криптовалют та їх впливу на соціальні, економічні та технологічні сфери у сучасному світі. У роботі надаються та обґрунтовуються висновки про ефективність застосування конкретних методів машинного та глибокого навчання у прогнозуванні динаміки ціни криптовалюти Bitcoin. Практична значимість роботи полягає у можливості застосування алгоритму розрахунку прогнозу цін криптовалют інвесторами при формуванні оптимального портфеля криптовалют.

Рік виконання кваліфікаційної магістерської роботи – 2023.

Рік захисту роботи – 2023.

Ключові слова: криптовалюта, прогнозування часового ряду, машинне навчання, глибоке навчання, випадковий ліс, багатошаровий перцептрон, індекс настрою.

В і д г у к
про кваліфікаційну магістерську роботу
здобувача навчально-наукового інституту
«Інститут інформаційних технологій в економіці»
освітньо-професійної програми «Економічна кібернетика і Дата Сайнс»
Каїки Олексія Павловича
на тему «Аналіз та прогнозування динаміки цін на криптовалютному ринку
методами машинного та глибокого навчання»

1. Актуальність теми дослідження обумовлена тим, що ринок криптовалют постійно зазнає значних коливань цін, і вирішення питань, пов'язаних з аналізом та передбаченням цих коливань, стає надзвичайно важливим завданням для інвесторів, трейдерів, фінансових аналітиків та економістів. А для України ця тема є особливо актуальною, оскільки наша держава є одним зі світових лідерів у впровадженні інформатизації в суспільство, зокрема і у фінансовій сфері.

2. Позитивні риси кваліфікаційної бакалаврської роботи: у роботі серед інших факторів, які зазвичай застосовують у подібних дослідженнях, автор застосував цікавий інтегральний показник рівня сентименту F&G (Fear & Greed), який вказує на загальний емоційний стан учасників ринку.

3. Наявність самостійних розробок автора: у роботі проведено ґрунтовний економічний аналіз впливу обраних факторів на динаміку ціни BTC. В процесі дослідження проведено порівняння прогностичних властивостей методів RF та MLP, на основі яких будувалися прогнози.

4. Цінність теоретичних висновків та практичних рекомендацій: у роботі автор показує, що методи машинного та глибокого навчання стали ключовими для прогнозування цін криптовалют, доповнюючи традиційні методи аналізу. Практичні рекомендації стосуються доцільності застосовувати ці методи як для короткострокового, так і для довгострокового прогнозування ціни криптовалюти.

5. Загальна оцінка кваліфікаційної магістерської роботи та її допущення до захисту перед ЕК: кваліфікаційна робота Каїки О.П. виконана на належному професійному рівні та свідчить про високий рівень сформованості компетентностей її автора, передбачених освітньою-професійною програмою «Економічна кібернетика і Дата Сайнс» другого (магістерського) рівня вищої освіти за спеціальністю «Економіка». Вважаю, що робота може бути допущена до захисту перед ЕК та заслуговує високої оцінки.

Науковий керівник: професор кафедри
математичного моделювання та статистики,
доктор фізико-математичних наук, професор
«___» _____ 2023 р.

О.М. Приতোманова

Рецензія
на кваліфікаційну магістерську роботу
здобувача вищої освіти
Каїки Олексія Павловича

Тема «Аналіз та прогнозування динаміки цін на криптовалютному ринку методами машинного та глибокого навчання»

Актуальність теми кваліфікаційної магістерської роботи і доцільність її розроблення. Дослідження ринку криптовалют є актуальним, оскільки Україна є одним зі світових лідерів у впровадженні інформатизації в суспільство та активно здійснює заходи з цифровізації ринку фінансових послуг.

Якість проведеного дослідження. У роботі проведено комплексний аналіз світових тенденцій у сфері криптовалют та технології блокчейну. Відзначені можливості, наявні на криптовалютних біржах для вітчизняного інвестора.

На основі цього аналізу автор обґрунтовує вибір криптовалюти та інформаційну базу для моделювання динаміки її ціни. Виконано ґрунтовний економічний аналіз впливу обраних факторів на динаміку ціни BTC. В процесі дослідження проведено порівняння прогностичних властивостей методів RF та MLP, на основі яких будувалися прогнози.

Позитивні риси кваліфікаційної магістерської роботи. Цікавою особливістю роботи є застосування серед факторів, що впливають на ціну BTC, показника рівня сентименту F&G (Fear & Greed), який вказує на загальний емоційний стан учасників ринку.

Зауваження. Серед зауважень можна відзначити обмеженість дослідження однією криптовалютою. Цікаво було б порівняти динаміку її ціни з іншими, можливо і не такими популярними, криптовалютами та вплив на них показника рівня сентименту F&G.

Практична значимість висновків і рекомендацій. Практична значимість роботи полягає у побудові алгоритму розрахунку прогнозу ціни криптовалюти.

Загалом, кваліфікаційна робота Каїки О.П. виконана на належному науковому і професійному рівні та свідчить про високий рівень компетентностей її автора, передбачених освітньою програмою другого рівня вищої освіти за спеціальністю «Економіка».

Рецензент:

к.т.н., доцент кафедри моделювання складних систем
Факультету комп'ютерних наук та кібернетики
Київського національного університету

ЗМІСТ

ВСТУП.....	3
РОЗДІЛ 1. ТЕОРЕТИЧНІ ТА МЕТОДИЧНІ ОСНОВИ ПРОГНОЗУВАННЯ ДИНАМІКИ ЦІН НА КРИПТОВАЛЮТНОМУ РИНКУ.....	5
1.1 Огляд літератури про прогнозування цін криптовалют.....	5
1.2 Структура ринків криптовалют та особливості їх функціонування.....	8
1.3 Математичні методи та інформаційні технології, що використовуються для побудови прогнозів на ринку криптовалют	15
РОЗДІЛ 2. ЗАСТОСУВАННЯ МЕТОДІВ МАШИННОГО ТА ГЛИБОКОГО НАВЧАННЯ ДЛЯ АНАЛІЗУ ЦІНОВИХ КОЛИВАНЬ КРИПТОВАЛЮТИ.....	34
2.1 Збирання та обробка історичних даних про ціни криптовалюти Bitcoin	34
2.2 Аналіз динаміки ціни криптовалюти	51
РОЗДІЛ 3. ЗАСТОСУВАННЯ МЕТОДІВ МАШИННОГО ТА ГЛИБОКОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ДИНАМІКИ ЦІН КРИПТОВАЛЮТ	64
3.1 Прогнозування часового ряду методами машинного навчання	64
3.2 Прогнозування часового ряду методами глибокого навчання	75
ВИСНОВКИ	82
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	84

ВСТУП

На тлі постійного розвитку сучасних технологій та динамічних змін у фінансовому секторі, криптовалютний ринок визначає нові перспективи для інвесторів, трейдерів та дослідників. Зростання популярності криптовалют та технології блокчейну свідчить про необхідність ретельного аналізу їхньої динаміки цін. Широкий спектр факторів, що впливають на цінові коливання, вимагає високоточних та ефективних методів аналізу та прогнозування. Аналіз та прогнозування динаміки цін на криптовалютному ринку методами машинного та глибокого навчання доповнює традиційні ринкові методи і створює можливості для покращення результатів дослідження. Дослідження криптовалютного ринку для України є актуальним, оскільки наша держава є одним зі світових лідерів у впровадженні інформатизації в суспільство, а також впродовж останніх років поступово звертає увагу на цифровізацію грошей, легалізацію та регуляцію криптовалют. Останні наукові дослідження та публікації у сфері аналізу криптовалютного ринку визначають перспективи використання методів машинного та глибокого навчання для досягнення більш точних прогнозів цін. Іноземні та українські науковці вже досягли певних успіхів у використанні цих методів, які в умовах сучасного світу впроваджуються для вирішення актуальних завдань у сфері фінансів, економіки, бізнесу, а також застосовуються для прогнозування динаміки на криптовалютному ринку.

Метою виконання даного дослідження є розробка та застосування ефективних моделей машинного та глибокого навчання для аналізу та прогнозування динаміки цін на криптовалютному ринку. Відповідно до поставленої мети визначені такі завдання:

- опрацювати джерела з аналізу та прогнозування цін на криптовалютному ринку;

- провести огляд криптовалютних бірж, інструментів та засобів для аналізу та прогнозування динаміки цін на криптовалютному ринку, відзначити правові аспекти для українських користувачів для обґрунтування актуальності обраної теми дослідження;
- охарактеризувати інформаційну базу для моделювання та провести графічний аналіз динаміки ціни обраної криптовалюти;
- провести аналіз факторів та їх впливу на ціну криптовалют;
- розробити та налаштувати моделі машинного та глибокого навчання для прогнозування цінових коливань криптовалют;
- оцінити ефективність та порівняти прогностичну здатність розроблених моделей.

Об'єктом дослідження є ціни криптовалют на криптовалютних біржах.

Предметом дослідження є методи машинного та глибокого навчання, які застосовуються для аналізу та прогнозування динаміки цін на криптовалютному ринку.

Для досягнення поставлених завдань використовуватимуться методи статистичного аналізу, машинного та глибокого навчання, а також методи обробки та аналізу великих обсягів даних. Отримані результати покликані поглибити розуміння факторів, що впливають на динаміку цін на криптовалютному ринку, та визначити ефективні методи їхнього аналізу та прогнозування. Це може виявити практичне застосування для інвесторів, трейдерів та регуляторів ринку.

Дослідження в роботі базується на актуальних наукових статтях, публікаціях у наукових журналах та Інтернет-сайтах, а також статистичних даних та інших інформаційних ресурсах, що стосуються як ринку криптовалют загалом, так і конкретних валют на ньому.

Робота складається зі вступу, трьох розділів: теоретико-методичний, аналітико-дослідницький, конструктивний; висновків, списку використаних джерел.

РОЗДІЛ 1

ТЕОРЕТИЧНІ ТА МЕТОДИЧНІ ОСНОВИ ПРОГНОЗУВАННЯ ДИНАМІКИ ЦІН НА КРИПТОВАЛЮТНОМУ РИНКУ

1.1 Огляд літератури про прогнозування цін криптовалют

Тематика аналізу та прогнозування цін криптовалют привертає увагу багатьох науковців, фінансистів та економістів в Україні та за кордоном.

У роботі [1] представлено глибокий аналіз теоретичних підходів і моделей, які використовуються для прогнозування ринку криптовалют. Автор висвітлює узагальнені переваги та недоліки структурних методів і моделей, що застосовуються у цьому контексті, зокрема нейромережових підходів. Основний фокус статті спрямований на порівняльний аналіз різних моделей штучних нейронних мереж (ШНМ), таких як CNN-2l (Convolutional Neural Network with 2 Layers), CNN-3l (Convolutional Neural Network with 3 Layers), LSTM (Long Short-Term Memory), sLSTM (Stacked Long Short-Term Memory), BiLSTM (Bidirectional Long Short-Term Memory), GRU (Gated Recurrent Unit), CLSTM (Convolutional Long Short-Term Memory), MLP (Multilayer Perceptron), RFBNN (Radial Basis Function Neural Network). Автор детально розглядає переваги та недоліки кожної з цих моделей, а також проводить тестування їхньої ефективності на реальних даних ринку.

Наукова цінність статті полягає в тому, що автор досліджує використання нейромережових методів для прогнозування ринку криптовалют. Результати тестування свідчать про високий рівень ефективності моделей LSTM і CLSTM, особливо враховуючи додавання макроекономічних показників. У висновках автор підкреслює потенційні переваги гібридних LSTM мереж, які об'єднують інформацію

про макроекономічні фактори з даними ринку криптовалют. Зазначається, що дослідження вказує на перспективи розвитку і вдосконалення інструментів для прогнозування ринку криптовалют, що може бути важливим в контексті стрімкого розвитку цього сектору фінансового ринку.

Стаття [2] спрямована на дослідження прогнозування курсу криптовалют, зокрема біткоїна, використовуючи вейвлет-аналіз та нейронні мережі LSTM. Автор використовує дискретне вейвлет-перетворення (ДВП) та його модифікацію (MODWT) для попередньої обробки фінансового часового ряду. Дослідник розглядає два види вейвлетів - Хаара і Добеші, а також MODWT максимального перекриття. Перед побудовою рекурентної мережі LSTM дані піддаються нормалізації та центруванню. Після підготовки даних створюється модель LSTM з певними параметрами, такими як кількість шарів, кількість блоків, оптимізатор, функція втрат і кількість епох.

Результати дослідження включають порівняння прогнозів, отриманих за допомогою LSTM без вейвлету та з використанням вейвлет-аналізу. Зокрема, досліджується вплив різних вейвлетів та рівнів декомпозиції на точність прогнозування. Для оцінки ефективності використовуються показники, такі як коефіцієнт детермінації, середня абсолютна похибка, відсоткова похибка, корінь з середньоквадратичної похибки та метрика Dynamic Time Warping (DTW).

Висновки дослідження підкреслюють ефективність використання вейвлет-аналізу для покращення точності прогнозування. Зазначається, що оптимальний результат досягається з використанням MODWT з рівнем декомпозиції 2. Також відзначається важливість правильного вибору вейвлет-перетворення та рівня декомпозиції для досягнення кращих результатів. Автор також вказує на можливості покращення моделі за допомогою врахування зовнішніх факторів у майбутніх дослідженнях.

Стаття [3] спрямована на вивчення ефективності моделей машинного навчання (ML), таких як Binary Auto Regressive Trees (BART), Multilayer Perceptron

(MLP) та Random Forest (RF), у прогнозуванні короткострокової динаміки цін на Bitcoin (BTC), Ethereum (ETH) та Ripple (XRP). Автори використовують часові ряди даних криптовалют для тренування та тестування моделей. Величина лагу визначається одним з ключових параметрів у дослідженні. Автори порівнюють результати прогнозів та класифікації напрямку змін цін, враховуючи переваги та недоліки кожної моделі. Зазначається, що алгоритми ML (зокрема, штучні нейронні мережі, C&RT та їх ансамблі) набагато менш чутливі до статистичних даних часового ряду, ніж класичні статистичні та економетричні методи.

Висновки показують, що моделі BART та MLP мають вищу точність прогнозування порівняно з RF.

У статті [4] досліджується перетин тенденцій криптовалюти та соціальних медіа, зосереджуючись на Біткоїні (BTC). Дослідження пропонує інтегровану модель, яка використовує історичні ціни, аналіз настроїв твітів і різні показники, такі як кількість твітів, підписки користувачів і статус перевірки, для прогнозування цін на BTC. Для аналізу настроїв використовується FinBERT, модель на основі BERT, тоді як для прогнозування ціни використовується Gated Recurrent Unit (GRU). Набір даних містить історичні ціни BTC від Yahoo Finance і твіти про криптовалюту в реальному часі від Kaggle. Настрої твітів позначаються за допомогою VADER, і дослідження вирішує класовий дисбаланс за допомогою збільшення кількості даних. Хід дослідження включає збирання і попередню обробку даних, прогнозування настроїв за допомогою FinBERT і прогнозування цін за допомогою GRU. Набір даних містить такі функції, як найвища та найнижча ціна, ціни відкриття та закриття, текст твітів, інформація, пов'язана з користувачами, і показники настроїв. Попередня обробка даних передбачає видалення нерелевантних твітів, усунення дисбалансу класів і розділення даних для навчання, тестування та перевірки.

Аналіз настрою на основі FinBERT досягає середньої абсолютної відсоткової помилки (MAPE) у 9,45%. Прогноз ціни на основі GRU дає кращий результат - 3,6%. Дослідження порівнює FinBERT з іншими алгоритмами прогнозування настрою

та представляє порівняльний аналіз продуктивності GRU. Тестування в реальному часі передбачає збирання твітів, що відповідають щоденним цінам поза набором даних.

Стаття демонструє доцільність використання соціальних мереж, зокрема Twitter, для прогнозування руху цін на BTC. Інтегрована модель демонструє багатообіцяючі результати, і в майбутньому аналіз настроїв може бути розширений, щоб включити в нього численні медіа-джерела, як-от головні новини, онлайн-статті та Reddit.

1.2 Структура ринків криптовалют та особливості їх функціонування

Криптовалюта є формою цифрової валюти, її емісія та управління відбувається через децентралізовану платіжну систему, яка працює автоматично без централізованого адміністрування [5]. Основна відмінність криптовалют полягає в тому, що їхня інформація зберігається в блокчейні, і для підтвердження транзакцій використовується асиметричне шифрування для перевірки повноважень виконувати її, а також інші криптографічні методи, такі як доказ виконаної роботи (Proof-of-Work) або доказ володіння активами (Proof-of-Stake). Тобто, це форма цифрових активів, заснована на мережі, яка розподілена між великою кількістю комп'ютерів [6].

Більшість криптовалют існують у децентралізованих мережах, які використовують технологію блокчейн — розподілений ланцюг, який підтримує розрізнена мережа комп'ютерів [7]. Визначальною особливістю криптовалют є те, що вони, як правило, не випускаються жодним центральним органом влади, що робить їх теоретично несприйнятливими до державного втручання чи маніпуляцій. Різноманітність та велика кількість криптовалют пояснюється різними технічними

характеристиками, функціональністю та цілями їхніх створювачів. Кожна криптовалюта може виконувати унікальні завдання та мати особливості, які роблять їх привабливими для конкретних груп користувачів або використання в певних галузях. Це варіювання може включати в себе різницю в алгоритмах консенсусу, обсязі блокчейну, швидкості транзакцій, приватності, розмірі блоку та інших технічних параметрах. Деякі криптовалюти спрямовані на конкретні галузі, такі як смарт-контракти, де Ethereum є одним із провідних гравців, або приватні транзакції, де Monero відомий своєю фокусованістю на анонімності. Інші криптовалюти можуть мати за мету полегшити швидкі та ефективні глобальні транзакції, як от Ripple. Таке розмаїття надає користувачам інструменти для вибору та використання тих криптовалют, які найкращим чином відповідають їхнім потребам та вимогам. Як наслідок, конкуренція та інновації в цьому просторі сприяють розвитку нових технологій та вдосконаленню існуючих блокчейн-платформ [8].

Криптовалюти піддаються значній волатильності [9], тобто мінливості своєї вартості. Це пояснюється рядом причин. Хоча криптовалюти вважаються способом досягти абсолютної анонімізації та нерегулювання, в останні роки державні та міжнародні інституції намагаються впливати на криптовалютний ринок законодавчими ініціативами та моніторингом [10]. Проте процеси, пов'язані з криптовалютами, а це торгівля, обмін, випуск тощо, повною мірою досі не регулюються юридичними особами та державними інституціями. Ціни на більшість криптовалют піддаються впливу за рахунок новин, прогнозів експертів та учасників ринку. До того ж, оскільки такі активи є відносно новим явищем в нашому світі, багато учасників ринку не розуміють ринкових та технологічних процесів, які стоять за криптовалютами, а отже, створюється атмосфера мінливості та зміни настроїв. Така природа криптовалют приваблює спекулятивних інвесторів [11], які хочуть отримати гроші на коливаннях цін. Оскільки жоден фізичний актив не підтверджує їх цінність, їхня ціна залежить в більшості випадків від віри. Це одна з головних

причин, чому ціни криптовалют настільки мінливі. Ціна знизиться, якщо віра гравців на ринку буде втрачена, і зросте, якщо криптовалюта набере популярності.

Технологічна складність [12] криптовалют та їх видобутку пов'язана з використанням розподіленої технології блокчейн. Криптовалюти, такі як Bitcoin, використовують криптографічні методи для забезпечення безпеки та консенсусу в мережі. Видобуток (майнінг) криптовалют включає в себе розв'язання складних математичних завдань, які підтверджують та обробляють транзакції. Для цього майнерам необхідно вирішити криптографічні головоломки, що вимагає значних обчислювальних ресурсів. Складність головоломок регулюється протоколом криптовалюти і змінюється періодично для забезпечення сталої швидкості блоків. Цей процес також контролює загальну кількість монет, які можуть бути видобуті, що часто робить їх децентралізованими та обмежує інфляцію. Технічна складність зумовлює необхідність великих обчислювальних потужностей для видобутку, а також впливає на вартість обладнання та споживання електроенергії [13]. Однак ця складність є важливою для забезпечення безпеки та стійкості криптовалютних мереж та операцій в середині них.

Інформація та технології пов'язані з криптовалютами доступні для використання та перегляду громадськістю. Це може охоплювати відкритий вихідний код блокчейн-протоколів, публічні баланси та історії транзакцій криптовалют, а також відкриті API для розробки додатків та сервісів. Крім того, криптовалютні біржі та гаманці доступні для будь-кого, хто бажає стати трейдером, криптоінвестором, або просто зберігати криптоактиви. Відкритий доступ сприяє прозорості, довірі та розвитку криптогалузі. Із кожним новим блоком блокчейну зростає обчислювальна потужність, необхідна майнеру для розрахунку всього ланцюжка з нуля, а чим довше ланцюг, тим важче «зламати» мережу. Нині біткоїн — це децентралізована обчислювальна мережа, продуктивність якої більш ніж у вісім разів перевищує сумарну обчислювальну потужність усіх суперкомп'ютерів у світі. Для того щоб мати над нею навіть обмежений контроль, потрібні величезні ресурси та витрати в

сотні мільйонів доларів. Проте, використання криптовалюти час від часу піддається хакерським атакам [14], що викликає недовіру у потенційних користувачів та не спонукає інвесторів вкладати значні кошти в цифрові активи.

Основним дискусійним моментом у питанні контролю і регулювання обігу криптовалюти у світі є її децентралізованість [15]. Перевагою децентралізації є те, що за її наявності легше впровадити певні стандарти випуску валюти та її розповсюдження. Проте, децентралізація надає широкий доступ до контролю персональних даних користувачів криптовалют, контролю їхніх грошових потоків, а відповідно й до технологічної небезпеки та руйнування світової економіки. Високорозвинені країни прагнуть врегулювати електронні платежі, взявши їх під контроль держави, обклавши податками, та зменшити ризики в процесі її застосування. Держави зі слабкою економікою не готові до впровадження подібних платіжних систем. Тому більшість урядів вибрали лінію нейтралітету, уникаючи будь-яких рішень щодо віртуальної валюти. Ставлення країн до регуляції криптовалют різняться і можуть коливатися від позитивного прийняття до обмежень чи повної заборони [16]. Ось деякі приклади ставлення деяких країн:

Японія визнає криптовалюту як законний засіб платежу. Країна ввела регуляторні механізми для обмінників криптовалют, щоб забезпечити безпеку користувачів. Сполучені Штати зазвичай розглядають криптовалюту як власність, а не валюту. А регулювання значно варіюється від штату до штату. Китай вважає криптовалюту загрозою своїй економіці. В країні заборонені Initial Coin Offerings (ICO), обмін криптовалют та деякі аспекти майнінгу. Індія розглядає можливість створення власної цифрової валюти, але виражає обурення стосовно використання криптовалют для незаконних операцій. Швейцарія відома своєю криптовалютною дружбою, ставленням до криптовалют і блокчейн-проектів як до інновацій. Що стосується країн ЄС, то вони мають різні підходи до регулювання, але загалом дивляться на блокчейн-технології як на потенційну інновацію.

