

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»**

**Навчально-науковий інститут прикладного системного аналізу
Кафедра математичних методів системного аналізу**

До захисту допущено:

Завідувач кафедри

_____ Оксана ТИМОЩУК

«__» _____ 2025 р.

Дипломна робота

на здобуття ступеня бакалавра

за освітньо-професійною програмою «Системний аналіз і управління»

спеціальності 124 «Системний аналіз»

на тему: «Аналіз впливу блокчейн транзакцій

на курс криптовалюти»

Виконав:

студент IV курсу, групи КА-15

Кривоус Максим Олександрович _____

Керівник:

Ас. каф. ММСА, Канцедал Георгій Олегович _____

Консультант з економічного розділу:

Доцент, к.е.н., Рощина Надія Василівна _____

Консультант з нормконтролю:

Доцент каф. ММСА, к.ф.-м.н., Статкевич Віталій Михайлович _____

Рецензент:

Проф. каф. ШІ., д.т.н., проф. Данилов Валерій Якович _____

Засвідчую, що у цій дипломній роботі
немає запозичень з праць інших авторів
без відповідних посилань.

Студент _____

Київ – 2025 року

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Навчально-науковий інститут прикладного системного аналізу
Кафедра математичних методів системного аналізу

Рівень вищої освіти – перший (бакалаврський)

Спеціальність – 124 «Системний аналіз»

Освітньо-професійна програма «Системний аналіз і управління»

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ Оксана ТИМОЩУК

«___» _____ 2025 р.

ЗАВДАННЯ

на дипломну роботу студенту

Кривоусу Максиму Олександровичу

1. Тема роботи «Аналіз впливу блокчейн транзакцій на курс криптовалюти», керівник роботи Канцедал Георгій Олегович, асистент кафедри ММСА, затверджені наказом по університету від 26.05.2025 № 1759-с.
2. Термін подання студентом роботи: 11.06.2025 р.
3. Вихідні дані до роботи: 272-денна історія ончейн-показників та відповідні щоденні ціни BTC.
4. Зміст роботи: дослідження предметної області; теоретичні основи аналізу впливу блокчейн транзакцій на курс криптовалюти; програмна реалізація та експериментальні результати; функціонально-вартісний аналіз програмного забезпечення для прогнозування курсу BTC.
5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо): ілюстрації до теоретичної частини, ілюстрація кореляційних залежностей, ілюстрації порівняння прогнозування різних моделей з реальними значеннями, таблиці та графіки до функціонально-вартісного аналізу.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Економічний	Рощина Н.В., доцент, к.е.н.		

7. Дата видачі завдання _____

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1.	Вибір теми та постановка задачі	14.04.2025	Виконано
2.	Аналіз актуальності, постановка мети та завдань дослідження	21.04.2025	Виконано
3.	Ознайомлення з літературними джерелами та сучасними підходами	28.04.2025	Виконано
4.	Збір даних: ончейн-метрики, історія курсів	05.05.2025	Виконано
5.	Побудова та тренування моделей прогнозування	12.05.2025	Виконано
6.	Аналіз результатів, оцінка точності, оформлення висновків	19.05.2025	Виконано
7.	Оформлення пояснювальної записки	26.05.2025	Виконано
8.	Оформлення презентації	02.06.2025	Виконано

Студент

Максим КРИВОУС

Керівник

Георгій КАНЦЕДАЛ

РЕФЕРАТ

Дипломна робота: 124 с., 6 рис., 5 табл., 2 дод., 38 джерел.

БЛОКЧЕЙН, ТРАНЗАКЦІЇ, КРИПТОВАЛЮТА, ПРОГНОЗУВАННЯ, МАШИННЕ НАВЧАННЯ, LIGHTGBM, SHAP, VAR.

Тема: Аналіз впливу блокчейн транзакцій на курс криптовалюти.

Робота присвячена дослідженню взаємозв'язку між ончейн-метриками (характеристиками блокчейн-транзакцій) та зміною курсу біткоїна. Проведено комплексний аналіз набору даних, що містить 272-денну історію ончейн-показників та відповідні щоденні ціни BTC.

Об'єкт дослідження – вплив ончейн-метрики (характеристик блокчейн-транзакцій) на динаміку ринкової ціни біткоїна.

Предмет дослідження – методи статистичного аналізу (кореляція, VAR) і алгоритми машинного навчання для побудови та інтерпретації прогнозних моделей курсу BTC на основі блокчейн-даних.