Різні криптовалюти можуть мати різний характер ліквідності, що зумовлюється такими факторами, як розмір ринку, популярність активу, кількість угод на ринку, якість того чи іншого активу та його перспективи та баланс між покупцями та продавцями. Найбільша ліквідність на криптовалютному ринку міститься в біткоїні, так можна відразу набирати актив на сотні тисяч або навіть на мільйони доларів, не чинячи істотного впливу на ціну. На другому місці — Ethereum (ETH). Далі йдуть такі активи як LTC, XRP, XLM, ATOM, TRX, ETC тощо. Свою нішу займають стейблкоїни [17], що функціонують із прив'язанням до певної фіатної криптовалюти, найчастіше – американського долара. Найбільш популярні — USDT, USDC. Стейблкоїни також вважаються криптовалютами, і в найширшому розумінні вони мають значно більшу ліквідність, але мають відносно стабільну вартість, що не дає змоги розглядати їх як прибуткові активи. В сучасному світі стейблкоїни можуть використовуватися як засіб зберегти гроші в віртуальних американських доларах, не використовуючи банки та готівку. Середня ліквідність є в старих проєктах, які торгуються на різних біржах, але не на всіх, наприклад, DASH, IOTA, DOGE, LINK, EOS та інші, що входять у топ 50-100 за капіталізацією. Найменша ліквідність у криптовалют, які торгуються на одній або декількох біржах, у рейтингу за капіталізацією 100+ і вище, їх тисячі. Сюди ж можна віднести криптопроекти-меми, фантокени тощо. Такі активи вельми ризиковано розглядати як інвестування на значущі суми, а ось для коротких спекуляцій і на невеликі суми вони можуть бути актуальними для гравців ринку, що можуть працювати з ризиком втрат.

Індустрія криптовалют швидко розвивається, тому з'являється безліч нових проєктів, для розвитку яких розробники залучають зовнішнє фінансування. Для цього вони здійснюють емісію своєї криптовалюти (створюють криптовалюту) та влаштовують токенсейл. Токенсейл - це процес продажу нової криптовалюти перед її лістингом на біржі. Ранні інвестори отримують цю криптовалюту від команди за найнижчою ціною, фінансуючи проєкт та його вихід на ринок. Поширені форми токенсейлів включають ICO (первинна пропозиція монет) [18], IEO (первинна

пропозиція на централізованій біржі) та IDO (первинна пропозиція на децентралізованій біржі). ICO стали популярними, але з часом втратили популярність через зростання шахрайських проєктів. Згодом почали з'являтися IEO, проведені на біржах, як їхній заміник. Але попри високі ризики для інвесторів, IDO, які відбуваються на децентралізованих біржах, стали зручним методом для проєктів залучати кошти.

Особливості функціонування криптовалютного ринку в Україні:

Дослідження криптовалютного ринку для України є актуальним, оскільки наша держава є одним зі світових лідерів у впровадженні інформатизації та технологій у суспільство, а також впродовж останніх років поступово звертає увагу на цифровізацію грошей, легалізацію та регуляцію криптовалют.

В Україні на криптовалютному ринку здійснюються операції купівлі, продажу, обміну та конвертації криптовалют. В Україні функціонують криптовалютні біржі, такі як KUNA, Binance, WhiteBit, Exmo та BTC TRADE UA, кожна з яких пропонує різні можливості вибору валют для торгівлі, різниця також присутня і в комісіях, швидкості реєстрації, інтерфейсу тощо. Наприклад, біржа Exmo проводить транзакції в доларах США, євро, польських злотих, тоді як KUNA та BTC TRADE UA спеціалізуються на гривневих транзакціях. Українські регулятори намагаються вирішити правовий статус криптовалют та регулювання операцій з ними, але відсутність єдиного підходу на світовому рівні ускладнює цей процес. Метою регулювання виступає захист прав споживачів, протидія відмиванню грошей та оподаткування отриманих доходів. Однак глобальна практика показує, що вироблення єдиного стандарту у цьому питанні є складним завданням, оскільки країни визначають власний підхід до регулювання криптовалют.

Законодавство, що повністю регулювало б ліцензування, податковий контроль та інші правові аспекти регулювання криптовалютного ринку ще не визначено повною мірою. Деякі ініціативи та процеси зі створення законопроектів вже

проводяться [19]. До прикладу, на початку 2020 року в Україні прийнято закон "Про валюту та валютні операції". У 2022 році був прийнятий закон Про віртуальні активи, де вказано, що віртуальні активи не є засобом платежу на території України та не можуть бути предметом обміну на майно (товари), роботи (послуги). У той же час віртуальні активи, до яких відносять і криптовалюти, визначені як нематеріальне благо, що є об'єктом цивільних прав, має вартість і виражене сукупністю даних в електронній формі. Також було опубліковано законопроект "Про внесення змін до Податкового кодексу України та інших законодавчих актів України щодо врегулювання обороту віртуальних активів в Україні", в якому визначається правовий статус віртуальних активів, а також розглядається введення ліцензій на здійснення валютних операцій. Отже, ліцензія може стати обов'язковою для діяльності юридичних осіб на ринку криптовалют. Станом на дату написання роботи, наразі законодавчими органами розглядаються податкові обов'язки для українських користувачів криптовалют, що включає оподаткування доходів від торгівлі криптовалютами. Національний банк України, законодавчі та правоохоронні органи прагнуть знайти підходи до взаємодії з криптовалютним ринком для забезпечення контролю та безпеки. Діяльність на нерегульованому криптовалютному ринку пов'язана з суттєвими ризиками, такими як втрата грошей внаслідок кібератак, діяльність недобросовісних посередників, шахрайство, відсутність інвестиційних гарантій та високий рівень волатильності ринку. Відсутність державного регулювання та недостатній розвиток інфраструктури також є серйозними проблемами для функціонування криптовалютного ринку в Україні.

1.3 Математичні методи та інформаційні технології, що використовуються для побудови прогнозів на ринку криптовалют

Для прогнозування цінових коливань криптовалют дослідники, трейдери та інвестори використовують різноманітні підходи [20] та методи, тому що завдання, які стоять перед ними, можуть варіюватися впродовж часу залежно від конкретного становища на криптовалютному ринку, а також очікуваних та неочікуваних змін. Розглянемо існуючі теорії і моделі для прогнозування цін криптовалют.

Аналіз технічних індикаторів

Аналіз технічних індикаторів є одним з ключових методів аналізу ринку в торгівлі фінансовими інструментами, включаючи криптовалюти [21]. Технічний аналіз базується на ідеях про те, що ціни вже враховують всю інформацію, доступну на ринку, і що патерни цін та інші технічні показники можуть надати інформацію про те, яким чином ціни можуть змінитися в майбутньому.

Ключові технічні індикатори та їхні основні концепції включають ряд показників та індикаторів. Середня рухома (Moving Average) є індикатором, що визначає середнє значення ціни за певний період часу. Є різні види середніх рухомих, такі як проста (SMA) та експоненційна (EMA) [22], що використовуються з різною метою. RSI (Індекс відносної сили) вимірює силу та інтенсивність змін цін для визначення перекупленості чи перепроданості актива. RSI має значення від 0 до 100. MACD (Розбіжність конвергенції ковзної середньої) показує відмінність між експоненційною середньою рухомою на коротший та довший періоди часу. Цей показник може слугувати сигналом для визначення змін тренду. Для вимірювання волатильності активу на ринку використовуються лінії Болінджера. Вони складаються з трьох ліній: середньої смуги, верхньої смуги (верхня межа волатильності) та нижньої смуги (нижня межа волатильності). У технічному аналізі

також використовується Технічний індекс Сили. Цей індикатор вказує на силу та напрямок тренду ринку. Чим вищий він, тим сильніший тренд.

У контексті криптовалют, технічний аналіз часто використовується для прогнозування майбутніх цін. Трейдери та аналітики вивчають графіки цін, намагаючись виявити патерни та тренди, що можуть вказувати на можливі зміни в цінах. Крім того, вони використовують технічні індикатори для отримання сигналів покупки або продажу.

Аналіз технічних індикаторів для прогнозування цін криптовалют має свої переваги та недоліки, які важливо враховувати при використанні цього методу аналізу. Технічні індикатори зазвичай досить прості для розуміння та застосування, навіть для тих, хто тільки вчиться торгувати. Вони часто ефективно працюють при короткостроковому аналізі, що може бути важливо для денних трейдерів та інших, хто шукає можливості для швидких операцій з криптовалютами та іншими активами. Технічні індикатори можуть допомагати виявляти тренди та патерни цін, що може бути корисним для прийняття рішень щодо входу або виходу з ринку. Технічні індикатори легко вбудовуються в автоматизовані торгові стратегії, що дозволяє трейдерам автоматизувати процес прийняття рішень. Проте, технічні індикатори мають залежність від минулих даних, тобто технічний аналіз базується на історичних даних, індикатори розраховуються на основі минулих цін. Це може призвести до того, що аналіз може бути менш ефективним в умовах ринку, які суттєво відрізняються від минулих періодів. Також, технічний аналіз ігнорує фундаментальні фактори, такі як новини, події на ринку або фінансовий стан компанії, що може призвести до втрати важливої інформації. Іноді, виникають проблеми з тим, що трейдери можуть використовувати занадто багато технічних індикаторів або аналізувати діаграми занадто детально, що може вести до аналізу "перетримки", де індикатори можуть показувати суперечливі сигнали. Варто також зазначити, що технічний аналіз не завжди ефективний в умовах великих та непередбачуваних змін на ринку, таких як геополітичні події або інші несподівані фактори. Технічний аналіз

може бути менш ефективним для ринків з низькою ліквідністю чи відсутністю чіткої трендової динаміки.

Аналіз фундаментальних факторів

Фундаментальний аналіз в криптовалютному ринку [23] вивчає основні фактори, які впливають на ціну криптовалют та їхній ринковий потенціал. Відмінною особливістю фундаментального аналізу від технічного є те, що він зосереджений на оцінці реальної вартості актива, використовуючи економічні, фінансові та технічні показники.

Фундаментальний аналіз криптовалют включає в себе комплексну оцінку різноманітних аспектів для формування повної карти, необхідної для прийняття обґрунтованих інвестиційних рішень. Один із ключових факторів – це якість команди та стадія розвитку проєкту, де аналізується не лише досвід та експертиза команди, але й її стратегічний план та потенціал для майбутнього росту. Технологічна інноваційність також аналізується і є важливим аспектом, оскільки вона визначає технічні характеристики криптовалюти та її інноваційний вплив на сферу блокчейн-технологій. Детальний аналіз партнерств та співпраці з іншими компаніями допомагає зрозуміти можливі взаємовигідні зв'язки, які можуть впливати на успіх проєкту. Ліквідність і обсяг торгів дозволяють визначити рівень активності на ринку та його вартість. Під час дослідження аналізується регулятивний контекст, оскільки він впливає на правове середовище криптовалют та може визначати їхню придатність для регульованого фінансового середовища. Аналіз економічних та фінансових показників зумовлює оцінку фінансової стійкості проєкту та його здатності до довгострокового розвитку. Рівень прийняття та використання криптовалюти в реальному світі служить індикатором її практичного застосування. Нарешті, слід враховувати новини та події, оскільки вони можуть вплинути на ринкову динаміку і спричинити значні зміни в цінності криптовалют. Ретельне вивчення всіх цих аспектів допомагає інвесторам уникнути ризиків та здійснювати

свої рішення на основі об'єктивних даних. Фундаментальний аналіз надає трейдерам більше контексту для прийняття рішень. Наприклад, якщо команда проєкту має сильний досвід і розвиток технології є інноваційним, це може свідчити про потенціал росту ціни криптовалюти в майбутньому.

Аналіз фундаментальних індикаторів для прогнозування цін криптовалют також має свої переваги та недоліки, і їх важливо враховувати при використанні цього методу аналізу. Фундаментальний аналіз особливо корисний для інвесторів, які планують тримати актив довгий час і спрогнозувати його потенційну вартість на основі фундаментальних факторів. Даний різновид аналізу дозволяє трейдерам та інвесторам оцінювати реальну вартість активу, використовуючи різноманітні фінансові та економічні показники. Використання фундаментального аналізу дозволяє враховувати глобальні економічні та політичні події, які можуть вплинути на ринок криптовалют. Для криптовалют, які представляють блокчейн-проєкти, фундаментальний аналіз дозволяє докладно оцінювати технічний потенціал, команду, партнерство та інші фундаментальні аспекти. Фундаментальний аналіз дозволяє аналізувати фінансовий стан та стійкість проєкту, що є важливим при оцінці ризиків. Проте, застосування фундаментального аналізу має свої недоліки, наприклад, те, що він часто вимагає більш тривалого часу для проведення дослідження та збору необхідної інформації порівняно з технічним аналізом. Також, оцінка фундаментальних факторів може бути суб'єктивною і залежати від тлумачення інвестора чи аналітика. Фундаментальний аналіз не завжди бере до уваги технічні показники та тенденції, які можуть бути важливими для короткострокових трейдерів, а інформація, яку використовують для фундаментального аналізу, може бути невірно інтерпретована або втрачати актуальність через час.

Соціальний аналіз та відслідковування новин

Соціальний аналіз [27] та відслідковування новин криптовалютного ринку включають моніторинг та аналіз активності та відгуків у соціальних мережах,

форумах, блогах, та новинних сайтах. Цей підхід може допомогти виміряти сентимент (настрій) громадськості, виявляти тренди та прогнозувати ринкову активність.

Соціальний аналіз має особливе застосування в прогнозуванні цін криптовалют. Оскільки ціни криптовалют можуть змінюватися залежно від емоцій гравців на ринку, визначення та врахування настроїв учасників ринку щодо конкретної криптовалюти може покращувати якість прогнозів. Позитивні відгуки та події можуть призвести до зростання цін, тоді як негативний може викликати спад. В одному з досліджень сентиментного аналізу [28], застосовуючи методи машинного навчання, використовуються Python бібліотеки, створені саме з цією метою. Моніторинг новин та подій використовується криптосередовищі для передбачення впливу цих подій на ціни криптовалют через новизну ринку, певну недовіру у нього, яка може проявлятися у нестандартних та непосліовних рішеннях. Також соціальний аналіз може виявити аномальні ситуації, такі як широкомасштабні маніпуляції чи атаки на блокчейн проекти.

Методи соціального аналізу виявляють свої переваги у контексті аналізу криптовалютного ринку. Зокрема, вони дозволяють отримати інформацію з перших рук, враховуючи відгуки та обговорення як надійні джерела, що можуть перевершувати традиційні новини. Крім того, соціальний аналіз сприяє врахуванню громадської думки та реакції громадськості на ключові події, що важливо для розуміння суспільного сприйняття криптовалют.

З іншого боку, соціальний аналіз також має свої недоліки. Інформація може бути неоднорідною та піддаватися впливу штучних змін, що може впливати на точність аналізу. Існує ризик маніпулювання громадським настроєм для створення штучного попиту чи пропозиції на ринку. Також, часто буває важко визначити причинно-наслідкові зв'язки між змінами в громадському настрої та конкретними ціновими змінами. Крім того, спрогнозування непередбачуваних подій чи ринкових реакцій може бути важким завданням у рамках соціального аналізу.

Аналіз волатильності і ризику

Аналіз волатильності та ризику в контексті фінансів оцінює коливання цін, а також можливість збитків або непередбачуваних змін на ринку. Він використовується для оцінки ступеня небезпеки і непередбачуваності в цінових рухах.

Ключові аспекти в аналізі волатильності та ризику - це:

- волатильність, тобто міра коливань цін, яка вказує на рівень ризику та нестабільності валюти. Волатильність розраховується за формулою:

$$\text{Волатильність} = \sqrt{\left(\frac{1}{N-1} \sum_{i=1}^N (\ln(P_i) - \ln(P_{i-1}))^2 \right)}, \quad (1.1)$$

де N - кількість спостережень (періодів часу);

P_i - ціна криптовалюти в i -му періоді;

$\ln$ – натуральний логарифм.

- бета-коефіцієнт [29] - визначає ступінь залежності активу від ринкових змін. Бета-коефіцієнт розраховується за формулою:

$$\beta = \frac{\text{cov}(R_{\text{crypto}}, R_{\text{market}})}{\text{var}(R_{\text{market}})}, \quad (1.2)$$

де β - бета-коефіцієнт;

$\text{cov}(R_{\text{crypto}}, R_{\text{market}})$ - коваріація між дохідністю криптовалюти та ринковою дохідністю;

$\text{var}(R_{\text{market}})$ - дисперсія ринкової дохідності.

- управління ризиками [30] - допомагає інвесторам та трейдерам ефективно керувати ризиками, приймаючи рішення на основі очікуваної волатильності;

- оцінка ефективності стратегій, тобто аналіз того, наскільки добре різні стратегії працюють у змінливих умовах ринку;
- прогнозування небезпечних зон, тобто вказування на можливі періоди високого ризику, що дозволяє інвесторам уникати діяльності в певний момент або навпаки - приймати необхідні рішення.

Переваги аналізу волатильності і ризику включають можливість зменшення ризику втрат для інвесторів, дозволяючи їм своєчасно реагувати на зміни та зменшувати можливі збитки. Крім того, цей аналіз сприяє підвищенню ефективності управління портфелем, допомагаючи розподіляти активи так, щоб мінімізувати загальний ризик портфеля, а також покращує прогнозування, забезпечуючи краще розуміння ринкових умов. З іншого боку, аналіз волатильності і ризику має свої недоліки. Наприклад, він не завжди може точно передбачити високі рівні волатильності. Результати також можуть змінюватися в залежності від змін у ринкових умовах. Поганий прогноз волатильності, спростованість в моделях та непередбачувані події можуть впливати на точність аналізу. Таким чином, застосування аналізу волатильності та ризику важливо супроводжувати усвідомленим підходом до ризиків та обмежень моделі.

Машинне навчання та штучний інтелект

Із розвитком як криптовалютного ринку, так і інформаційних технологій, штучний інтелект (ШІ) і машинне навчання (МН) поступово починають відігравати все більшу роль у прогнозуванні цін криптовалют [24, 25]. Характер сфер застосування машинного навчання і штучного інтелекту для прогнозування криптовалют включає різні аспекти ринку та аналізу, де ці технології можуть бути ефективно використані. машинне навчання та штучний інтелект відіграють ключову роль в аналізі та прогнозуванні динаміки цін на криптовалютному ринку в різних сферах застосування. По-перше, в області технічного аналізу і виявлення патернів МН та ШІ можуть аналізувати обширні обсяги історичних даних цін, визначаючи

технічні патерни для формування прогнозів. Інший аспект стосується прогнозування цінових трендів, де алгоритми МН можуть аналізувати часові ряди цін, допомагаючи трейдерам та інвесторам приймати обґрунтовані рішення щодо входу або виходу з ринку. Машинне навчання також використовується для оцінки ризиків на криптовалютному ринку, враховуючи фактори, такі як волатильність, ліквідність та новини. Сентиментний аналіз за допомогою ШІ дозволяє визначити реакцію гравців на ринку на події та новини, впливаючи на ринковий настрій. Машинне навчання також допомагає під час розробки та оптимізації торгових стратегій, автоматизуючи процеси прийняття рішень. Важливою областю є аналіз ринкової ліквідності, де алгоритми МН можуть допомагати прогнозувати рівень ліквідності на ринку. Завдяки машинному навчанню можна виявляти аномалії або незвичайні ринкові поведінки, що може слугувати сигналом про можливі зміни в цінах. Моделі МН також можуть використовуватися для прогнозування обсягів торгів, що є важливим для оцінки активності ринку. Усі ці аспекти допомагають покращити аналітику та прийняття рішень на криптовалютному ринку. Ці технології дозволяють ефективно обробляти величезні обсяги даних, розпізнавати складні взаємозв'язки та забезпечувати трейдерів та інвесторів цінною інформацією для прийняття своїх рішень на криптовалютному ринку.

Машинне навчання (МН) та штучний інтелект (ШІ) відіграють значущу роль у прогнозуванні цін криптовалют, але вони також мають свої переваги та недоліки [26], які важливо враховувати при побудові прогностичних моделей. МН та ШІ приносять численні переваги в аналізі криптовалютного ринку. Вони відзначаються адаптивністю до змін в ринкових умовах, здатністю ефективно опрацьовувати великі обсяги даних та автоматизувати процеси прийняття рішень. Алгоритми МН можуть виявляти складні зв'язки та навіть слабкі патерни та тренди на криптовалютному ринку, що є важливим для трейдерів та інвесторів. Однак використання машинного навчання та штучного інтелекту може мати ряд перешкод та недоліків. Залежність від якості вхідних даних може призвести до неточностей у прогнозах, а багато моделей

МН можуть ігнорувати фундаментальні фактори. Також слід враховувати схильність до перенавчання, що може впливати на універсальність моделей. Деякі моделі МН, особливо глибокі нейронні мережі, можуть бути важко інтерпретованими, ускладнюючи пояснення прийняття рішень. Найважливіше, МН та ШІ можуть бути неспроможними передбачити непередбачувані події чи великі зміни на ринку, такі як глобальні кризи чи зміни в регулюванні, тобто розроблені моделі необхідно перенавчати, враховуючі нові обставини на ринку криптовалют.

Найбільш розповсюдженим в контексті прогнозування цінової динаміки криптовалют є методи прогнозування часових рядів. Методи прогнозування часових рядів використовують історичні дані про ціни криптовалют для розбудови моделей прогнозування майбутніх цінових траєкторій. Вони базуються на припущенні, що минулі цінові патерни і зміни можуть надати інформацію для передбачення майбутніх рухів. Прикладами таких методів та алгоритмів можуть слугувати ARIMA (Autoregressive Integrated Moving Average), що використовує попередні значення та їхні відмінності для побудови моделі, та LSTM (Long Short-Term Memory) [31] - рекурентні нейронні мережі, спроектовані для роботи з послідовностями, включаючи часові ряди. Ці методи є ефективними для виявлення цінових патернів, вони добре пристосовуються до врахування часових залежностей. Проте, варто зазначити, що дані методи можуть бути чутливі до великої кількості шумів, а покладаючись лише на них може бути важко врахувати несподівані події, що в контексті криптовалютного ринку є дуже важливим, але складним завданням.

Важливим етапом перед застосуванням моделей для прогнозування часових рядів є перевірка стаціонарності. Часовий ряд вважається стаціонарним, якщо його статистичні властивості, такі як середнє та дисперсія, залишаються сталими з часом. Якщо часовий ряд не є стаціонарним, можуть виникнути проблеми при визначенні параметрів моделі та прогнозуванні. Багато методів аналізу та моделювання часових рядів працюють краще зі стаціонарними даними. Існує кілька тестів на

стаціонарність, зокрема Тест Дікі-Фуллера (ADF), що перевіряє наявність одного або більше елементів, що забезпечують нелінійний тренд у даних.

Формула тесту:

$$\Delta y_t = \rho y_{t-1} + \alpha + \beta t + \gamma D_t + \xi_t, \quad (1.3)$$

де Δy_t - перша різниця часового ряду;

y_{t-1} - значення ряду на попередньому кроці часу;

t - часовий індекс

D_t - дампінг-фактор, який дорівнює 0, якщо t менше або рівне критичного значення, і дорівнює 1, якщо t більше критичного значення;

$\rho, \alpha, \beta, \gamma$ - параметри моделі;

ξ_t - білий шум.

Гіпотеза тесту полягає в тому, що $\rho = 1$ (є одиничний корінь) проти альтернативної гіпотези $\rho < 1$ (немає одиничного кореня, тобто ряд є стаціонарним). Статистика тесту порівнюється з критичними значеннями для визначення статистичної значущості результатів тесту. Якщо статистика менше критичного значення, то гіпотеза про наявність одиничного кореня відхиляється, що свідчить про стаціонарність ряду. Характерними ознаками стаціонарного часового ряду є постійність середнього значення та дисперсії, тобто вони залишаються постійними для всіх часових періодів. Отже в стаціонарному часовому ряді не має бути трендів або систематичних змін в середньому значенні або дисперсії з часом.

В аналізі часового ряду важливим етапом є дослідження автокореляції та часткової автокореляції. Автокореляція - це міра ступеня кореляції між значеннями змінної та її власними попередніми значеннями у часовому ряді. Іншими словами, автокореляція у часовому ряді вказує, наскільки кожен елемент ряду корелює з

попередніми значеннями ряду в той самий часовий момент. Коефіцієнти автокореляції використовуються для вимірювання ступеня кореляції між значеннями одного і того ж часового ряду на різних часових затримках. Обчислення коефіцієнтів автокореляції залежить від конкретного методу, але основна ідея полягає в обчисленні кореляції між часовим рядом та його власними лаговими значеннями. Одним з тестів на дослідження автокореляції є тест Льюнга-Бокса. Під час тесту Льюнга-Бокса тестова статистика обчислюється як сума коефіцієнтів автокореляції залишків на різних затримках, піднесених до квадрата. Формула тесту виглядає наступним чином:

$$(1.4) \quad Q = n(n+2) \sum_{k=1}^h \frac{r_k^2}{n-k},$$

де Q - значення тестової статистики;

n - кількість спостережень в часовому ряді;

h - кількість лагів для яких обчислюється автокореляція (зазвичай визначається емпірично);

r_k - значення автокореляційної функції для лагу k .

Отримана тестова статистика порівнюється з критичним значенням з розподілу хі-квадрат. Критичне значення визначається на основі рівня значущості та ступеня свободи. Якщо тестова статистика перевищує критичне значення, то нульова гіпотеза відкидається, і це може свідчити про наявність статистично значущої автокореляції в залишках моделі, що може вказувати на неадекватність моделі врахувати всі важливі патерни в даних.

Часткова автокореляція - це міра лінійної залежності між значеннями часового ряду на різних часових затримках, враховуючи вплив проміжних лагів. Іншими словами, часткова автокореляція відображає взаємозв'язок між двома точками в

часовому ряді, враховуючи внесок усіх проміжних затримок. Для обчислення часткової автокореляції використовується аналіз регресії, а саме метод часткової регресії. Цей метод вимагає врахування впливу попередніх лагів, щоб знайти чистий внесок конкретного лагу в автокореляцію. Це дозволяє визначити, наскільки великою є автокореляція між двома точками після усунення впливу проміжних точок. Часткова автокореляція зазвичай використовується для визначення оптимального порядку авторегресійного (AR) компонента в моделях часових рядів.

Важливе місце у розумінні та прогнозуванні динаміки цін посідає класифікація цінових тенденцій. Методи класифікації тенденції або тренду використовують алгоритми для визначення того, чи буде ринок рости, падати або залишатися стабільним. Вони використовують різні показники та характеристики наявних даних для класифікації цінових тенденцій. Прикладами методів та алгоритмів класифікації тенденції є метод опорних векторів, що використовується для вирішення задач класифікації, де модель навчається розділяти дані на два класи, а також методи дерев рішень та випадкового лісу, що можуть бути корисними для прогнозування напрямку руху ціни. Перевагами методів класифікації тенденцій є здатність прогнозувати напрямок ринку, а також вони добре працюють на основі різноманітних характеристик(екзогенних факторів). Серед недоліків методів класифікації тенденцій є те, що вони можуть бути чутливі до періодів з низькою волатильністю і не завжди ефективні при великому обсязі даних.

Для визначення можливих екзогенних факторів, що можуть мати вплив на цінову динаміку, є аналіз текстових даних із великої кількості онлайн-ресурсів. Методи аналізу текстових даних використовують обробку природної мови (NLP) для аналізу новин, соціальних мереж, результатів опитувань та інших текстових джерел для розуміння впливу інформації, що супроводжує ринкові події, на ціни криптовалют. Основні методи аналізу текстових даних включають зазначений раніше sentimentний аналіз та групування тематик. Ці методи допомагають визначати тон та настрій тексту (позитивний, негативний, нейтральний) або іншої інформації, яку

можна виміряти в контексті криптовалют, що розглядаються, окремих компаній та проектів, або ринку загалом. Дані методи допомагають виявляти вплив новин та громадської думки, але є вразливими, наприклад, до сарказму та складних висловлювань, а також їх аналіз може бути складним завданням великі обсяги неструктурованих даних з різних джерел.

Більш повний, науково-дослідницький підхід до аналізу ринку криптовалют включає в себе використання методу кластерного аналізу, який дозволяє систематизувати та групувати схожі дані для отримання важливої інформації про структуру ринку. Цей метод визначає групи активів, що мають спільні характеристики, що робить його корисним інструментом для ідентифікації подібних активів та розуміння їх взаємодії. Один із ключових методів кластерного аналізу - K-Means, дозволяє розділити дані на K кластерів, призначаючи кожен об'єкт кластеру з найближчим середнім. Цей підхід створює чітку структуру, де об'єкти в межах кластерів подібні один до одного, але відрізняються від об'єктів інших кластерів. У підходах кластерного аналізу також може використовуватися ієрархічний кластерний аналіз, що будує структуру кластерів у вигляді ієрархії, де об'єкти, в даному випадку, можливі групи криптоактивів, об'єднуються на основі їхньої подібності. Це дозволяє визначати не лише самі кластери, але й масштаб взаємозв'язків між ними.

Аналіз ринку криптовалют за допомогою методу кластерного аналізу має численні переваги. Він допомагає виявляти групи активів зі спільними характеристиками, що може бути важливим для розробки стратегій диверсифікації портфеля. Однак, важливо враховувати його недоліки, такі як залежність від вибору кількості кластерів, чутливість до початкових умов та необхідність в глибоких знаннях щодо стану та тенденцій криптовалютного ринку. Такий аналіз необхідний для глибокого розуміння динаміки ринку, де велика кількість активів може бути ефективно структурована та вивчена, щоб приймати обґрунтовані рішення у сфері управління активами.

Інструменти та програмне забезпечення, що використовуються для прогнозування цін криптовалют

Для аналізу криптовалютного ринку та динаміки цін використовується спеціалізоване програмне забезпечення. В сучасному світі постійно створюються та розвиваються продукти, які допомагають інвесторам та трейдерам приймати рішення щодо взаємодії з тим чи іншим ринком, або певною криптовалютою. Серед таких інструментів виділяють:

1. Торгівельні платформи і біржі [33]

Торгівельні платформи та криптовалютні біржі є важливими складовими інфраструктури криптовалютного ринку. Вони виступають як основні середовища для торгівлі різними криптовалютами, забезпечуючи учасникам можливість купувати та продавати активи. Наприклад, Binance, одна з найбільших та найпопулярніших криптовалютних бірж у світі, пропонує широкий вибір торгових пар та високу ліквідність. Coinbase, яка базується у Сполучених Штатах, слугує важливим сервісом для новачків, надаючи зручний інтерфейс та широкий вибір підтримуваних активів. Kraken відзначається своєю безпекою та різноманіттям торгових пар, а Bitstamp, яка є однією з найстаріших криптовалютних бірж у Європі, славиться своєю довіреністю та надійністю. Ці та інші торговельні платформи не тільки дозволяють виконувати торгівлю, але й надають широкий спектр інструментів для аналізу ринку, таких як графіки цін впродовж обраних часових проміжків, статистика та інші аналітичні засоби. Трейдери можуть використовувати ці інструменти для ретельного вивчення рухів цін та обсягів, а деякі платформи також пропонують API для автоматизації торгівлі та створення власних стратегій. Здійснення торгівлі на цих платформах дозволяє використовувати різні типи ордерів та торгові пари, що створює різноманітні можливості для трейдерів та інвесторів у взаємодії з криптовалютним ринком.

2. Інструменти для фінансового аналізу

Для фінансового аналізу використовуються спеціалізовані платформи, як наприклад, Bloomberg Terminal – інструмент для аналізу фінансових ринків, включаючи криптовалютні дані та новини. Bloomberg Terminal є комплексним інструментом, що об'єднує в собі різноманітні фінансові дані та новини, зокрема в галузі криптовалют. Він створений для проведення глибокого аналізу фінансових ринків та надає користувачам можливість взаємодії з широким спектром інформації.

Основна сила Bloomberg Terminal полягає в його здатності надавати відомості, недоступні на більшості інших платформ. Відстеження ринкових показників, аналіз новин, прогнозування тенденцій - усе це забезпечується в одному інтегрованому середовищі. Однак, слід відзначити, що вартість Bloomberg Terminal досить значна, що ставить його в розряд ексклюзивних інструментів, доступних переважно для фінансових професіоналів та великих корпорацій. Цей високий вартісний поріг утруднює доступ до цього інструменту для широкого кола користувачів.

Інші онлайн-платформи, такі як TradingView та CryptoCompare, грають важливу роль у візуалізації ключових показників криптовалютного ринку. TradingView славиться широкою спільнотою трейдерів та надає можливість аналізувати графіки цін, а також володіє власними інструментами для вдосконаленого технічного аналізу. З іншого боку, CryptoCompare відрізняється своєю різноманітністю аналітичних інструментів, які допомагають в детальному вивченні та порівнянні криптовалют. Ця платформа враховує різноманітні аспекти ринку, надаючи користувачам можливість здійснювати інформовані рішення при торгівлі та інвестуванні. Використання цих онлайн-платформ не лише спрощує доступ до актуальної інформації, але і створює середовище для взаємодії та обміну думками великої групи учасників ринку. Їхня здатність агрегувати та представляти дані у зручний спосіб робить їх цінними інструментами для тих, хто прагне отримати повний обсяг інформації та здійснювати обдумані рішення на криптовалютному ринку.

3. Інструменти для самостійного аналізу даних

Аналіз даних та статистика є ключовим етапом в процесі розуміння та використання інформації. Для маніпулювання даними, в тому числі криптовалютного ринку, використовуються різноманітні інструменти, такі як середовище MS Excel, мови програмування Python та R, а також відповідні бібліотеки для програмування.

Ці інструменти спрямовані на те, щоб зробити аналіз даних зручним та ефективним. Вони дозволяють створювати та взаємодіяти з таблицями, будувати графіки, виконувати різні операції статистичного аналізу та створювати моделі різної складності. Наприклад, мова програмування Python, що використовується далі у дослідженні, разом із відповідними бібліотеками стає потужним інструментом для створення моделей прогнозування. Гнучкість цих інструментів та можливість програмування дозволяють проводити глибокий та індивідуальний аналіз даних, враховуючи усі особливості конкретного завдання. Як результат, аналітики можуть використовувати ці інструменти для отримання цінної інформації з великих обсягів даних, розробки ефективних стратегій та прийняття обґрунтованих рішень.

На даний момент спільнота криптовалютного ринку може використовувати більш складні для розуміння, але не менш ефективні методи машинного та глибокого навчання. Наприклад, у мові програмування Python створені спеціалізовані бібліотеки, такі як Scikit-Learn, що надає ефективні інструменти для класифікації, регресії, кластеризації та інших завдань машинного навчання та аналізу даних, TensorFlow - для створення та тренування глибоких нейронних мереж та інших моделей машинного навчання, PyTorch, що є фреймворком для машинного навчання та глибокого навчання, відомим своєю гнучкістю та зручним інтерфейсом, що полегшує розробку та експерименти з нейронними мережами. Для людей із невеликим технічним та науковим досвідом можуть стати у нагоді такі інструменти, як Facebook Prophet, CryptoPredictions. Дані інструменти допомагають використовувати методи машинного та глибокого навчання для прогнозування цін у більш спрощеній для сприйняття формі. До прикладу, CryptoPrediction [35] -

спеціалізована платформа для прогнозування криптовалютних цін. Система аналізує обширні історичні дані, включаючи обмінні курси та ринкові тенденції, для конкретної криптовалюти. Використовуючи цю історичну інформацію, алгоритм використовує комбінацію лінійних та поліноміальних регресій для генерації прогнозів щодо можливих майбутніх змін цін. Акцент робиться на забезпеченні точних прогнозів, позбавлених надто оптимістичних прогнозів, що часто зустрічаються у криптовалютному просторі. Підхід заснований на даних, надає користувачам числові висновки на основі аналізу історичних даних цінової динаміки.

Для більш повного розуміння ринку також аналізуються дані за тематикою криптовалют та процесів, пов'язаних з ними, під час аналізу соціальних мереж, відкритих джерел даних, показники стосовно блокчейн-транзакцій та матеріали про актуальний становище відносно нових проектів. Для відстеження новин та реакцій спільноти на соціальних мережах можуть використовуватися спеціалізовані платформи, до прикладу CryptoSlate, CoinMarketCal. Такі інструменти надають інформацію про громадську думку і нові події у сфері криптовалют. Крім того, для пришвидшеного отримування необхідних для аналізу даних використовуються такі інструменти, як Web Scraping, тобто збирання даних з веб-сайтів за різними техніками, з використанням програмування. Також пришвидшити процес збирання даних допомагає використання публічних API від компаній, які пропонують вже готові рішення. Інструменти, як наприклад, Blockchain Explorers, Etherscan або Bitquery, що охоплює більше 20 криптовалют, надають дані про транзакції та статистику блокчейну створені для аналізу виконаних транзакцій, огляд та збирання статистики смартконтрактів.. Такі інструменти забезпечують детальний огляд транзакцій та структури блокчейну. Відомими прикладами ресурсів, завдяки яким можна отримати інформацію про нові та існуючі проекти, запуски tokenів тощо є ICObench, CoinGecko. Додаючи фундаментальних знань особі, що приймає рішення

про взаємодію з криптовалютним ринком, вони дозволяють оцінювати нові проекти та їхні потенційні вигоди.

У подальшому дослідженні використовується Neural Prophet [36] – фреймворк для навчання моделей та прогнозування часових рядів. Neural Prophet є високотехнологічною та науково розробленою моделлю прогнозування, яка може враховувати важливі зовнішні фактори у процесі прогнозування. Модель має автоматичний механізм вибору гіперпараметрів, пов'язаних із навчанням, що підвищує її ефективність на тестових даних. Крім того, Neural Prophet використовує сезонність зі змінними Фур'є, що надає можливість враховувати сезонні закономірності на різних інтервалах, таких як щорічні, щоденні, щотижневі та щогодинні періоди. Також цей інструмент дозволяє моделі адаптуватися до нелінійних тенденцій у даних. Крім того, Neural Prophet надає можливості візуалізації, що дозволяє користувачам оглядати компоненти прогнозу, коефіцієнти моделі та кінцеві прогнози. Під час навчання модель адаптується до лагових та майбутніх регресорів, дозволяючи враховувати часові залежності для отримання більш точних прогнозів. Підсумовуючи, це зручний і потужний пакет для мови програмування Python, що поєднує легкість використання з високим рівнем функціональності, надаючи зручний інтерфейс та високорівневі можливості у сфері прогнозування.

РОЗДІЛ 2

ЗАСТОСУВАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ АНАЛІЗУ ЦІНОВИХ КОЛИВАНЬ КРИПТОВАЛЮТИ

2.1 Збирання та обробка історичних даних про ціни криптовалюти Bitcoin

Для дослідження була обрана криптовалюта Bitcoin (BTC). Таке рішення пояснюється тим, що вона займає приблизно 50% усього криптовалютного ринку [37]. Дослідження не включає аналіз альткоїнів, оскільки більшість з них змінюються в часі в залежності від змін головного гравця ринку - Біткоїна. Таким чином, обрана криптовалюта задає тенденції функціонування та динаміки всього ринку, є найбільш розповсюдженим та найбільш відомим криптоактивом. Аналіз та прогнозування часового ряду буде перевірятися та порівнюватися на декількох моделях, тому для майбутніх досліджень, можна буде замінити цінові дані Біткоїна на дані певних альткоїнів, перевіряючи ефективність нових побудованих прогнозів.

Збирання історичних даних BTC відбувалося на відомому ресурсі Yahoo.Finance [38]. Ця платформа дозволяє користувачам досліджувати різноманітні фінансові показники, котирування валют, а також отримувати актуальні новини з усього світу фінансів, у тому числі з ринку криптовалют. Було обрано проміжок часу впродовж останніх п'яти років, тобто з листопада 2018 року по листопад 2023 року. Частота даних була обрана у вигляді щоденних показників. Платформа дозволяє завантажити датасет у форматі CSV із інформацією щодо таких показників:

- Дата спостереження - Date
- Ціна відкриття - Open
- Найвища ціна за час дати спостереження - High
- Найнижча ціна за час дати спостереження - Low

- Ціна закриття – Close
- Скоригована ціна закриття – Adj Close
- Обсяг торгів – Volume

Для додаткових експериментів із аналізом впливу рівня сентименту, що вказує на загальний емоційний стан учасників ринку – показника Fear and Greed Index (Індекс страху та жадібності), було використано дані з веб-ресурсу Alternative [39]. Індекс страху та жадібності використовується зазвичай, як спосіб описати одним числом - умовним балом від 0 до 100 емоції гравців на ринку. В тому числі за ним можна зробити припущення, про те, що трейдери, інвестори, та загалом, спільнота відчують зараз у зв'язку з ринковим становищем: невизначеність, оптимізм, паніку або розчарування, піднесення тощо. Визначенням індексу займаються декілька інтернет-ресурсів, проте підходи до визначення дуже схожі. За методикою визначення та розрахунку цього показника веб-сайтом Alternative, це інтегральний показник, який поєднує:

- Волатильність активу
- Об'єм ринку
- Згадування в соціальних медіа
- Опитування учасників ринку
- Домінація криптовалюти
- Тренди запитів до пошукових мереж

Оцінюється поточна волатильність та максимальний відсоток втрат для біткоіна порівняно із середніми значеннями протягом останніх 30 та 90 днів. Надмірний ріст волатильності розглядається як ознака ринку, що переживає страх. Також оцінюється поточний обсяг та динаміка ринку порівняно із середніми значеннями за останні 30/90 днів, об'єднуючи ці дані. Зазвичай великі обсяги покупок на позитивному ринку свідчать про надмірно жадібну поведінку ринку. Додатково проводиться аналіз настроїв Reddit та Twitter. Враховуються та

підраховуються повідомлення за різними хештегами, перевіряючи, наскільки швидко та скільки взаємодій вони отримують. Платформа проводить щотижневі криптоопитування, ретельно вивчаючи результати як виявлення настроїв групи криптоінвесторів. Визначається також і домінантність монети, що відображає ринкову частку в крипторинку. Зокрема для Bitcoin вважається, що зростання домінантності спричинене страхом від (і, отже, зменшенням) занадто спекулятивних інвестицій в альткойни, оскільки Bitcoin поступово стає більш безпечним активом. Для дослідження інтересів учасників ринку використовуються дані Google Trends для різних запитань, пов'язаних з Bitcoin, зокрема зміни обсягів пошуків та рекомендованих популярних запитань.

Для аналізу впливу показника Fear and Greed Index було зібрано дані його значень впродовж останніх п'яти років.

У середовищі MS Excel датасет було підготовано для подальших дій шляхом коректного сортування даних за датою спостереження.

Далі за допомогою середовища Google Colaboratory було імпортовано датасет, використовуючи бібліотеку Pandas (скорочення у коді «pd»). Загалом, для трансформації обробки та розбиття даних, використовувались як вбудовані функції Python, так і такі бібліотеки та модулі:

- **Pandas** – Використовується для завантаження та збереження даних. Pandas надає структури даних, такі як DataFrame, що полегшує завантаження та маніпулювання даними. Також Pandas дозволяє виконувати різноманітні операції з даними, такі як сортування, групування, фільтрація тощо. Pandas дозволяє об'єднувати дані з різних джерел та виконувати операції з'єднання.

- **NumPy**

NumPy надає широкий функціонал для виконання математичних операцій на масивах чисел. NumPy пропонує об'єкт масиву, який полегшує векторизовані операції та має велику кількість вбудованих функцій для масивів. Також NumPy має вбудовані функції для генерації випадкових чисел та масивів.

На відміну від акцій, які торгуються у робочі дні в окремі години, криптовалютний ринок функціонує безперервно. Щоб згладити щоденні коливання ціни, було побудовано зважену ціну по кожному дню за формулою:

$$\text{Зважена ціна} = \frac{\text{Ціна відкр.} + \text{Ціна закр.} + \text{Найв.ціна за день} + \text{Найн.ціна за день}}{4} \quad (2.1)$$

Подальші дії над даними відбувалися з таким набором показників:

- Дата - Date
- Обсяг торгів - Volume
- Індекс сентименту - FG_Index
- Зважена ціна – weighted_price

Виділення чотирьох стовпців для дослідження із трансформованими цінами у зважену ціну за день представлені прикладом з перших п'яти рядків нової таблиці на рис. 2.2.

	Date	Volume	FG_Index	weighted_price
0	2018-11-15	7.032140e+09	28	5629.344848
1	2018-11-16	5.279320e+09	23	5594.207397
2	2018-11-17	4.303150e+09	24	5557.762573
3	2018-11-18	4.159680e+09	26	5599.157592
4	2018-11-19	7.039560e+09	28	5238.989990

Рисунок 2.2 – Формування датасету зі зваженою ціною

Джерело: розроблено автором у програмі Google Colaboratory на основі [37,38]

Проведено аналіз змінних та дослідження їх взаємозв'язків задля розглядання можливості відноситися до даних, як до регресійної моделі із незалежними змінними FG_Index та Volume та залежною – weighted_price. Для дослідження більш широких часових проміжків дані було згруповано помісячно, обираючи середньомісячне

значення по кожному показнику, маючи на меті зменшити вплив волатильності активу на якість моделі. Потім було виведено основну дескриптивну характеристику датасету за допомогою функції `describe()`. Результати функції включають визначення наступних показників:

- `count` - кількість непорожніх значень.
- `mean` - Середнє (середнє) значення.
- `std` – стандартне відхилення.
- `min` - мінімальне значення.
- 25% - 25% процентиля.
- 50% - 50% процентиля.
- 75% - 75% процентиля.
- `max` - максимальне значення.

Результати побудови статистичних показників за допомогою функції `describe()` представлено на рис. 2.3.

	Volume	FG_Index	weighted_price
count	6.100000e+01	61.000000	61.000000
mean	2.836292e+10	45.312463	23595.247986
std	1.528512e+10	19.625916	16084.398895
min	5.397926e+09	11.266667	3705.384626
25%	1.722532e+10	28.806452	9484.572298
50%	2.680464e+10	44.800000	21483.779156
75%	3.643816e+10	57.096774	34581.706182
max	8.096972e+10	92.064516	60703.981609

Рисунок 2.3 – Дескриптивна характеристика датасету

Джерело: розроблено автором у програмі Google Colaboratory на основі [37,38]

З побудованого набору показників видно, що дані згруповано у середньомісячні спостереження, таким чином, будуть досліджуватися показники шістдесяти одного місяця. Варто відзначити, що за 5 останніх років, емоційний стан

учасників ринку, виражений у індексі сентименту (FG_Index) змінювався в проміжку від умовного відчаю та безнадії (найменше значення - 11.2), до майже повного оптимізму та віри у криптовалюту (максимальне значення – 92), а середнє значення впродовж останніх 5 років знаходилося на рівні 45 пунктів, що свідчить про те, що ринок в основному відчував нейтральні емоції, з невеликим переважанням страху перед жадібністю.

Задля визначення можливих взаємозв'язків показників на рис. 2.4 провізуалізовано розподіл значень змінних, додаючи до аналізу точкові та частотні діаграми.

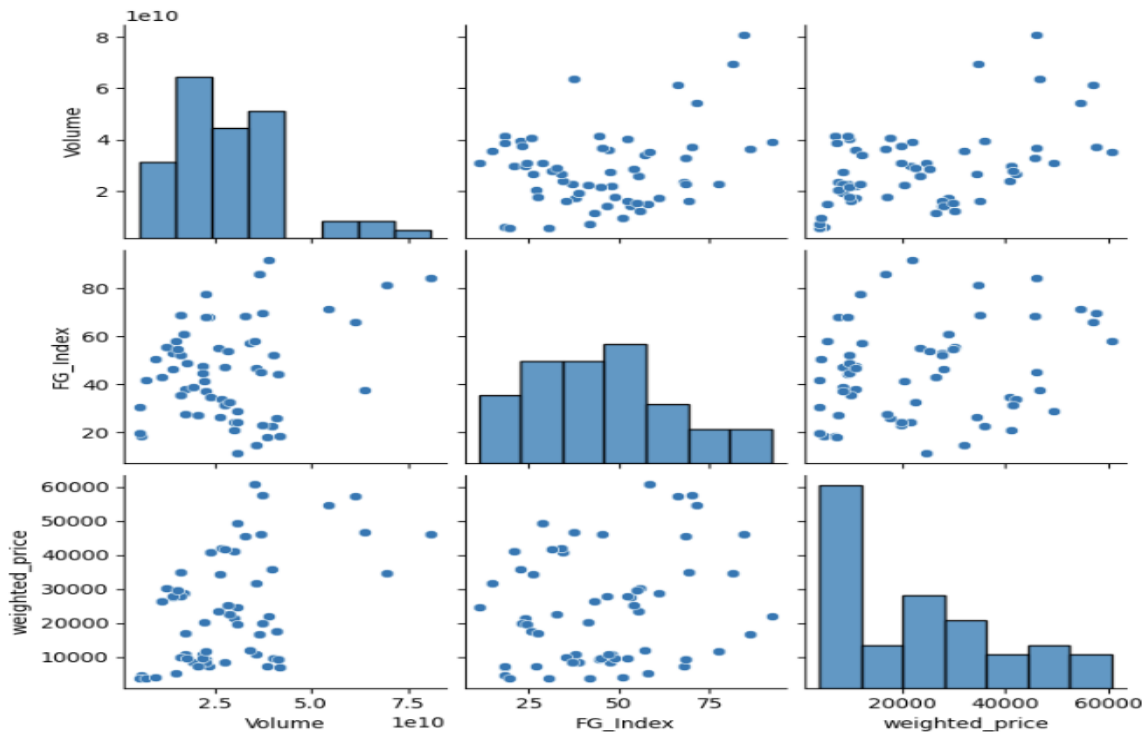


Рисунок 2.4 – Діаграми розподілу та графік розсіювання

Джерело: розроблено автором у програмі Google Colaboratory на основі [37,38]

З візуалізації частотних діаграм можна помітити, що дані розподілено досить нерівномірно, лише показник індексу сентименту має ближчий до нормального розподіл, ніж у інших показників, що можна пояснити тим, що цей показник має

чіткі межі від 0 до 100, а медіанне значення близьке до п'ятдесяти. Наразі за графіками розміювання не можна помітити чіткі та явні залежності між величинами усіх пар змінних. Дані було стандартизовано методом `StandardScaler()` [40]. Його основне призначення полягає в тому, що від кожної незалежної змінної віднімається середнє значення, а потім результат ділиться на стандартне відхилення цієї змінної:

$$Z = \frac{x - \mu}{\delta} \quad (2.5)$$

де z – стандартизоване (нове) значення незалежної змінної;

x – значення незалежної змінної;

μ - середнє значення незалежної змінної;

δ - стандартне відхилення.

Для дослідження можливої статистичної залежності між змінними, досліджено їхній кореляційний зв'язок з використанням коефіцієнту кореляції. Цей показник може змінюватися від -1 до 1. При додатній кореляції залежність між величинами буде прямою: при збільшенні однієї величини, збільшується й інша. При від'ємній кореляції залежність обернена: збільшення однієї величини пов'язано зі зменшенням другої. Коефіцієнти кореляції між змінними обсягу торгів(`Volume`), індексу сентименту (`FG_Index`) та зваженої ціни(`weighted_price`) зображено на кореляційній матриці на рис. 2.6.

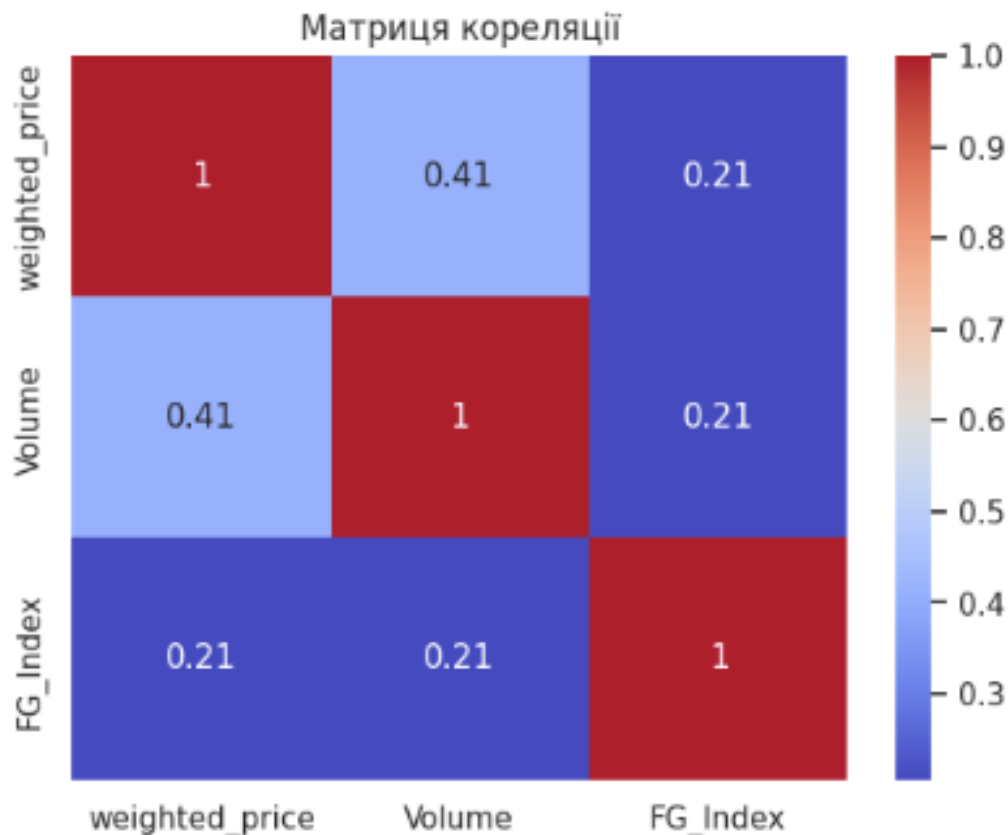


Рисунок 2.6 – Кореляційна матриця

Джерело: розроблено автором у програмі Google Colaboratory на основі [37,38]

Умовна сила зв'язку індекса настрою зі зваженою ціною складає 21%. Таке ж значення для зв'язку індекса настрою зі змінною обсягу торгів. Трохи більше статистичного зв'язку – 41% - є між показниками обсягу торгів та ціною. Але втім, значення коефіцієнтів кореляції свідчать про слабкий статистичний зв'язок між кожною з пар змінних. Проте, є необхідність проводити більш глибокий аналіз, щоб спростувати або підтвердити взаємозалежності факторів.

Стандартизовані дані було використано для побудови множинної регресійної моделі для визначення спроможності незалежних змінних описувати цільову. Моделювання здійснювалося з використанням бібліотек sklearn та statsmodels [41].

Регресійна модель описується формулою:

$$Y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \dots + \beta_p * x_p + \varepsilon \quad (2.7)$$

де Y – залежна змінна;

$x_1, x_2, \dots, x_p$ – незалежні змінні;

$\beta_0, \beta_1, \dots, \beta_p$ – невідомі параметри, які потрібно оцінити;

ε – випадкова величина.

Для оцінки якості побудованої моделі був використаний коефіцієнт детермінації, що вимірює ступінь варіації залежної змінної, яку можна пояснити моделлю. В інших словах, він вказує, наскільки точно модель відображає залежну змінну від незалежних змінних. Коефіцієнт детермінації може приймати значення від 0 до 1, де 0 вказує на жодного пояснення, а 1 - на повне пояснення.

Результати побудови моделі множиної лінійної регресії на тренувальній вибірці показали статистичну значущість параметру Volume, і статистичну незначущість параметру моделі FG_Index. Якість моделі за показником коефіцієнта детермінації становить лише 0,267. Повний результат представлений на рис. 2.8:

```

=====
                        OLS Regression Results
=====
Dep. Variable:          weighted_price      R-squared:                0.267
Model:                  OLS                 Adj. R-squared:           0.234
Method:                 Least Squares       F-statistic:             8.178
Date:                   Thu, 30 Nov 2023    Prob (F-statistic):      0.000935
Time:                   18:44:41           Log-Likelihood:          -525.67
No. Observations:      48                 AIC:                    1057.
Df Residuals:          45                 BIC:                    1063.
Df Model:              2
Covariance Type:       nonrobust
=====
                        coef      std err          t      P>|t|      [0.025      0.975]
-----
const          2.337e+04    2069.405     11.295     0.000     1.92e+04     2.75e+04
Volume         7865.3277    2058.307      3.821     0.000     3719.685     1.2e+04
FG_Index        56.5729    2138.020      0.026     0.979    -4249.620     4362.766
=====
Omnibus:                 3.870    Durbin-Watson:           2.012
Prob(Omnibus):           0.144    Jarque-Bera (JB):        3.462
Skew:                    0.576    Prob(JB):                0.177
Kurtosis:                2.365    Cond. No.                1.42
=====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

R^2: 0.26656852166720657
Mean Squared Error: 203132824.41540527
F-Statistic: 8.177712458082768, p-value: 0.0009345225308554511

T-Scores and P-Values:
      T-Score  P-Value
Volume   3.821261  0.000405
FG_Index  0.026460  0.979007

```

Рисунок 2.8 – Результат побудови моделі лінійної регресії

Джерело: розроблено автором у програмі Google Colaboratory на основі [37,38]

Згідно з оцінкою параметрів моделі, змінна FG_Index виявилася статистично незначущою, фактор обсягу торгів та вільний параметр – статистично значущі, як і модель загалом.

Далі для більш глибокого дослідження було використано методи MLP(багатошаровий перспетрон) та RF(Випадковий Ліс). Багатошаровий персептрон — це нейронна мережа, що з'єднує кілька шарів нейронів, що означає, що шлях сигналу через вузли проходить лише в один бік. Кожен нейрон, крім вхідних, має нелінійну функцію активації. MLP використовує зворотне розповсюдження як методику навчання з учителем. Випадковий ліс — це широко використовуваний алгоритм машинного навчання, який поєднує результати кількох дерев рішень для досягнення єдиного результату. Його простота використання та гнучкість сприяли його прийняттю у науковій спільноті, оскільки він обробляє як проблеми класифікації, так і регресії.

Візуалізація роботи методів MLP та RF, переваги та недоліки цих методів представлені на рис. 2.9.

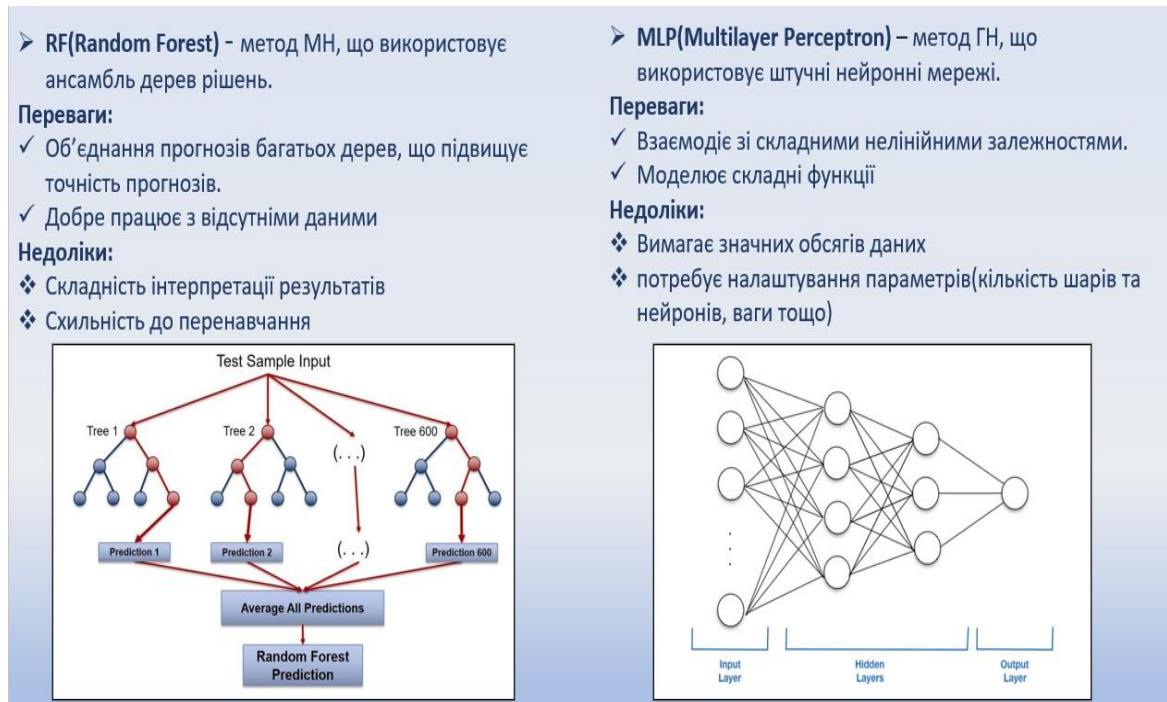


Рисунок 2.9 – Візуалізація роботи методів MLP та RF, їх переваг та недоліків

Джерело: розроблено автором на основі [42, 43]

Багатошаровий перцептрон можна побудувати, завантаживши необхідні методи з пакету *sklearn* у мові програмування *Python*. Основні методи, які використовувалися, перелічені нижче.

`train_test_split` – метод, що використовується для розбиття вибірки на тренувальну та тестову. Під час розбиття прийнято обирати розмір тестової вибірки близько 20 – 30 відсотків, щоб перевірити ефективність моделі на достатній кількості невідомих значень. В регресійних моделях також варто задавати параметр `random_state`, задля того, щоб дані для розбиття вибиралися випадковим чином. В даному етапі роботи було обрано розмір тренувальної вибірки – 80%, розмір тестової вибірки – 20% від загальної кількості спостережень.

MLPRegressor – метод, що використовується для ініціалізації моделі багатошарового персептрону. Його основними параметрами є `hidden_layer_sizes`, що визначає кількість нейронів у кожному прихованому шарі мережі. Параметр `activation` вказує на функцію активації, яка використовується у прихованих шарах, такі як 'logistic' – логістична функція активації, 'tanh' - тангенсна, або функція, що частіше за інші використовується для навчання нейронних мереж - 'relu' - це одна з функцій активації, яка часто використовується у штучних нейронних мережах. Фактично, це просто математична функція, яка обчислюється як максимум між нулем і вхідним значенням. До прикладу, якщо вхід до нейрона менше нуля, то функція поверне нуль; якщо вхід більше або дорівнює нулю, то вона поверне саме це значення. ReLU є популярним вибором через свою простоту та ефективність. Вона дозволяє передавати позитивні значення без змін, а негативні призводить до нуля, виконуючи при цьому нелінійну трансформацію. Ця нелінійність допомагає моделі нейронної мережі навчатися складнішим завданням, оскільки додає гнучкість у представленні внутрішніх шарів мережі. Параметр `solver` визначає оптимізатор для навчання ваг мережі, де 'lbfgs' використовує метод квазі-Ньютона, 'sgd' - стохастичний градієнтний спуск, і 'adam' - стохастичний оптимізатор. Також можна налаштувати параметр `alpha`, що відповідає за силу L2 регуляризації, що допомагає уникнути перенавчання.

В дослідженні було використано можливість перебору параметрів для виведення тих параметрів, що дають найвищий показник ефективності моделі. Для оцінки точності моделі. Для порівняння ефективності розробленого багатошарового персептрону із моделлю множинної регресії, побудованої раніше, використовується показник коефіцієнт детермінації, або R^2 .

Найкращими показниками після навчання моделі виявилася кількість нейронів 2 та 1 у прихованих шарах, функція активації - relu, та параметр регуляризації alpha, що дорівнює 0,001. Після зазначеного перебору, модель перевірялася на тестових даних. Результат коефіцієнту детермінації становить 0,16,

що вказує на те, що якість побудованої моделі знизилась, порівняно з множинною лінійною регресією.

Порівняння результатів порівняння реальних та прогнозованих значень представлено у вигляді точкової діаграми на рис. 2.10.

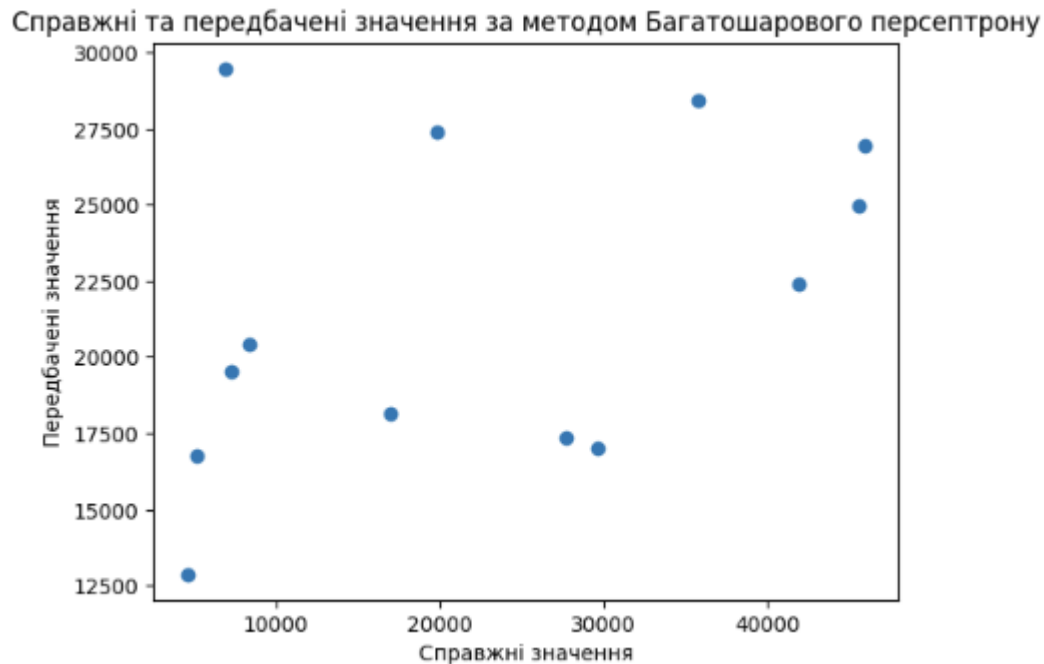


Рисунок 2.10 – Діаграма розсіювання реальних та прогнозованих значень за методом MLP

Джерело: розроблено автором у програмі Google Colaboratory на основі [37,38]

З рис. 2.10 видно, що модель дійсно значно помиляється у передбаченні ціни за регресорами. Якісним результатом вважався би той, де точки були б розповсюджені більш лінійно, в один ряд. Можна зробити висновок, що метод багатошарового перцептрону не допомагає досягти кращої точності. Це може бути пов'язане з невеликою кількістю даних через їх попереднє групування.

Для побудови випадкового лісу за допомогою Python також необхідно завантажити необхідний пакет – RandomForestRegressor. Метод випадкового лісу є ансамблем методом машинного навчання, який використовує багато дерев прийняття рішень для здійснення прогнозів у регресійних задачах. Основна ідея

полягає в тому, щоб випадковим чином обирати підвибірки даних та використовувати їх для тренування окремих дерев. Кожне дерево прийняття рішень будується на випадковому піднаборі факторів і випадковій підвибірці даних. Коли всі дерева побудовані, їхні прогнози комбінуються, часто шляхом усереднення, для отримання кінцевого регресійного прогнозу. Даний метод також дозволяє оцінити важливість факторів. Метод випадкового лісу може допомогти у визначенні важливості факторів на основі того, як вони сприяють точності передбачень моделі. Оцінка важливості факторів випадкового лісу може варіюватися, але загальний підхід зазвичай базується на двох основних аспектах, це зменшення критерію вузла (Node Impurity Reduction), коли дерево приймає рішення на кожному вузлі, воно розглядає різні фактори та їх значення для поділу даних. Важливість фактора вимірюється тим, як сильно він зменшує загальну "нечистоту" (impurity) вузла. Чим більше зменшення, тим важливіше фактор для розділення даних. Також оцінюється приріст важливості фактора (Mean Decrease in Impurity, MDI). При обчисленні важливості факторів для випадкового лісу, оцінка MDI враховується для кожного дерева в лісі, а середнє зменшення важливості фактора для всіх дерев оцінюється як важливість цього фактора в рамках всього лісу. Коли важливість факторів визначена, вона може бути нормалізована, щоб сума важливостей дорівнювала 1 або 100%. Це полегшує порівняння важливості різних факторів. Важливість факторів у випадковому лісі може бути використана для ранжування факторів за їх впливом на модель, а також для відбору факторів при необхідності покращення ефективності моделі або зменшення складності. На рис. 2.11 зображено порівняння реальних та прогнозованих значень за методом Random Forest, а також виведені розраховані значення важливості факторів та коефіцієнт детермінації.

R² Score: 0.11187277259624118

Важливість факторів:

Volume: 0.617742922830821

FG_Index: 0.3822570771691791

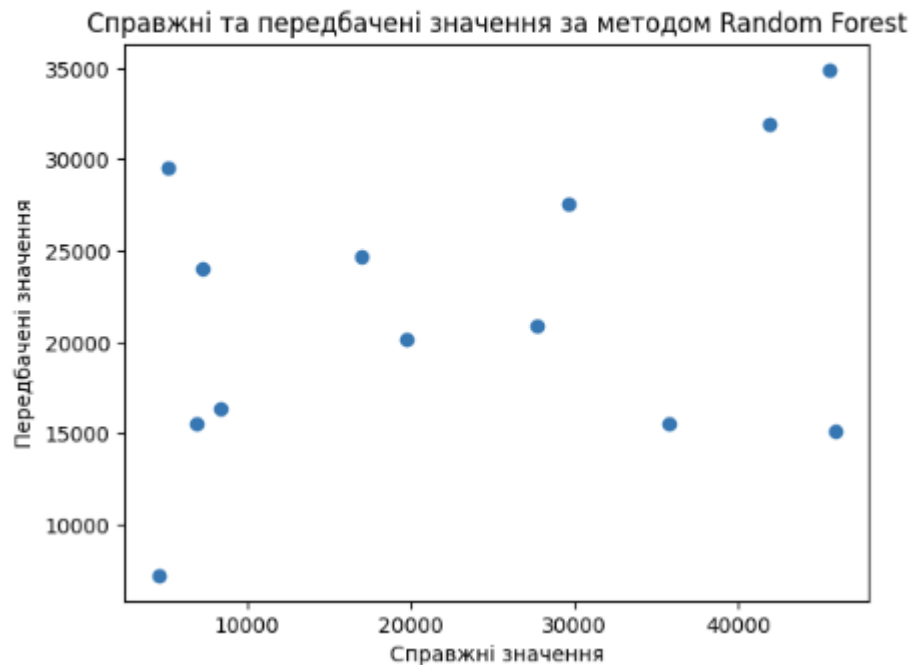


Рисунок 2.11 – Діаграма розсіювання реальних та прогнозованих значень за методом Random Forest

Джерело: розроблено автором у програмі Google Colaboratory на основі [37,38]

З розрахованих значень важливості факторів для обсягу торгів – 62% та FG_Index – 38, можна лише порівняти саме їх, а не їхній вплив на динаміку ціни. Коефіцієнт детермінації, що обраний за показник якості моделі, виявився ще нижчим, ніж у двох попередніх методів. Точкова діаграма вказує на те, що моделі важко розпізнати динаміку ціни на основі регресорів.

Незважаючи на більш складні обчислення і налаштування параметрів [44], дані методи зменшили якість моделі до значень коефіцієнту детермінації рівного 0.16 та 0.11 відповідно. Це можна пояснити малою кількістю даних вибірки та слабким зв'язком між змінними, про що було зазначено вище. Отже, точність моделі занадто низька для того щоб використовувати її як метод прогнозування.

Після спроби побудувати модель з двома незалежними змінними, було проведено аналіз взаємодії ціни (`weighted_price`) окремо з кожною змінною. Спершу була розглянута змінна `Volume` наразі не як фактор моделі, а як цільова змінна, щоб дослідити, як зміна саме ціни впливає на зміну об'єму торгів. Результати моделювання представлені на рис. 2.11.

OLS Regression Results						
Dep. Variable:	Volume	R-squared:	0.250			
Model:	OLS	Adj. R-squared:	0.237			
Method:	Least Squares	F-statistic:	19.62			
Date:	Tue, 12 Dec 2023	Prob (F-statistic):	4.16e-05			
Time:	20:40:07	Log-Likelihood:	-1507.8			
No. Observations:	61	AIC:	3020.			
Df Residuals:	59	BIC:	3024.			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	1.716e+10	3.05e+09	5.622	0.000	1.11e+10	2.33e+10
weighted_price	4.747e+05	1.07e+05	4.429	0.000	2.6e+05	6.89e+05
Omnibus:	10.355	Durbin-Watson:	0.424			
Prob(Omnibus):	0.006	Jarque-Bera (JB):	10.124			
Skew:	0.927	Prob(JB):	0.00633			
Kurtosis:	3.739	Cond. No.	5.09e+04			

Рисунок 2.11 – Результати моделі для дослідження впливу ціни активу на обсяги торгів

Джерело: розроблено автором у програмі Google Colaboratory на основі [37,38]

В результаті, побудована модель, хоч і була статистично значущою, вказала на значення коефіцієнту детермінації рівному 0,25, що не дозволяє нам робити висновок про суттєву залежність об'єму торгів від ціни даного активу.

Для дослідження взаємодії змінної сентименту і ціни криптовалют дані були знов розбиті у щоденні значення. Спершу була розглянута змінна `FG_Index` із різними значеннями лагу – 1-7,14 та 21 день. Метою такого аналізу було дослідити вплив попередніх значень сентиментного фактору на формування даної ціни.

Результати представлені на рис. 2.12.

OLS Regression Results						
Dep. Variable:	weighted_price	R-squared:	0.051			
Model:	OLS	Adj. R-squared:	0.050			
Method:	Least Squares	F-statistic:	96.45			
Date:	Tue, 12 Dec 2023	Prob (F-statistic):	3.25e-22			
Time:	20:39:34	Log-Likelihood:	-20092.			
No. Observations:	1813	AIC:	4.019e+04			
Df Residuals:	1811	BIC:	4.020e+04			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	1.614e+04	848.552	19.016	0.000	1.45e+04	1.78e+04
FG_Index_Lagged	166.3623	16.940	9.821	0.000	133.138	199.586
Omnibus:	256.411	Durbin-Watson:	0.007			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	143.446			
Skew:	0.549	Prob(JB):	7.10e-32			
Kurtosis:	2.167	Cond. No.	115.			
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly sp						

Рисунок 2.12 - Результати моделі для дослідження залежності ціни від індекса сентименту з різними значеннями лагу та відображення результатів

Джерело: розроблено автором у програмі Google Colaboratory на основі [37,38]

Результатом експерименту стало те, що за всіх розглянутих значеннях лагу, коефіцієнт детермінації складав не більше 0,06.

Експеримент щодо залучення двох незалежних змінних - обсягу торгів та індексу сентименту для опису цільової змінної не дав позитивних результатів стосовно якості побудованих моделей. Побудовані моделі мають низький рівень якості. Оскільки додаткові змінні мають дуже низький вплив на цільову змінну, їх було видалено з датасету для того, щоб надалі проводити дослідження класичного часового ряду з даними у вигляді Час – Значення.

Тобто подальше дослідження проводилося з датасетом, який має стовпці:

- Date
- weighted_price

2.2 Аналіз динаміки ціни криптовалюти

Для загального відображення динаміки ціни BTC необхідно побудувати графік часового ряду для первинного огляду. Для цього датасет було повернено у вигляд, що визначає щоденні показники зваженої ціни та дати спостереження. На рис. 2.13 представлено динаміку ціни по днях впродовж останніх п'яти років.

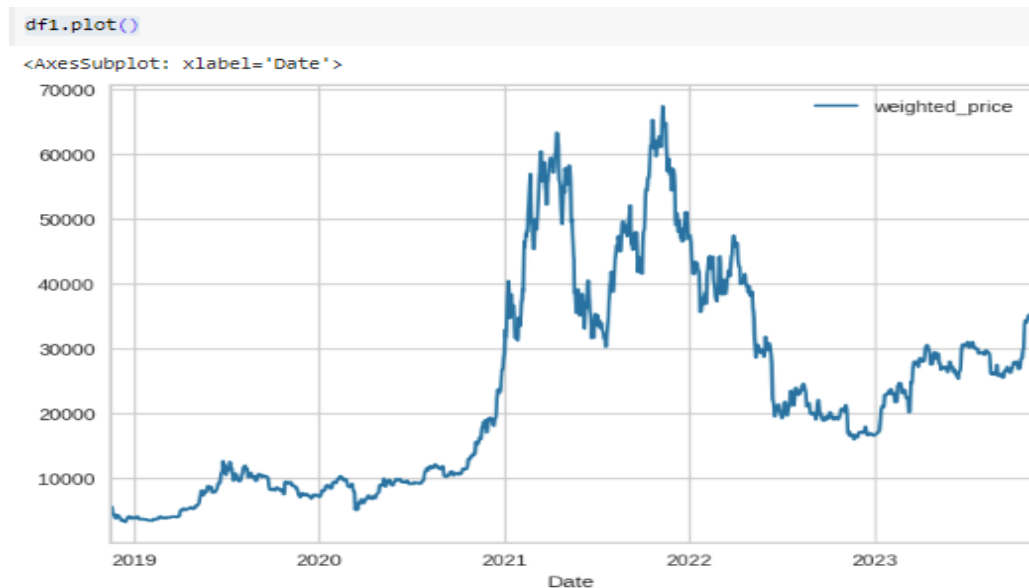


Рисунок 2.13 – Ціни BTC з листопада 2018 року по листопад 2023 року

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

Для аналізу часового ряду, постає завдання визначити його ключові компоненти [45]. Зазвичай, серед ключових компонентів виділяють тренд, сезонність, циклічність та залишкову складову.

Тренд відображає загальний напрямок розвитку часового ряду, вказуючи на довгострокові тенденції або зміни. Тренд можна визначити, наприклад, за допомогою лінійної апроксимації (наприклад, метод найменших квадратів), експоненційного згладжування або фільтрів, таких як ковзне середнє. Сезонність вказує на циклічні зміни часового ряду, які повторюються протягом фіксованого періоду (наприклад, щоденно, щомісяця або щороку). Сезонність можна визначити шляхом порівняння значень у тому ж самому періоді різних циклів, або застосовуючи методи аналізу гармонік, такі як фур'є-аналіз. Циклічність вказує на більш довгострокові циклічні зміни, які не мають фіксованого періоду, і можуть бути менше регулярними, ніж сезонні зміни. Циклічність може бути визначена, наприклад, за допомогою фільтрів або статистичних методів, таких як детрендінг (видалення тренду). В даному дослідженні не розглядається циклічний компонент, оскільки, як і історичний проміжок, що береться для аналізу - п'ять років, так і саме явище - криптовалюти - з'явилося відносно нещодавно, і це не дає можливість проаналізувати можливі цикли. Випадковий компонент включає в себе випадкові або непрогнозовані зміни, які не пояснюються трендом, сезонністю або циклічністю. Він розраховується відніманням прогнозованих значень (тренду, сезонності, циклічності) від спостережень.

Математичне представлення значення часового ряду у момент t – представлено у формулі.

$$Y(t)=T(t)+S(t)+C(t)+\varepsilon(t), \quad (2.14)$$

де $Y(t)$ - значення часового ряду у момент часу t ;

$T(t)$ - тренд у момент часу t ;

$S(t)$ - сезонність у момент часу t ;

$C(t)$ - циклічність у момент часу t ;

$\varepsilon(t)$ - випадковий або резидуальний компонент у момент часу t .

Дослідження часового ряду проводилось за допомогою пакету *statsmodels.tsa.seasonal*, що здійснює визначення компонент автоматично. З рис. 2.15 можна помітити загальний тренд на зростання, попри різке зростання і падіння ціни в проміжок з середини 2021 року по початок 2022 року, коли популярність усього криптовалютного ринку була дуже висока, гравці ринку були сповнені надій щодо прогресу багатьох активів, а ціна BTC злетіла за 67 тисяч американських доларів.

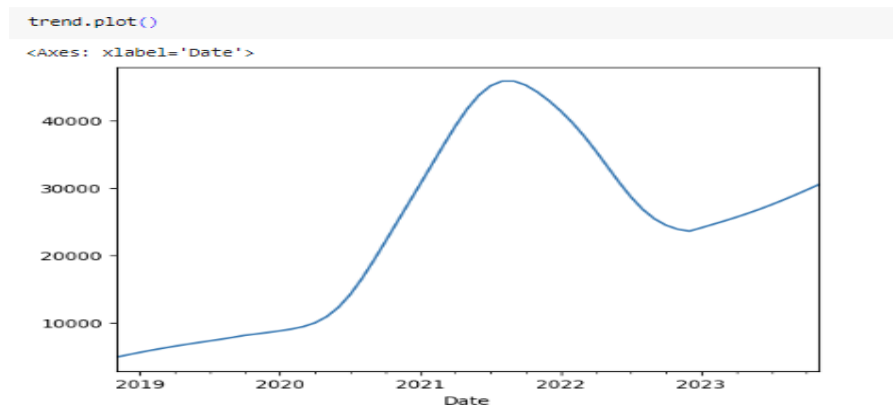


Рисунок 2.15 – Трендова складова часового ряду

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

Учасники ринку також можуть звертати увагу на динаміку ціни впродовж року для виявлення певних патернів поведінки. На рис. 2.16 досліджено характер річних коливань.

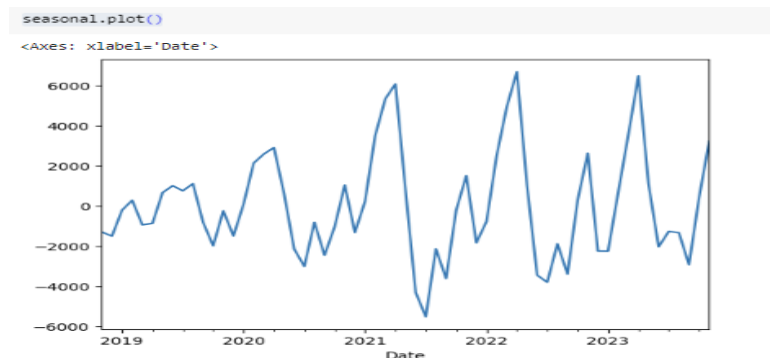


Рисунок 2.16 – Сезонна складова часового ряду

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

Дослідження річних коливань вказує на те, що в середньому значення ціни зростає від початку року до середини весни. Влітку в середньому відбуваються просідання, а осінь характеризується чергуваннями підняття та зменшення ціни. Висновки з візуалізації цього показника можуть бути корисними, до прикладу, якщо трейдер займається середньостроковими операціями. Знаючи характер поведінки ціни криптовалюти у різні періоди року, трейдер може ухвалювати рішення про купівлю, продаж, або уникнення дій.

Щодо днів тижня теж можна побачити цікаві характеристики часового ряду. З рис. 2.17 помітно, що найвищі ціни досягаються в середині тижня – з вівторка по четвер. А період з п'ятниці до понеділка є в середньому спадним, з невеликим підняттям ціни в середині вихідних днів.

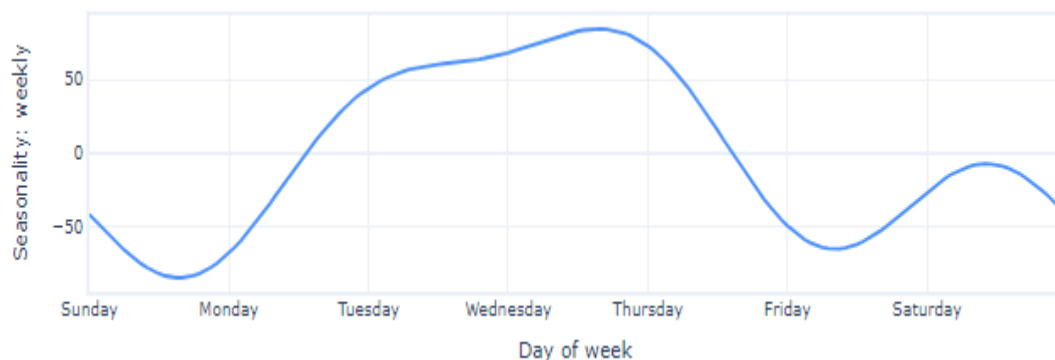


Рисунок 2.17 – Динаміка ціни криптовалюти в залежності від днів тижня

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

В одному з досліджень [46], де в інформаційній базі присутні щогодинні прогнози, також вказується те, що Bitcoin має тенденцію зростати о близько 22-23 годині доби. Такі спостереження можуть бути важливими для короткострокових трейдерів, проте, для середньо- і довгострокових прогнозів такі невеликі проміжки часу не відіграють суттєву роль.

Часовий ряд із виключенням сезонної та трендової компонент, маючи на меті зобразити залишки, представлено на рис. 2.18.

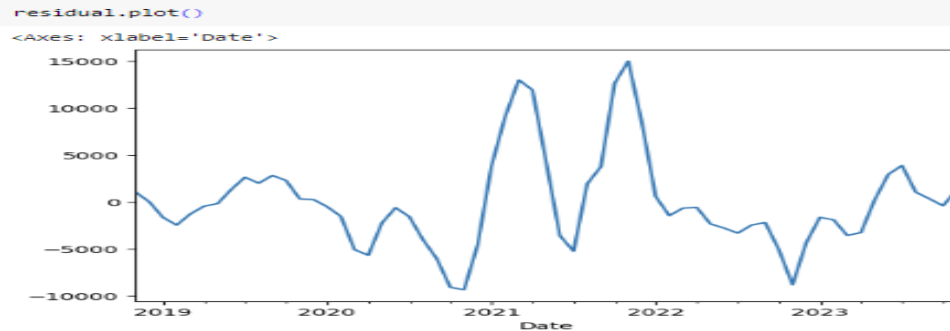


Рисунок 2.18 – Залишки часового ряду після виключення трендової та сезонної складової

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

Видно, що залишки також мають певний непослідовний, хаотичний характер. Це може свідчити про те, що дані можуть знаходитися під значним впливом зовнішніх чинників, які не враховані в моделі, або що часовий ряд взагалі важко передбачити або має велику ступінь випадковості, що може бути пояснено значною волатильністю такої криптовалюти, як Bitcoin.

Статистична перевірка на стаціонарність

Для роботи з часовим рядом, зокрема для побудови прогностичних моделей методами машинного навчання була використана бібліотека Rucaret [47]. Вона дозволяє завдяки невеликій кількості коду дізнатися основні показники часового ряду, а також провести тести на його стаціонарність. З візуалізації часового ряду видно, що він не є стаціонарним, проте було проведено додаткові тести на перевірку за допомоги функції `check_stats()`. Для часового ряду змінної `weighted_price` було отримано описові характеристики, результати представлено на рис. 2.19.

check_stats()

	Test	Test Name	Data	Property	Setting	Value
0	Summary	Statistics	Transformed	Length		1819.0
1	Summary	Statistics	Transformed	# Missing Values		0.0
2	Summary	Statistics	Transformed	Mean		23577.041135
3	Summary	Statistics	Transformed	Median		20496.82666
4	Summary	Statistics	Transformed	Standard Deviation		16152.95769
5	Summary	Statistics	Transformed	Variance		260918042.124528
6	Summary	Statistics	Transformed	Kurtosis		-0.541192
7	Summary	Statistics	Transformed	Skewness		0.682661
8	Summary	Statistics	Transformed	# Distinct Values		1819.0
9	White Noise	Ljung-Box	Transformed	Test Statistic	{'alpha': 0.05, 'K': 24}	41254.134568
10	White Noise	Ljung-Box	Transformed	Test Statistic	{'alpha': 0.05, 'K': 48}	77401.649872
11	White Noise	Ljung-Box	Transformed	p-value	{'alpha': 0.05, 'K': 24}	0.0
12	White Noise	Ljung-Box	Transformed	p-value	{'alpha': 0.05, 'K': 48}	0.0
13	White Noise	Ljung-Box	Transformed	White Noise	{'alpha': 0.05, 'K': 24}	False
14	White Noise	Ljung-Box	Transformed	White Noise	{'alpha': 0.05, 'K': 48}	False
15	Stationarity	ADF	Transformed	Stationarity	{'alpha': 0.05}	False
16	Stationarity	ADF	Transformed	p-value	{'alpha': 0.05}	0.47168
17	Stationarity	ADF	Transformed	Test Statistic	{'alpha': 0.05}	-1.622002
18	Stationarity	ADF	Transformed	Critical Value 1%	{'alpha': 0.05}	-3.43397
19	Stationarity	ADF	Transformed	Critical Value 5%	{'alpha': 0.05}	-2.863139
20	Stationarity	ADF	Transformed	Critical Value 10%	{'alpha': 0.05}	-2.567621
21	Stationarity	KPSS	Transformed	Trend Stationarity	{'alpha': 0.05}	False
22	Stationarity	KPSS	Transformed	p-value	{'alpha': 0.05}	0.01
23	Stationarity	KPSS	Transformed	Test Statistic	{'alpha': 0.05}	0.931799
24	Stationarity	KPSS	Transformed	Critical Value 10%	{'alpha': 0.05}	0.119
25	Stationarity	KPSS	Transformed	Critical Value 5%	{'alpha': 0.05}	0.146
26	Stationarity	KPSS	Transformed	Critical Value 2.5%	{'alpha': 0.05}	0.176
27	Stationarity	KPSS	Transformed	Critical Value 1%	{'alpha': 0.05}	0.216
28	Normality	Shapiro	Transformed	Normality	{'alpha': 0.05}	False
29	Normality	Shapiro	Transformed	p-value	{'alpha': 0.05}	0.0

Рисунок 2.19 – Результат виконання коду на перевірку показників датасету, включаючи перевірку на стаціонарність

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

Тест Дікі-Фулера (ADF) показав значення тесту = -1,62, при критичних значеннях для 1% = -3.43; для 5% = -2.86; для 10% = -2.56.

Розраховане значення вище, ніж критичні значення, тобто можна вважати даний часовий ряд нестационарним. Додатково, за Квятковскі-Філіпс-Шмідт-Шин (KPSS) тестом також можна визначити нестационарність ряду, оскільки результат тесту 0.93 є вищим за критичні значення. Нестационарність часового ряду свідчить про те, що статистичні властивості ряду змінюються з часом, а це передбачає складність можливого ефективного моделювання.

Автокореляція та часткова автокореляція

Для дослідження автокореляції було проведено тест Льюнга-Бокса (Ljung-Box test). Він є статистичним тестом для перевірки автокореляції в залишках часового ряду після застосування моделі, наприклад, авторегресійної (AR) чи ковзного середнього (MA) моделі. Сутність цього тесту полягає в тому, щоб визначити, чи залишки мають статистично значущу автокореляцію, що може свідчити про те, що модель не враховує всі важливі патерни в даних. Для даного датасету було обрано значення кількості лагових днів – тридцять. Результати виконання тесту представлені на рис. 2.20.

Lag	Ljung-Box Statistic	p-value
1	1816.328379	0.0
2	3624.749181	0.0
3	5425.310394	0.0
4	7217.826953	0.0
5	9001.843794	0.0
6	10777.011765	0.0
7	12543.101146	0.0
8	14300.448502	0.0
9	16049.380893	0.0
10	17789.595537	0.0
11	19520.683124	0.0
12	21242.622340	0.0
13	22955.578431	0.0
14	24659.667770	0.0
15	26355.004785	0.0
16	28041.939606	0.0
17	29720.755758	0.0
18	31391.444060	0.0
19	33054.241308	0.0
20	34709.554771	0.0
21	36357.275480	0.0
22	37997.186070	0.0
23	39629.445609	0.0
24	41254.134568	0.0
25	42870.943188	0.0
26	44479.739155	0.0
27	46080.123371	0.0
28	47671.315398	0.0
29	49252.909674	0.0
30	50825.110266	0.0

Рисунок 2.20 – Результати тесту Льюнга Бокса

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

Результати дослідження автокореляції показали присутність не випадкових змін в значеннях часового ряду для всіх значень лагу від 1 до 30 днів, оскільки кожне

значення p-value менше за граничне значення (Ljung-Box Statistics) для тесту. Тобто зміну даних не можна вважати білим шумом, а дані тим чи іншим чином впливають на майбутні значення ціни. Часткова автокореляція для даного часового ряду представлена на рис. 2.21.

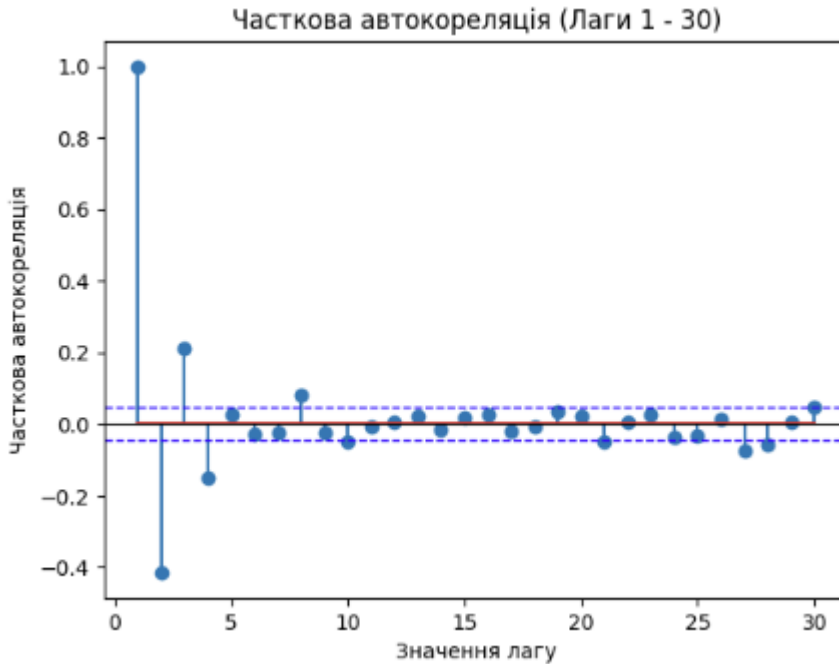


Рисунок 2.21 – Часткова автокореляція часового ряду зі значеннями лагу 1 – 30 днів

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

Дослідження часткової автокореляції вказує на значну залежність даних від попереднього дня, трохи менший, але присутній зв'язок між лагами у 2-4 днів, і низький зв'язок із даними більших значень лагу. Аналіз часового ряду вказує на те, що передбачати коливання цільової змінної досить складно через відсутність чітко вираженої сезонності та тренду. Дослідження часового ряду на автокореляцію вказало на наявність зв'язку між різними лаговими значеннями, особливо зі значеннями попередніх одного – чотирьох днів.

Опис прогностичних моделей

За допомогою пакету Rycaret було побудовано 26 моделей машинного навчання. Далі наведено опис деяких з них.

1. Theta Forecaster [48]

Метод Theta — це підхід до однофакторного прогнозування, який змінює локальну кривизну часового ряду за допомогою коефіцієнта Theta, застосованого до другої різниці даних. Цей коефіцієнт, позначений як θ , коливається від 0 до 1, де $\theta < 1$ вказує на дефляцію кривої, а $\theta > 1$ вказує на розширення. Оригінальна тета-модель передбачає створення двох тета-ліній ($\theta = 0$ і $\theta = 2$), а прогнози генеруються шляхом екстраполяції цих ліній. Динамічно оптимізована тета-модель забезпечує гнучкість, оновлюючи параметри моделі в кожен період часу, щоб мінімізувати помилки прогнозу в короткострокових і довгострокових компонентах. Метод застосовується до сезонних часових рядів після десезоналізації, і його ефективність оцінюється в сценаріях короткострокового прогнозування навантаження.

2. Extreme Gradient Boosting w/ Cond. Deseasonalize & Detrending

XGBoost може виконувати як завдання класифікації, так і регресії. Метод Extreme Gradient Boosting з умовною десезоналізацією та видаленням тренду є підходом до прогнозування, що використовує потужні градієнтні ансамблі дерев рішень для покращеного виділення сезонних та трендових патернів у даних. Застосування умовної десезоналізації та детрендингу дозволяє покращити якість прогнозів, враховуючи структурні особливості часових рядів. Хоча XGBoost має хорошу продуктивність, вона має також і недолік - ймовірність перенавчання, що обробляється шляхом експериментування з параметрами моделювання.

3. Random Forest w/ Cond. Deseasonalize & Detrending

Випадковий ліс — це широко використовуваний алгоритм машинного навчання, який поєднує результати кількох дерев рішень для досягнення єдиного результату. Його простота використання та гнучкість сприяли його популярності, оскільки він виконує як задачі класифікації, так і регресії. Метод Random Forest з умовною десезоналізацією та видаленням трендів є підходом до прогнозування, який

використовує потужні ансамблі випадкових дерев рішень для кращого виділення сезонних та трендових залежностей в даних.

4. Extra Trees w/ Cond. Deseasonalize & Detrending

Метод Extra Trees (Extremely Randomized Trees або Extremely Randomized Forest) є варіацією алгоритму випадкового лісу, де дерева рішень формуються за допомогою випадкового вибору підмножин ознак для розділу вузла. Відмінність Extra Trees в тому, що він використовує випадкові значення для вибору порогів розділення, навіть більш випадкові, ніж у звичайного Random Forest. Це може підвищити варіативність моделі та поліпшити узагальнення, зокрема в контексті аналізу часових рядів.

Багато з вбудованих моделей мають методи десеզонування та детрендування [49]. Десеզонування та детрендування часових рядів - це техніки, які можуть бути використані для видалення певних компонентів даних, які не є об'єктом інтересу при прогнозуванні майбутніх значень. Це може сприяти підвищенню точності прогнозів за рахунок зменшення впливу цих зайвих компонентів на модель. Десеզонування полягає в вилученні сезонних ефектів часового ряду, таких як регулярні коливання, що повторюються на певних інтервалах, наприклад, щорічно, щомісяця чи щотижня. Це дозволяє зосередити увагу на більш важливих тенденціях та паттернах у даних, які не пов'язані із сезонністю. Детрендування передбачає вилучення трендових компонентів, які вказують на загальні зміни у часовому ряді. Це може бути особливо корисно в тих випадках, коли дані мають тренди чи відомі тенденції зростання або зменшення з часом. Використання цих технік дозволяє зробити часовий ряд більш стаціонарним, що у свою чергу полегшує роботу моделі та допомагає отримати більш точні прогнози.

Перевагою бібліотеки Rucaret є те, що можна одночасно побудувати усі вищезазначені моделі для часового ряду. Для подальшого дослідження також було використано багатокроковий прогноз за допомогою LSTM (Long Short-Term Memory) - архітектури рекурентної нейронної мережі (RNN), яка широко використовується в

глибокому навчанні. Відносна нечутливість методу до довжини проміжків є його перевагою перед іншими RNN, прихованими моделями Маркова та іншими методами навчання послідовностей значень. Він має на меті забезпечити короткочасну пам'ять для RNN, яка може тривати багато кроків у часі, отже, «довгу короткочасну пам'ять». Він застосовується для класифікації, обробки та прогнозування даних на основі часових рядів.

Також з використанням глибокого навчання було проведено навчання моделі та побудова прогнозу методом Neural Prophet - гібридному фреймворку прогнозування, заснованому на PyTorch. Це розкладна модель часових рядів із компонентами, трендом, сезонністю, авторегресією, спеціальними подіями, майбутніми регресорами та лаговими регресорами. Майбутні регресори – це зовнішні змінні, які мають відомі майбутні значення для прогнозованого періоду, тоді як лагові регресори – це ті зовнішні змінні, які мають значення лише для періоду спостереження. Тренд можна змодельовати як лінійну або покроково-лінійну тенденцію за допомогою точок зміни. Сезонність моделюється за допомогою рядів Фур'є і, таким чином, може обробляти численні сезонності для високочастотних даних. Автоматична регресія обробляється за допомогою реалізації AR-Net [50], авторегресивної нейронної мережі прямого зв'язку для часових рядів. Лагові регресори також моделюються за допомогою окремих нейронних мереж прямого зв'язку. Інструмент автоматично вибирає гіперпараметри, що робить його зручним і доступним для початківців у роботі з машинним навчанням, а також він може бути корисним для учасників криптовалютного ринку, які планують включати методи машинного та глибокого навчання для покращення результатів своїх операцій.

Вибір показників оцінки якості моделей

Щоб порівняти ефективність різних моделей, необхідно виділити критерій оцінки. Використання для часового ряду показника коефіцієнта детермінації, який був обрахований при задачі множинної регресії, є недоцільним, оскільки наразі

датасет не має незалежних змінних, що впливають на модель. Для об'єктивного порівняння результатів прогнозу за різними методами було обрано показник середньої відсоткової абсолютної похибки (MAPE), середню абсолютну похибку (MAE) та середньоквадратична похибка (RMSE). MAPE є показником якості моделі, який вимірює середню абсолютну відсоткову похибку між фактичними і прогнозованими значеннями в часовому ряді. MAPE вимірює похибку у відсотках, що полегшує інтерпретацію точності прогнозу. MAE, в свою чергу середню абсолютну різницю між прогнозованими і фактичними значеннями, що також є корисним, для того, щоб дізнатися, в яких величинах помиляються моделі. RMSE також є метрикою, яка використовується для вимірювання точності прогнозів. Вона вказує на середню квадратичну помилку між прогнозованими значеннями і фактичними значеннями у тестовому наборі даних. Показники RMSE, MAE та MAPE розраховуються за формулами:

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

(2.22)

де n - кількість спостережень;

y_i - фактичне значення для i-го спостереження;

ŷ_i - прогнозоване значення для i-го спостереження.

$$MAPE = \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad ,$$

(2.23)

де n - кількість спостережень;

y_i - фактичне значення для i-го спостереження;

ŷ_i - прогнозоване значення для i-го спостереження.

$$RMSE = \sqrt{(\sum_{i=1}^n (y_i - \hat{y}_i)^2)}, \quad (2.23)$$

де n - кількість спостережень;

y_i - фактичне значення для i -го спостереження;

$\hat{y}_i$ - прогнозоване значення для i -го спостереження.

РОЗДІЛ 3

ЗАСТОСУВАННЯ МЕТОДІВ МАШИННОГО ТА ГЛИБОКОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ДИНАМІКИ ЦІН КРИПТОВАЛЮТ

3.1 Прогнозування часового ряду методами машинного навчання

В даному розділі представлений процес прогнозування ціни криптовалюти Bitcoin за допомогою методів машинного і глибокого навчання. Для прогнозування на один крок вперед використовуються методи Випадковий ліс та Багатошаровий пресептрон, описаними в другому розділі. Для прогнозування на декілька кроків вперед у часі, використовуються моделі з бібліотеки Pycaret, такі як Auto ARIMA [51], XGBoost, Theta Forecaster, Decision Tree, Polytrend та метод Prophet з фреймворку NeuralProphet мови програмування Python.

Спершу проведено прогнозування ціни криптовалюти Bitcoin технікою прогнозування на один крок вперед. Техніка однокрокового прогнозування розглядається лише для прогнозу на один період, і може бути неефективною для побудови значущих середньо- і довгострокових прогнозів. За цією технікою, модель навчається з використанням історичних даних до певного моменту часу, а потім використовується для створення прогнозів на наступний часовий крок. Після кожного передбачення фактичне спостережене значення для цього кроку часу стає частиною навчальних даних для наступного прогнозу. Він менш вимогливий до обчислень і забезпечує прогнози в реальному часі, що робить його придатним для сценаріїв, коли рішення потрібно приймати послідовно та швидко.

На відміну від задачі регресії, що розглядалася у попередньому розділі, в даному розділі розглядається саме задача прогнозування, в якій методи, що

використовуються, використовують попередні значення змінної, що прогнозується, в якості умовних предикторів невідомого значення.

В даному есперименті моделі будувалися, включаючи зазначену в другому розділі екзогенну змінну - індекс сентименту (FG_Index), щоб дослідити, чи покращить це прогностичну здатність моделей, а також без неї (враховуючи лише попередні значення ціни). Дані було розбито на тренувальну та тестову вибірки, а в якості предикторів розглядалися лагові значення змінних зі значеннями 7, 14 та 21 день відповідно.

Результати показників якості моделі – середньоквадратичні похибки (RMSE) та середні абсолютні відсоткові похибки (MAPE), представлені у табл. 1.

Таблиця 1 – Оцінка точності моделей методами Випадкового лісу та Багатошарового пресептрону

Метод	lag = 7		lag = 14		lag = 21	
	MAPE,%	RMSE	MAPE,%	RMSE	MAPE,%	RMSE
RF(Ціна+F&G)	2.97	878	3.11	898	3.83	1071
MLP(Ціна+F&G)	2.43	820	2.73	894	3.21	999
RF(Ціна)	2.99	885	3.29	932	4.15	1184
MLP(Ціна)	2.71	985	3.39	1174	3.53	1203

Як видно з табл.1, метод MLP, модель якого включає додаткову змінну сентименту F&G, має найменші MAPE та RMSE похибки при розглянутих величинах лагу. З цього можна зробити припущення, що по-перше, сентиментний фактор впливає на ціну криптовалюти і дозволяє краще спрогнозувати її значення, а по-друге, що метод MLP має перевагу над методом RF в задачі прогнозування на 1 крок вперед. Графік зі значеннями тестових(реальних) у порівнянні з прогнозованими значеннями ціни криптовалюти за методами Випадкового лісу (RF)

та Багатошарового персептрону (MLP) на проміжку 90 днів, представлений на рис. 3.1.



Рисунок 3.1 – Порівняння тестових та передбачених методами RF s MLP значень ціни

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

Як видно з рис. 3.1, обидві моделі досить непогано описують зміну криптовалюти в часі. Це зумовлюється саме тим, що під час моделювання, кожне наступне значення навчається на повному наборі тренувальних даних, додаючи до них вже відомі попередні тестові значення, а не передбачені, як у техніці прогнозування на декілька кроків вперед.

У багатокроковому прогнозуванні мета полягає в тому, щоб передбачити кілька майбутніх значень у часовому ряді, окрім лише наступного часового кроку. Подібно до однокрокового прогнозування, модель навчається на історичних даних, але в цьому випадку вона використовується для створення прогнозів для кількох майбутніх часових кроків. Прогнозовані значення потім використовуються як вхідні

дані для прогнозування наступних часових кроків. Техніка багатокрокового прогнозування може надати більш повне уявлення про майбутню тенденцію та корисна, коли потрібне довгострокове планування або прийняття рішень. Для прогнозування на багато кроків вперед, було використано бібліотеку `Rusaret` для одночасної побудови моделей машинного навчання. Для цього було встановлено необхідні пакети, а часовий ряд було розділено на навчальну і тестову вибірки та проведено ініціалізацію базових налаштувань для коректної роботи.

У початковій ініціалізації було вказано перехресну перевірку для покращення якості моделей і попередження перенавчання. У модулі `PyCaret` перехресна перевірка (cv) виконується за допомогою спеціальної техніки під назвою «`TimeSeriesSplit`» або «`Expanding Window Split`». Традиційні методи перехресної перевірки, такі як k -кратна перехресна перевірка, можуть бути непридатними для даних часових рядів, оскільки вони припускають, що точки даних є незалежними та однаково розподіленими. `TimeSeriesSplit` вирішує цю проблему, зберігаючи часовий порядок даних. Набір для початкового навчання використовується для навчання моделі під час першого згинання. Після початкового навчання набір для тестування розширюється, щоб включити наступну частину точок даних. Потім модель повторно навчається на розширеному навчальному наборі, а її продуктивність оцінюється на новому тестовому наборі. Цей процес повторюється, доки весь набір даних не буде використано для тестування. Кожна ітерація додає більше даних до навчального набору та оцінює модель на новому наборі точок даних. `PyCaret` розраховує показники ефективності (наприклад, MAE) для кожної ітерації та надає підсумок середнього значення та стандартного відхилення цих показників для різних складок. Ключова ідея полягає в тому, що модель навчається на історичних даних і перевіряється на майбутніх даних, симулюючи, як модель працюватиме при розгортанні в реальному сценарії, де прогнози робляться на невідомих майбутніх даних. На рис. 3.2 продемонстровано процес валідації для часового ряду.

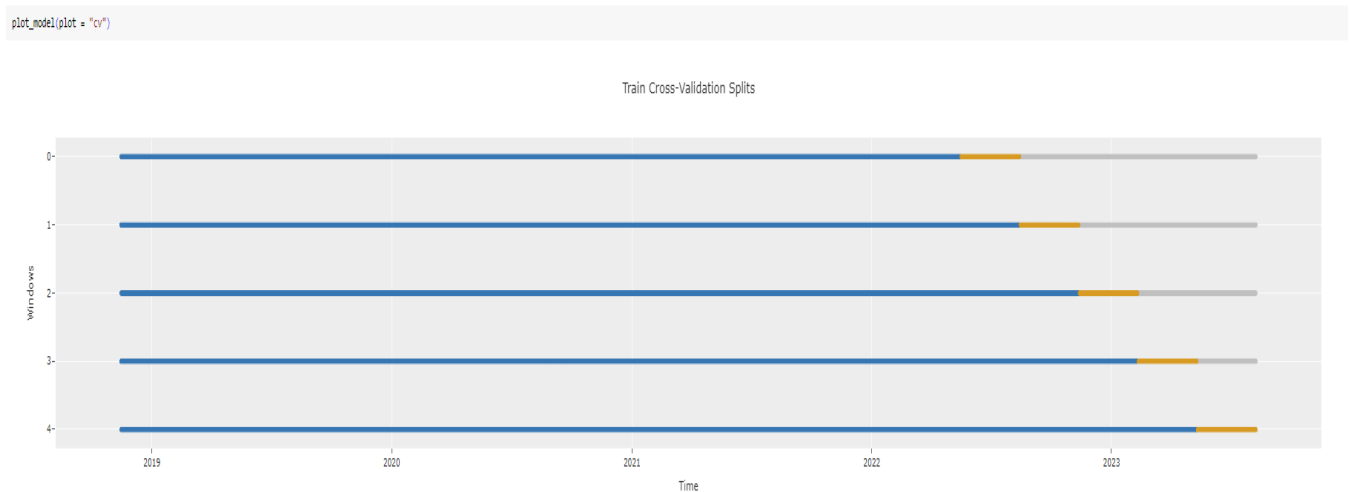


Рисунок 3.2 – Перехресна перевірка часового ряду

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

Вибір часового проміжку для прогнозування криптовалюти (або будь-якого фінансового інструменту) може бути обумовлений різними факторами. Часовим проміжком для прогнозування було обрано період в 90 днів. На криптовалютному ринку, так само як і в інших фінансових ринках, існують різні тривалості тренду та патернів. Проміжок в 90 днів дозволяє взяти до уваги достатньо даних для визначення змін в ринковому контексті, але не є надто довгим для того, щоб втратити актуальність у змінних умовах ринку. Дев'яносто днів є компромісом між короткостроковим та довгостроковим виглядами. Це дозволяє виявити короткострокові коливання, а також помітити тривалі тенденції. Тривалість в 90 днів може виявитися досить репрезентативною для взяття до уваги змін в цінах та поведінці ринку.

Визначивши базові параметри, такі як період прогнозування та кількість етапів крос-валідації, можна одразу ініціалізувати моделі в Rucaret. Після ініціалізації пакет дозволяє автоматично створити та порівняти 26 перелічених в другому розділі моделей машинного навчання. Результати побудованих моделей представлені на рис. 3.3.

	Model	MASE	RMSE	MAE	RMSE	MAPE	SMAPE	R2	TT (Sec)
auto_arima	Auto ARIMA	0.9887	0.7584	3771.1614	4491.7178	0.1638	0.1588	-3.6223	109.1940
naive	Naive Forecaster	1.0115	0.7698	3860.6509	4566.1696	0.1684	0.1602	-3.6201	1.0200
arima	ARIMA	1.0333	0.7803	3953.2466	4623.1902	0.1748	0.1706	-8.2640	1.9440
theta	Theta Forecaster	1.0495	0.7788	4015.3213	4631.0991	0.1783	0.1643	-4.4582	0.1260
grand_means	Grand Means Forecaster	1.0803	0.7858	4083.1094	4614.6892	0.1780	0.1748	-4.2125	0.0480
stlf	STLF	1.1130	0.7906	4265.0774	4716.9936	0.1978	0.1711	-4.6808	0.2260
croston	Croston	1.1172	0.7800	4263.1630	4638.1774	0.1898	0.1714	-2.9552	0.0620
huber_cds_dt	Huber w/ Cond. Deseasonalize & Detrending	1.1538	0.8688	4425.8951	5181.8493	0.2020	0.1845	-6.2273	5.5780
ada_cds_dt	AdaBoost w/ Cond. Deseasonalize & Detrending	1.1616	0.8360	4442.3501	4975.2770	0.2039	0.1825	-4.6158	6.5780
gbr_cds_dt	Gradient Boosting w/ Cond. Deseasonalize & Detrending	1.1715	0.8352	4489.9560	4979.6341	0.2038	0.1809	-5.8963	6.6100
rf_cds_dt	Random Forest w/ Cond. Deseasonalize & Detrending	1.2344	0.8758	4714.5202	5205.6827	0.2093	0.1932	-5.2167	9.6120
snaive	Seasonal Naive Forecaster	1.2632	0.9077	4820.6850	5403.0672	0.2139	0.1847	-3.9479	0.1300
omp_cds_dt	Orthogonal Matching Pursuit w/ Cond. Deseasonalize & Detrending	1.2682	0.9281	4894.8073	5568.4787	0.2303	0.1868	-13.9491	5.4120
exp_smooth	Exponential Smoothing	1.2788	1.0005	4863.0807	5918.5041	0.2102	0.2482	-5.7736	0.7960
xgboost_cds_dt	Extreme Gradient Boosting w/ Cond. Deseasonalize & Detrending	1.3559	0.9543	5193.3898	5696.6911	0.2373	0.2007	-6.1786	6.3580
lr_cds_dt	Linear w/ Cond. Deseasonalize & Detrending	1.3602	0.9952	5248.0884	5972.9998	0.2484	0.1979	-14.7853	5.8560
ridge_cds_dt	Ridge w/ Cond. Deseasonalize & Detrending	1.3602	0.9952	5248.0884	5972.9998	0.2484	0.1979	-14.7853	5.5760
llar_cds_dt	Lasso Least Angular Regressor w/ Cond. Deseasonalize & Detrending	1.3602	0.9952	5248.0881	5972.9978	0.2484	0.1979	-14.7853	5.5840
br_cds_dt	Bayesian Ridge w/ Cond. Deseasonalize & Detrending	1.3625	0.9963	5256.8718	5979.8877	0.2487	0.1981	-14.7835	5.4880
en_cds_dt	Elastic Net w/ Cond. Deseasonalize & Detrending	1.3688	0.9971	5281.6972	5985.6115	0.2497	0.1983	-14.8275	6.0940
lasso_cds_dt	Lasso w/ Cond. Deseasonalize & Detrending	1.3688	0.9971	5281.6957	5985.6093	0.2497	0.1983	-14.8275	5.5480
dt_cds_dt	Decision Tree w/ Cond. Deseasonalize & Detrending	1.3846	1.0105	5262.9762	5968.6507	0.2279	0.2156	-7.7954	5.6380
lightgbm_cds_dt	Light Gradient Boosting w/ Cond. Deseasonalize & Detrending	1.3898	0.9832	5305.0500	5852.7381	0.2391	0.2064	-5.0025	8.3780
et_cds_dt	Extra Trees w/ Cond. Deseasonalize & Detrending	1.4751	1.0460	5621.0995	6204.2497	0.2446	0.2221	-7.0579	9.1340
ets	ETS	1.4852	1.1533	5659.4857	6832.5587	0.2476	0.2939	-9.2105	1.2000
knn_cds_dt	K Neighbors w/ Cond. Deseasonalize & Detrending	1.6989	1.1940	6491.6095	7116.5301	0.2934	0.2373	-9.2664	5.7200
polytrend	Polynomial Trend Forecaster	5.8733	3.8065	22460.2734	22617.8403	1.0459	0.6488	-148.7261	0.0520

Рисунок 3.3 – Показники якості побудованих моделей у пакеті Rysaret

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

З результатів видно, що за більшістю показників модель Auto ARIMA мала перевагу над іншими моделями. Проте, наприклад, розглядаючи окремо значення середньої абсолютної відсоткової похибки (MAPE), використовуючи його, щоб об'єктивно оцінити якість моделі, бачимо, що помилка склала 16.4%, що є неідеальним, але цілком очікуваним результатом для багатокрокового прогнозу. AutoARIMA автоматизує процес визначення оптимальних значень для параметрів ARIMA (p, d, q) без ручного втручання. Вона використовує поетапний підхід, повторюючи різні комбінації цих параметрів і вибираючи модель, яка мінімізує вибраний критерій, такий як інформаційний критерій Акаїке (AIC) або байєсівський інформаційний критерій (BIC). Цей автоматичний пошук допомагає користувачам

знайти відповідну модель ARIMA для даних часових рядів без необхідності значного налаштування параметрів. Візуалізацію прогнозування даним методом та порівняння результатів з реальними даними представлено на рис. 3.4.

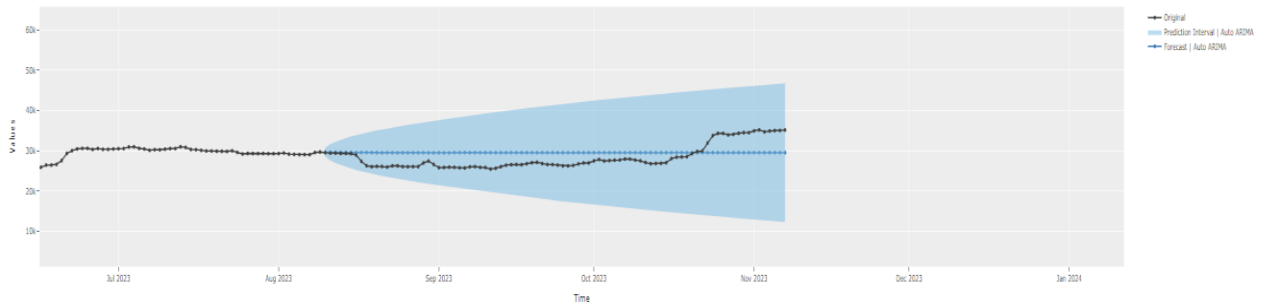


Рисунок 3.4 - Результати прогнозу за методом Auto ARIMA

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

Причина такого вигляду прогнозованих значень полягає в тому, що прогноз ARIMA для кожного часового кроку представляє собою лінійну комбінацію минулих значень часового ряду та попередніх прогнозів. У багатьох випадках прогнози можуть бути визначені лінійною залежністю від часу, що призводить до простої прямої лінії на графіку. Це особливо помітно, коли взяти до уваги, що ARIMA, по своїй суті, намагається знаходити лінійні залежності та патерни у часовому ряді. Такі моделі можуть бути ефективними для прогнозування лінійних чи лінійноподібних залежностей у даних. В той час вони можуть бути менш ефективним для деяких типів даних, які мають складні нелінійні структури або нестационарність, яку визначено в другому розділі.

Розглянемо декілька інших моделей та їх графіки прогнозованих значень ціни криптовалюти. Візуалізація результатів прогнозування методом XGBoost (Extreme Gradient Boosting w/ Cond. Deseasonalize & Detrending) представлена на рис. 3.5.

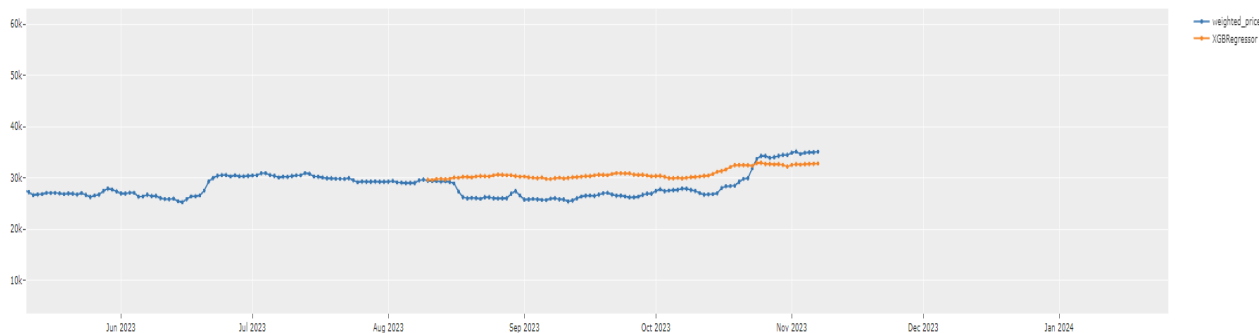


Рисунок 3.5 – Результати прогнозу за методом XGBoost

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

Як видно з рис. 3.5, модель непогано може помітити загальний напрямок руху ціни, але неточна в кількісній оцінці числових значень, в середньому помиляючись на близько п'ять тисяч доларів США від реальної ціни Bitcoin.

Розглянемо результати прогнозування методом Theta Forecaster, який серед інших моделей посів 4 місце за точністю результатів, на рис. 3.6.

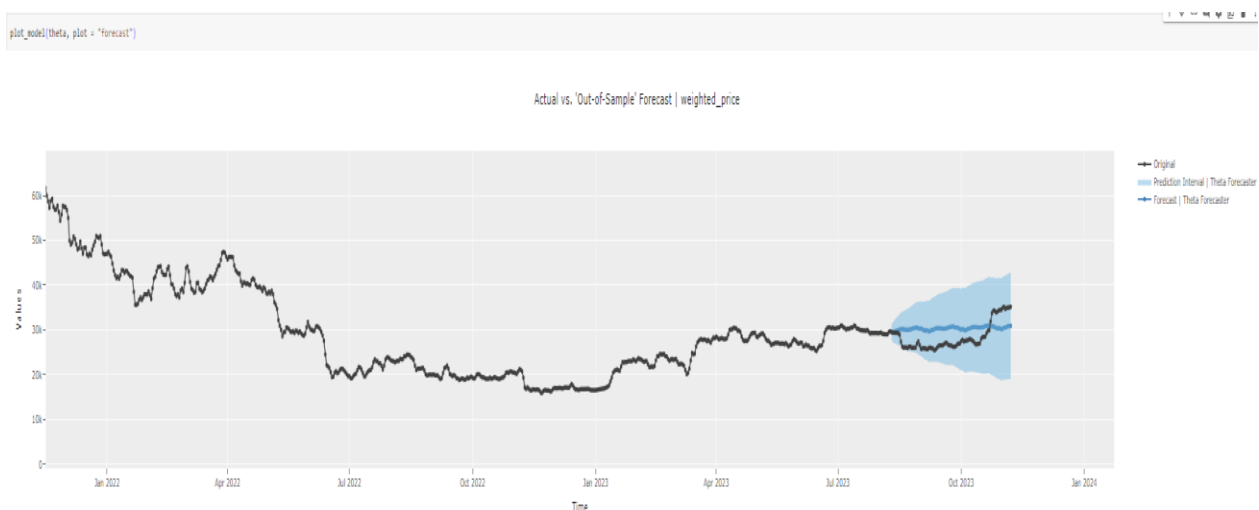


Рисунок 3.6 - Результати прогнозу за методом Theta Forecaster

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

В даному випадку, за візуальним аналізом, результат прогнозування є дещо схожим на результат попередньо розглянутої моделі. Проте на графіку видно менш стрімкий тренд на зростання ціни.

Розглянемо можливості покращення моделі Дерева рішень (Decision Tree w/ Cond. Deseasonalize & Detrending) за допомогою доналаштування (тюнінгу моделі), який включає підбір гіперпараметрів моделі [52], та без нього. Результати моделювання і графік прогнозу без тюнінгу представлені на рис. 3.7 та 3.8 відповідно.

```
model = create_model('dt_cds_dt')
```

	cutoff	MASE	RMSSE	MAE	RMSE	MAPE	SMAPE	R2
0	2022-05-16	2.1513	1.5711	8242.0719	9473.7133	0.3778	0.3008	-4.8662
1	2022-08-14	1.3954	0.9610	5517.4407	5843.8582	0.2822	0.2432	-19.2541
2	2022-11-12	0.5012	0.3599	1906.2910	2126.2850	0.1001	0.0996	0.3758
3	2023-02-10	1.5474	1.1118	5791.3313	6445.0682	0.2150	0.2472	-4.2184
4	2023-05-11	1.3274	1.0486	4857.7462	5954.3290	0.1642	0.1871	-11.0143
Mean	NaT	1.3846	1.0105	5262.9762	5968.6507	0.2279	0.2156	-7.7954
SD	NaT	0.5288	0.3877	2032.2632	2336.3215	0.0959	0.0682	6.7795

Рисунок 3.7 – Показники якості моделі Дерева рішень без тюнінгу

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

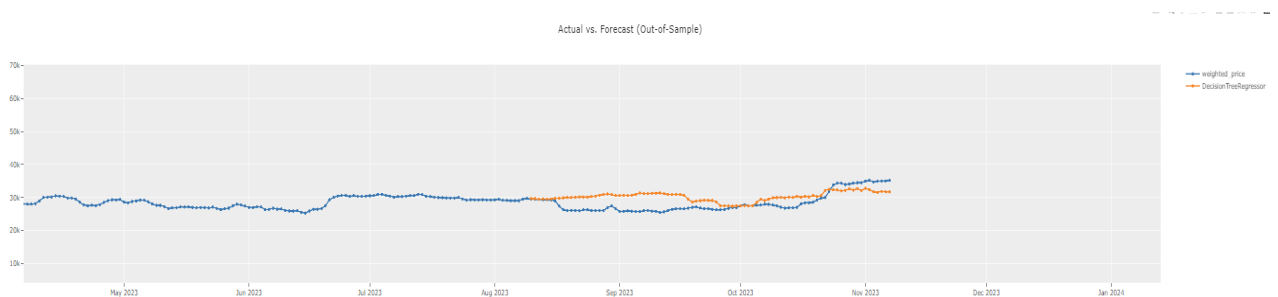


Рисунок 3.8 – Результати прогнозу за методом Дерева рішень без тюнінгу

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

Видно, що модель вловлює певні патерни динаміки ціни, а прогнозовані значення навіть наближаються до реальних значень приблизно на середині інтервалу прогнозування. Після цього рух ціни за прогнозом іде вгору, що дуже схоже на поведінку реальних показників ціни. Так, варто зазначити, що попри це, модель в більшості випадків помиляється за абсолютним значенням на декілька тисяч доларів, проте все ж таки, здатна описувати певну динаміку ціни в часі. Після проведення тюнінгу буде можливість порівняти його вплив на ефективність моделі.

На рис. 3.9 та 3.10 продемонстровано налаштування параметрів моделі та графік прогнозування за покращеною моделлю.

```
[ ] tuned_model = tune_model(model)
```

	cutoff	MASE	RMSSE	MAE	RMSE	MAPE	SMAPE	R2
0	2022-05-16	2.1248	1.5560	8140.3726	9382.4250	0.3736	0.2978	-4.7537
1	2022-08-14	1.4976	1.0243	5921.4946	6228.7195	0.3024	0.2586	-22.0097
2	2022-11-12	0.7013	0.5274	2667.4154	3115.9942	0.1511	0.1391	-0.3404
3	2023-02-10	0.8887	0.6570	3326.1202	3808.4346	0.1203	0.1295	-0.8221
4	2023-05-11	0.5126	0.4153	1875.7758	2358.0792	0.0632	0.0664	-0.8843
Mean	NaT	1.1450	0.8360	4386.2357	4978.7305	0.2021	0.1783	-5.7621
SD	NaT	0.5911	0.4143	2316.5129	2555.8634	0.1167	0.0862	8.2776

```

Fitting 5 folds for each of 10 candidates, totalling 50 fits
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 2 concurrent workers.
[Parallel(n_jobs=-1)]: Done 46 tasks | elapsed: 4.4min
[Parallel(n_jobs=-1)]: Done 50 out of 50 | elapsed: 4.8min finished

```

Рисунок 3.9 – Показники якості моделі Дерева рішень після тюнінгу

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

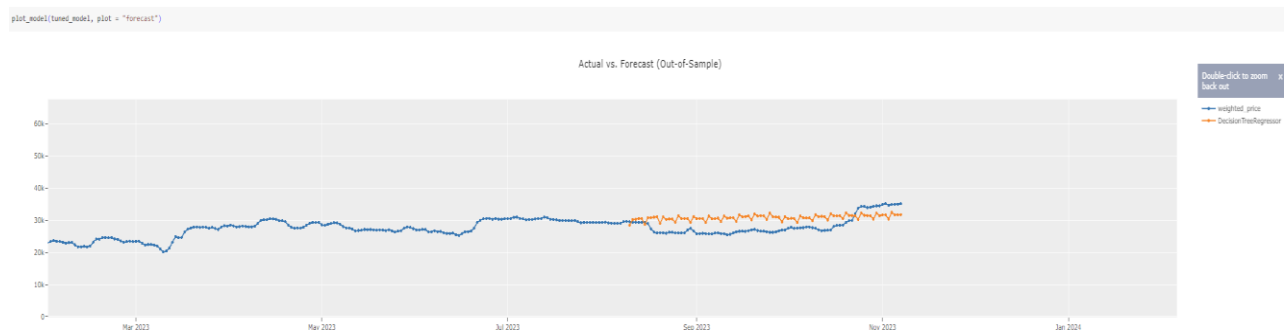


Рисунок 3.10 – Графік прогнозу методом Decision Tree після тюнінгу

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

З рис. 3.9 видно, що середня абсолютна відсоткова похибка покращилася з 22 % до 20%, а середня абсолютна похибка зменшилась майже на тисячу доларів. А з графіку на рис. 3.10 видно, що незважаючи на покращену якість моделі, прогноз майже не вловлює реальні зміни, а змінюється, скоріш випадковим чином. Проте, все одно, модель зважає на невеликий тренд ціни вгору.

Для порівняння якості прогнозів моделей, розглянемо менш точні моделі, наприклад найгіршу за показником MAPE модель Polytrend. Візуалізація результатів представлена на рис. 3.11.

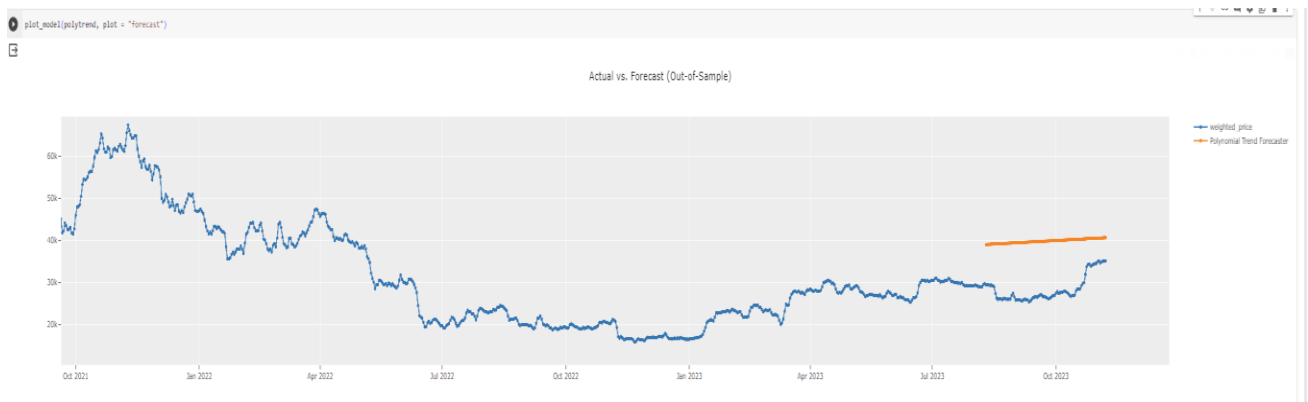


Рисунок 3.11 – Графік прогнозу за методом Polytrend

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

З рис. 3.11 видно, що модель помиляється на 5-10 тисяч доларів у різні часові проміжки у прогнозуванні ціни. Хоча, будучи найменш достовірною, модель все одно «вгадує» загальний тренд на незначне зростання ціни активу.

За результатами проведеного моделювання різними методами можна зазначити, що жодна модель не надає якісних кількісних вимірів у прогнозуванні ціни криптовалюти. Коефіцієнт детермінації кожної з моделей від'ємний, що вказує на нездатність достовірно кількісно описати зміну ціни криптовалюти. За критерієм MAE, кожна з моделей має абсолютну похибку прогнозу в декілька тисяч доларів, і навіть більше, що говорить про неефективність моделей у точності передбачення

ціни криптовалюти на довгих часових проміжках. Однак, більшість моделей може розпізнавати тренд на зростання, що може бути корисним у контексті довгострокового планування інвестиційного портфелю.

3.2 Прогнозування часового ряду методами глибокого навчання

Досліджуючи можливість техніки прогнозування на багато кроків вперед, постає питання можливості покращити результати, використовуючи методи глибокого навчання. Для цього було проведено розробку моделей LSTM – (Long Short-Term Memory) [53] та Neural Prophet.

LSTM або Довготривала Короткочасна Пам'ять (Long Short-Term Memory) — це тип рекурентної нейронної мережі (RNN), який спроектований для вирішення проблем довготривалих залежностей в даних. Це одна з архітектур глибокого навчання і застосовується, зокрема, у завданнях обробки послідовностей, прогнозування часових рядів, мовного моделювання та інших областях. LSTM використовує спеціальні структури, так звані вентиля, які допомагають регулювати потік інформації. Ці вентиля включають вентиль забуття (забуває непотрібну інформацію), вентиль входу (додає нову інформацію) та вентиль виходу (визначає, яка інформація буде передана далі). Завдяки своїй архітектурі, LSTM може ефективно працювати з послідовностями різної довжини та допомагає уникнути проблем обчислювальної неефективності. LSTM широко використовується в завданнях, де важлива обробка послідовностей, таких як прогнозування часових рядів, розпізнавання мови, машинний переклад та інші. Підсумовуючи, LSTM ставить за мету вирішення проблем, пов'язаних з довготривалими залежностями в даних, і дозволяє успішно моделювати та прогнозувати послідовності, що робить цей

метод придатним і для прогнозування ціни криптовалюти, незважаючи на її можливу волатильність.

При налаштуванні LSTM для прогнозування на більше ніж один крок вперед враховувалися декілька ключових аспектів. По-перше, кількість та конфігурація шарів нейронної мережі, що відіграють важливу роль у здатності моделі вивчати складні закономірності в даних. При побудові було використано два шари, кожен з яких містить 100 нейронів. Другий аспект - це крок вивчення (time steps), що визначає, скільки часових кроків використовується для кожного прикладу даних. Це може бути важливим для адекватного захоплення та урахування важливих патернів в часовому ряді. При побудові було використано проміжки кроків у співвідношенні 4 до 1, тобто 360 днів для вивчення до 90 для прогнозу. Функція активації також важлива - вона впливає на роботу внутрішніх воріт та виходів мережі. Було протестовано тангенсну функцію активації та `relu`, які описані в другому розділі. Додатково, налаштовуються параметри, такі як функція втрат і швидкість навчання, кількість епох для тренування.

Провівши необхідні налаштування моделі, можна розпочинати її виконання та аналіз результатів. На рис. 3.12 зображено проміжний датасет із результатами прогнозування (`yhat`), які співставляються реальним тестовим даним (`y`).

```

      y      yhat
0  29509.144045  33525.316
1  29398.477050  34601.883
2  29409.613280  34889.152
3  29351.687010  34725.830
4  29369.016602  34424.004
..      ...      ...
85  34687.677738  35352.450
86  34922.810547  33312.380
87  35018.487305  34903.820
88  35033.387693  33939.035
89  35149.707033  33475.945

[90 rows x 2 columns]
```

Рисунок 3.12 – Проміжний датасет для порівняння спрогнозованих LSTM значень ціни та реальних даних

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

Виведемо за допомогою бібліотеки Plotly графік для 90 днів прогнозу та реальні значення на рис. 3.13.

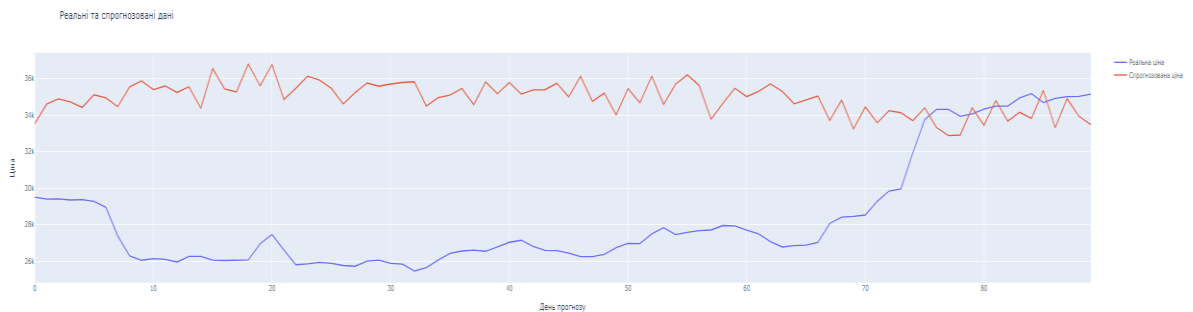


Рисунок 3.13 – Графік з реальними та спрогнозованими методом LSTM даними

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

З рисунку видно, що використаний метод, хоч і намагається знайти складні патерни динаміки ціни (верхня червона лінія), проте помиляється навіть у визначенні тренду – напрямку руху ціни, що свідчить про незначну ефективність, порівнюючи з методами, розглянутими раніше. Для того, щоб порівняти ефективність методу з іншими, потрібно визначити, наскільки вона помиляється у своїх прогнозах. Середня абсолютна похибка за методом LSTM розрахована на рис. 3.14.

```
df_compare['MAPE'] = (abs(df_compare['y'] - df_compare['yhat']) / df_compare['y']) * 100
mape = df_compare['MAPE'].mean()

print("Mean Absolute Percentage Error (MAPE):", mape)
```

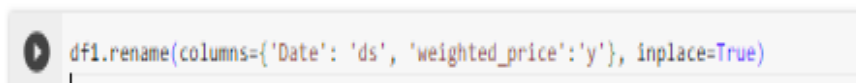
Mean Absolute Percentage Error (MAPE): 25.268325405332305

Рисунок 3.14 – Розрахунок MAPE для моделі LSTM

Джерело: розроблено автором у програмі Google Colaboratory на основі [37]

Результат MAPE у 25% свідчить про погіршення прогностичної здатності, порівнюючи з раніше розглянутими методами машинного навчання.

Для подальшого дослідження також був використаний фреймворк Neural Prophet, що був описаний в другому розділі. Він цілеспрямовано розроблений для прогнозування часових рядів. Для налаштування роботи Neural Prophet, стовпці датасету необхідно перейменувати: Дата має називатися “ds”, значення часового ряду – “y”, як продемонстровано на рис. 3.15.

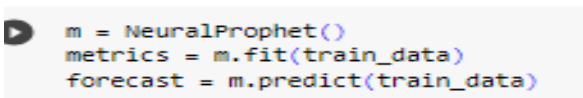


```
df1.rename(columns={'Date': 'ds', 'weighted_price': 'y'}, inplace=True)
```

Рисунок 3.15 – Підготовка датасету до роботи в Neural Prophet

Джерело: розроблено автором у програмі Google Colaboratory на основі [36,37]

Після підготовки даних до моделювання проводиться ініціалізація моделі, код для виконання цього зображений на рис. 3.16.



```
m = NeuralProphet()
metrics = m.fit(train_data)
forecast = m.predict(train_data)
```

Рисунок 3.16 – Код для ініціалізації Neural Prophet

Джерело: розроблено автором у програмі Google Colaboratory на основі [36,37]

В процесі ініціалізації, Neural Prophet підбирає гіперпараметри моделі по епохам – покроково зменшуючи похибки. Результати метрик після 162 кроків налаштування моделі провізуалізовано на рис. 3.17.

metrics					
	MAE	RMSE	Loss	RegLoss	epoch
0	56491.761719	68251.890625	0.743231	0.0	0
1	46051.148438	56102.992188	0.588987	0.0	1
2	36281.996094	44640.617188	0.443819	0.0	2
3	26741.982422	33266.156250	0.305181	0.0	3
4	18375.599609	22902.582031	0.184174	0.0	4
...	...	...	...	...	...
157	3548.333496	4723.614258	0.010169	0.0	157
158	3543.826416	4700.219238	0.010111	0.0	158
159	3534.047363	4689.639160	0.010057	0.0	159
160	3530.847900	4683.177246	0.010008	0.0	160
161	3551.796875	4696.371094	0.010142	0.0	161

162 rows × 6 columns

Рисунок 3.17 – Зображення покрокового покращення метрик моделі

Джерело: розроблено автором у програмі Google Colaboratory на основі [36,37]

Бачимо, що наприклад, середньоквадратична похибка під час навчання моделі на тренувальних даних зменшилася майже в 15 разів, а середня абсолютна похибка в результаті склала 3500. На рис. 3.18 зображено розраховані значення ціни BTC під час тренування (показник y_{hat1}):



Рисунок 3.18 – Розраховані значення ціни для тренувальних даних за методом NeuralProphet в порівнянні з реальними даними

Джерело: розроблено автором у програмі Google Colaboratory на основі [36,37]

Для оцінки точності прогнозування в період дев'яноста днів, оцінимо відмінність тестових даних від спрогнозованих. Для цього спочатку побудуємо прогноз, вказавши модель, тренувальні дані та період прогнозування (рисунок 3.19):

```
[ ]
df1_future = m.make_future_dataframe(train_data, periods=90)
forecast = m.predict(df1_future)
forecast90 = forecast['yhat1']
```

Рисунок 3.19 – Побудова прогнозу на 90 днів

Джерело: розроблено автором у програмі Google Colaboratory на основі [36,37]

Для зручності створимо окремий Pandas датасет із датою, реальним та спрогнозованим значенням ціни активу у стовпцях, результат виведено на рис. 3.20:

```
[34] actual_values.head()
```

	ds	y	yhat1
1729	2023-08-10	29509.144045	23773.093750
1730	2023-08-11	29398.477050	23782.216797
1731	2023-08-12	29409.613280	23876.039062
1732	2023-08-13	29351.687010	23918.472656
1733	2023-08-14	29369.016602	24253.939453

Рисунок 3.20 – Датасет тестових даних із включенням значень прогнозу

Джерело: розроблено автором у програмі Google Colaboratory на основі [36,37]

Необхідно визначити такі метрики, як значення середньоквадратичної, середньої абсолютної відсоткової похибки, та зобразити відмінності між реальними та спрогнозованими значеннями. Результат продемонстровано на рис. 3.21:

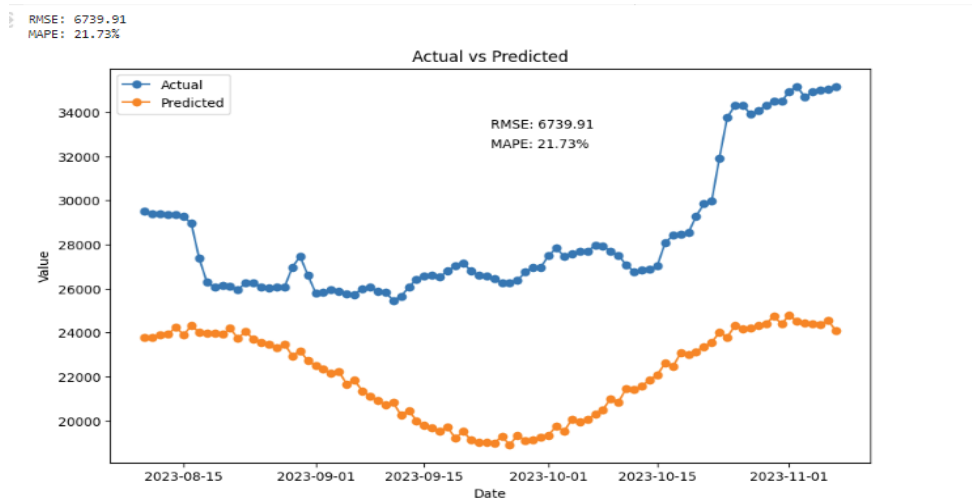


Рисунок 3.21 – Результат прогнозу за методом NeuralProphet

Джерело: розроблено автором у програмі Google Colaboratory на основі [36,37]

Синім кольором позначені реальні дані, помаранчевим – спрогнозовані значення. Середня абсолютна похибка склала близько 21%. Цей результат є кращим за деякі розглянуті вище методи, проте не свідчить про високу якість кількісного прогнозування. Модель змогла «вгадати» просадку ціни та наступне її збільшення, проте значно помилилась саме у величині цієї просадки та значеннях ціни.

Проведені дослідження щодо моделювання криптовалюти Bitcoin та результати багатокрокового прогнозування ціни свідчать про те, що в умовах значної волатильності ринку, середньо-і довгострокове прогнозування є складним процесом, що вимагає вибору правильних інструментів та методів. Розглянуті методи машинного та глибокого навчання в своїй більшості дають змогу побудувати достовірний прогноз лише динаміки руху активу, тобто визначати напрямок на зростання чи спадання ціни. Майже усі розглянуті методи не дають якісного прогнозування щодо оцінки значень ціни криптовалюти. Показники середніх абсолютних похибок прогнозованих значень варіюються між близько 3700\$ у найефективнішій моделі (Auto ARIMA) та 22000\$ у найменш ефективній (Polytrend). У відсотковому вираженні ці помилки становлять від 16.4% до більш ніж 100% у найменш ефективних. В медіані розглянуті методи помиляються на близько 20%.

ВИСНОВКИ

Створення та розвиток цифрових валют сприяє стрімким змінам у тенденціях грошового обігу. Процеси з регулювання та визначення правових аспектів криптовалютного ринку в більшості країн світу знаходяться у фазі розробок та доопрацювань, і ставлення офіційних органів країн різняться від повної підтримки таких нововведень, до абсолютного несприйняття та заборон на державному рівні. Дослідження та аналіз динаміки цін на криптовалютному ринку є важливим для осіб, що приймають участь в торгівлі, утримуванні та іншими операціями з даною формою цифрових активів, щоб приймати зважені та обґрунтовані рішення. Завдяки сучасним інструментам та технологіям, включаючи мови програмування, фреймворки та бібліотеки, веб-ресурси та додатки, - проведення таких досліджень може полегшити процес прийняття рішень про операції з криптовалютами. У роботі описуються ключові аспекти та процеси, що відбуваються на криптовалютному ринку та впливають на нього. У роботі детально досліджується, аналізується та прогнозується часовий ряд і визначаються ключові його компоненти, метрики та характеристики.

Традиційні методи аналізу динаміки ринку можуть використовуватися та доповнюватися методами машинного навчання. Вибір методу прогнозування залежить від таких факторів, як якість та достовірність даних, долучення додаткових (екзогенних) факторів в модель, обчислювальних можливостей, аналіз порівняння результатів підходів та моделей тощо.

Розглянуті методи аналізу та прогнозування часового ряду криптовалюти Bitcoin загалом можуть допомогти особам, що приймають рішення про взаємодію з тим чи іншим активом на криптовалютному ринку, побачити загальні патерни та динаміку руху ціни криптовалюти. Але необхідно зазначити, що використання їх в даному дослідженні в багатокроковому довгостроковому прогнозуванні може не бути корисним у передбаченні саме кількісного вираження цінових значень. Такі

особливості можна пояснити певною хаотичністю формування ціни більшості криптовалют, оскільки вони піддаються сильному впливу волатильності через їхні особливості. Серед них - новизна та недостатній регулювальний контроль, обмежена масштабність та ліквідність, вплив подій та новин, спекуляції та психологія учасників ринку, відсутність підтримки у вигляді реальних активів тощо.

Серед розглянутих методів найбільш ефективним на період 90 днів за технікою багатокрокового прогнозування виявився Auto ARIMA. Інші розглянуті методи мали більші похибки. В процесі дослідження також було проведено прогнозування ціни Bitcoin технікою прогнозу на один крок вперед методами Random Forest (RF) та Multilayer Perceptron (MLP). Хоча ефективність моделей була доведена непоганими оцінками точності, дану техніку не варто розглядати в якості ефективного рішення для середньо- і довгострокових інвестицій, оскільки за нею прогноз будується окремо кожен раз лише на 1 період часу.

Для дослідження того, чи будуть впливати певні екзогенні фактори на формування ціни криптовалюти, було проведено експеримент з додаванням двох додаткових змінних – індексу настрою та обсягу торгів. Моделювання, використовуючи ці фактори за допомогою моделі лінійної регресії, методів RF та MLP не дало якісних результатів. Отже, дані предиктори не сильно впливають на значення ціни, хоча значення обсягу торгів і має суттєвий статистичний вплив на значення ціни. Теоретично, показники емоційного стану гравців ринку, що дуже залежить від емоційної складової, мав би більше впливати на пояснення цінових рухів. Проте, волатильність криптовалют загалом і Bitcoin зокрема, не дає нам змоги покладатися на цей фактор. Показник щоденного об'єму торгів, хоча і має статистично значущий зв'язок, також не може достатньою мірою описувати зміну ціни. Для подальших досліджень можуть розглядатися такі варіанти, як побудова мультифакторного прогнозу з більшою кількістю релевантних дескрипторів, важливість яких для моделі, можна буде визначити. Для цього необхідно буде

виділяти якісні і кількісні чинники, які стосуються дотичних до криптовалютного ринку подій та показників, доповнюючи інформаційну базу дослідження.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Бебешко Б.Т. Аналіз методів та моделей прогнозування ринку цифрових криптовалют. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка»*. 2022. № 2.18. С. 163–174.
2. Міловська К.М., Мороз В.В. Вейвлетний аналіз та прогнозування фінансових часових рядів. *Вісник Національного технічного університету "ХПІ". Сер. : Інформатика та моделювання*. 2022. № 1-2 (7-8). С. 117-127.
3. Derbentsev V., Matviychuk A., Soloviev V.. Forecasting of Cryptocurrency Prices Using Machine Learning. *Advanced Studies of Financial Technologies and Cryptocurrency Markets*. Springer. 2020. P. 211-231.
4. Haritha GB & Sahana N. B, 2023. Cryptocurrency Price Prediction using Twitter Sentiment Analysis. URL: <https://doi.org/10.48550/arXiv.2303.09397>. (accessed 01.12.2023).
5. Galushka E., Pakon, O. The essence of cryptocurrencies and prospects for their development. *Young scientist*. 2017. № 4, (44). P. 634-638.
6. Nick Darlington. What is Blockchain Technology? A Step-by-Step Guide For Beginners. URL: <https://blockgeeks.com/guides/what-is-blockchain-technology/> (accessed 01.12.2023)
7. Bernard Marr. 35 Amazing Real World Examples Of How Blockchain Is Changing Our World. URL: <https://www.forbes.com/sites/bernardmarr/2018/01/22/35-amazing-real-world-examples-of-how-blockchain-is-changing-our-world/> (accessed 17.12.2023)

8. S. N.Khan, F. Loukil, C. Ghedira-Guegan, E. Benkhelifa, A. Bani-Hani. Blockchain smart contracts: Applications, challenges, and future trends. *Peer-to-Peer Netw. Appl.* 14. 2021. P. 2901–2925.
9. CERTA PAYMENTS LTD. Волатильність криптовалют: що це таке. URL: <https://cryptomus.com/uk/blog/cryptocurrencies-volatility-what-is-it> (дата звернення 01.12.2023)
10. Derevyanko B., Turkot O. Legal regulation of cryptocurrency relations with in Ukraine: EU experience. *Entrepreneurship, Economy and Law*. №3, 2022. P. 56–61.
11. Grobys K., Junttila J. Speculation and lottery-like demand in cryptocurrency markets. *Journal of International Financial Markets, Institutions and Money*. V. 71, 101289. 2021.
12. Tapscott D., Tapscott A. Blockchain Revolution: How the Technology Behind Bitcoin is Changing Money, Business, and the World. *New York: Penguin Random House*, 2016.
13. Varun K., Sombuddha C., Vinay C., Kuldip Singh S., Sherali Z. An Analysis of Energy Consumption and Carbon Footprints of Cryptocurrencies and Possible Solutions. *Digital Communications and Networks*. V. 9, I.1, 2023. P. 79-89.
14. Chainalysis team. 2022 Biggest Year Ever For Crypto Hacking with \$3.8 Billion Stolen, Primarily from DeFi Protocols and by North Korea-linked Attackers Beginners. URL: <https://www.chainalysis.com/blog/2022-biggest-year-ever-for-crypto-hacking/> (accessed: 01.12.2023)
15. Arner D., & Zetsche D., Buckley R., Kirkwood J.. The Financialization of Crypto: Lessons from FTX and the Crypto Winter of 2022-2023. *SSRN Electronic Journal*. №19. 2023.
16. Apolline Blandin, Cloots A.S., Hussain H., Rauchs M., Saleuddin R., Allen J.G., Brian. Z.Z., Cloud K. Global Cryptoasset Regulatory Landscape Study. *Cambridge Centre for Alternative Finance Reports*. Report. 2019.

17. Adachi M., Cominetta M., Kaufmann C., Van der kraaij A. (2020). A regulatory and financial stability perspective on global stablecoins, *Macroprudential Bulletin*. Issue 10. Publication. 2020.

18. FINMA - Swiss Financial Market Supervisory Authority. FINMA publishes ICO guidelines. URL: www.finma.ch/en/news/2018/02/20180216-mm-icowegleitung/. (accessed: 01.12.2023)

19. Lukianchuk R.. (2023). The current trends of the development of the domestic crypto sector in the conditions of the martial law. *Information and law*. №2(45). 2023. P. 113-124.

20. Sabry, Farida ; Labda, Wadha ; Erbad, Aiman ; Malluhi, Qutaibah; Cryptocurrencies and Artificial Intelligence: Challenges and Opportunities. *IEEE Access*. Vol.8. 2020. P. 175840-175858.

21. Naganjaneyulu G., Prashanth G., Revanth M., Narasimhadhan A.V. Multi Indicator based Hierarchical Strategies for Technical Analysis of Crypto market Paradigm. *International journal of electrical and computer engineering systems*. Vol.14. №7. 2023. P. 765-780.

22. Chen James. What is EMA? How to Use Exponential Moving Average With Formula. URL: <https://www.investopedia.com/terms/e/ema.asp> (accessed 01.12.2023)

23. Javier Pastor. What is fundamental analysis of cryptocurrency markets? URL: (accessed 01.12.2023)

24. Constantine Belov. How AI Will Affect Cryptocurrency Market? URL: <https://b2broker.com/news/how-ai-will-affect-cryptocurrency-market/> (accessed 01.12.2023)

25. Jaquart P., Köpke S., Weinhardt C. Machine learning for cryptocurrency market prediction and trading. *The Journal of Finance and Data Science*. Volume 8. 2022. P. 331-352.

26. Kala Philo. AI and Crypto Trading: How Machine Learning is Disrupting Financial Markets. URL: <https://www.zenledger.io/blog/ai-and-crypto-trading-how-machine-learning-is-disrupting-financial-markets/> (accessed 01.12.2023)
27. Hibaq O., Lester A. L. Uncover Social Media Interactions On Cryptocurrencies Using Social Set Analysis (SSA). *Procedia Computer Science*, Volume 219. 2023. P. 161-169.
28. Pérez J. Rajngewerc M., Giudici J., Furman D., Luque F., Alonso Alemany L., Martinez M. V. Pysentimiento: A Python Toolkit for Opinion Mining and Social NLP tasks. *arXiv preprint arXiv:2106.09462*. 2021.
29. Kenton W. Beta: Definition, Calculation, and Explanation for Investors. URL: <https://www.investopedia.com/terms/b/beta.asp>. (accessed 01.12.2023).
30. Бібліотека Posibniki. Графічний аналіз дохідності та ризику. URL: <https://posibniki.com.ua/post-grafichnii-analiz-dohidnosti-ta-riziku-rinku-kapitaliv-rinku-cinnih>. (дата звернення: 01.12.2023).
31. Toai T., Senkerik R., Zelinka I., Ulrich A., Hanh V., Huan V. ARIMA for Short-Term and LSTM for Long-Term in Daily Bitcoin Price Prediction. *Lecture Notes in Computer Science()*. Springer, Cham. V. 13588. 2023. P. 131–143
32. Hitam N. A., Ismail A. R., Saeed F. An Optimized Support Vector Machine (SVM) based on Particle Swarm Optimization (PSO) for Cryptocurrency Forecasting. *Procedia Computer Science*. V. 163. 2019. P. 427-433.
33. Li Y., Qiu T., Tang Q. Pisces: Private and Compliant Cryptocurrency Exchange. *Cryptology ePrint Archive 1317*. 2023. 27c.
34. Bloomberg Finance L.P. Bloomberg Crypto Data, Research & Tools. URL: <https://www.bloomberg.com/professional/solution/bloomberg-crypto-data-research-tools/>. (accessed 01.12.2023).
35. CryptoPredictions. Cryptocurrency Price Predictions for 2023, 2024-2026 Tools. URL: <https://cryptopredictions.com/>. (accessed 01.12.2023).

36. Triebe O. NeuralProphet. URL: <https://neuralprophet.com/contents.html>. (accessed 01.12.2023).
37. CoinGecko. Global Cryptocurrency Market Cap Charts. URL: <https://www.coingecko.com/en/global-charts> (accessed 15.11.2023).
38. Yahoo Finance/AOL Finance. URL: <https://finance.yahoo.com/quote/BTC-USD/history/>. (accessed 15.11.2023).
39. Alternative.me. Crypto Fear & Greed Index - Bitcoin Sentiment. URL: <https://alternative.me/crypto/fear-and-greed-index/>. <https://finance.yahoo.com/quote/BTC-USD/history/>. (accessed 15.11.2023).
40. Kharval A. StandardScaler in Machine Learning. URL: <https://thecleverprogrammer.com/2020/09/22/standardscaler-in-machine-learning/>. (accessed 15.11.2023).
41. Sutton B. Scikit-learn vs. StatsModels: Which, why, and how? URL: <https://www.thedataincubator.com/blog/2022/12/01/scikit-learn-vs-statsmodels/>. (accessed 15.11.2023).
42. Blakely L., Reno M.J., Broderick R. Evaluation and Comparison of Machine Learning Techniques for Rapid QSTS Simulations. Sandia National Laboratories. 2018.
43. Paola Benedetti. Multi Layer Perceptron. URL: <https://medium.com/@paolabenedetti1009/multi-layer-perceptron-63b3a1dd47af>. (accessed 15.11.2023).
44. Płoński P. Random Forests vs Neural Networks: Which is Better, and When? URL: <https://www.kdnuggets.com/2019/06/random-forest-vs-neural-network.html>. (accessed 15.11.2023).
45. Oloruntoba A., Ismail S. Break for Time Series Components (BFTSC) and Group for Time Series Components (GFTSC) in Identification of Time Series Components in Univariate Forecasting. *TEST Engineering & Management*. Volume 81,. 2019. P. 2826-2844.

46. Vojtko R., Javorská J. The Seasonality of Bitcoin. *Social science research network*. 2023.
47. Gitbook Inc. Time Series Forecasting with PyCaret Regression. URL: <https://pycaret.gitbook.io/docs/learn-pycaret/official-blog/time-series-forecasting-with-pycaret-regression>. (дата звернення: 01.12.2023).
48. Assimakopoulos V., Nikolopoulos K. The theta model: a decomposition approach to forecasting. *International Journal of Forecasting*. Volume 16, Issue 4. 2000. P. 521-530.
49. Vinay. Explaining the Math| Assumptions | Limitations | Examples of Popular Time Series 27 Forecasting Methods. URL: <https://medium.com/@vinay1996/explaining-the-assumptions-and-limitations-of-popular-time-series-27-forecasting-methods-3a6240ced7aa>. (дата звернення: 01.12.2023).
50. Triebe O., Laptev N., Rajagopal R. AR-Net: A simple Auto-Regressive Neural Network for time-series. *arXiv:1911.12436*. 2019.
51. Taylor G Smith. Pmdarima.arima.auto_arima. URL: https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.auto_arima.html (дата звернення: 01.12.2023).
52. Gitbook Inc. Optimize - Optimization functions in PyCaret. URL: https://pycaret.gitbook.io/docs/get-started/functions/optimize#tune_model (дата звернення: 01.12.2023).
53. Fjellström, C. Long Short-Term Memory Neural Network for Financial Time Series. *2022 IEEE International Conference on Big Data (Big Data)*. Institute of Electrical and Electronics Engineers (IEEE). 2022. P. 3496-3504.

КОРОТКИЙ ЗВІТ ПОДІБНОСТІ

Назва документа: Кайка_Притоманова_плагіат



Ім'я користувача:
Математичного моделювання та статистики Даценк...

ID перевірки:
1015974775

Дата перевірки:
06.12.2023 06:02:16 ЕЕТ

Тип перевірки:
Doc vs Internet + Library

Дата звіту:
06.12.2023 06:08:54 ЕЕТ

ID користувача:
100005744

Кількість сторінок: 76 Кількість слів: 11320 Кількість символів: 89396 Розмір файлу: 1.66 MB ID файлу: 1015654241

5.9%

Схожість

Найбільша схожість: 1.4% з Інтернет-джерелом (<https://incrypted.com/ua/scho-take-likvidnist-kriptoalyt>)

...С.Т.О.Р.І.Н.К.А...7.8...

0% Цитат

...С.Т.О.Р.І.Н.К.А...7.9...

Вилучення цитат вимкнено

Вилучення списку бібліографічних посилань вимкнено

0%

Вилучень

Немає вилучених джерел

Модифікації

Виявлено модифікації тексту. Детальна інформація доступна в онлайн-звіті.

Назва документа: Кайка_Притоманова_плагіат ID файлу: 1015654241