Мета роботи – розробити та апробувати знеперервні моделі прогнозування одномоментної зміни ціни біткоїна ($\Delta Price$) на базі ончейн-показників, а також оцінити їхню точність і пояснити внесок кожної ознаки в прогноз.

Виконано обробку ончейн-даних та історичних даних цін біткоїна. Порівняно класичні часові моделі (ARIMA, VAR) та методи машинного навчання (LightGBM, CatBoost), з використанням автокореляційних лагів і ковзних середніх. Експериментальне дослідження показало, що модель LightGBM демонструє найкращі результати ($RMSE \Delta price \approx 1201 \text{ USD}$), а SHAP-аналіз виявив ключовий вплив попередньої ціни та ончейн-метрики. Результати свідчать про обмежену здатність ончейн-факторів прогнозувати раптові стрибки курсу і окреслюють напрямки подальших досліджень.

ABSTRACT

Diploma thesis: 124 pages, 6 figures, 5 tables, 2 appendices, 38 references.
BLOCKCHAIN, TRANSACTIONS, CRYPTOCURRENCY, FORECASTING,
MACHINE LEARNING, LIGHTGBM, SHAP, VAR.

Topic: Analysis of the impact of blockchain transactions on the cryptocurrency exchange rate.

The thesis deals with the researching of relationship between on-chain metrics (characteristics of blockchain transactions) and changes in the Bitcoin exchange rate. A comprehensive analysis of a dataset containing 272 days of on-chain metrics and corresponding daily BTC prices was conducted.

Object of study: the impact of on-chain metrics (characteristics of blockchain transactions) on the dynamics of the market price of Bitcoin.

Subject of study: statistical analysis methods (correlation, VAR) and machine learning algorithms for building and interpreting predictive models of the BTC exchange rate based on blockchain data.

The purpose of the work: to develop and test continuous models for forecasting instantaneous changes in the price of Bitcoin (ΔPrice) based on on-chain indicators, as well as to evaluate their accuracy and explain the contribution of each indicator to the forecast.

On-chain data and historical Bitcoin price data were processed. We compared classical time series models (ARIMA, VAR) and machine learning methods (LightGBM, CatBoost) using autocorrelation lags and moving averages. The experimental study showed that the LightGBM model demonstrates the best results ($\text{RMSE } \Delta\text{price} \approx 1201 \text{ USD}$), and SHAP analysis revealed the key influence of the previous price and on-chain metrics. The results indicate the limited ability of on-chain factors to predict sudden price jumps and outline directions for further research.

ЗМІСТ

ВСТУП.....	8
РОЗДІЛ 1 ДОСЛІДЖЕННЯ ПРЕДМЕТНОЇ ОБЛАСТІ.....	10
1.1 Актуальність проблеми.....	10
1.2 Визначення та базові поняття.....	11
1.2.1 Блокчейн: принцип роботи та транзакції.....	11
1.2.2 Криптовалюта Bitcoin: економічна природа та курсоутворення....	12
1.3 Теоретичні зв'язки між активністю у блокчейні та ціною криптовалюти	14
1.4 Огляд досліджень і аналізів щодо взаємозв'язку ончейн-активності та	
ціни Bitcoin.....	17
1.5 Історичні приклади впливу активності блокчейну на курс Bitcoin.....	20
1.6 Висновки до розділу 1.....	23
РОЗДІЛ 2 ТЕОРЕТИЧНІ ОСНОВИ АНАЛІЗУ ВПЛИВУ	
БЛОКЧЕЙН-ТРАНЗАКЦІЙ НА КУРС КРИПТОВАЛЮТИ.....	24
2.1 Кореляційний аналіз: коефіцієнти Пірсона та Спірмена.....	24
2.2 Тести на стаціонарність: ADF та KPSS.....	26
2.3 Модель Vector Autoregression (VAR).....	30
2.4 Моделі машинного навчання для аналізу часових рядів і табличних	
даних.....	33
2.4.1 LightGBM (Light Gradient Boosting Machine).....	33
2.4.2 CatBoost (Categorical Boosting).....	36
2.4.3 Багатошаровий перцептрон (MLP).....	40
2.5 Метрики оцінки якості прогнозу.....	42
2.6 Висновки до розділу 2.....	49

РОЗДІЛ 3 ПРОГРАМНА РЕАЛІЗАЦІЯ ТА ЕКСПЕРИМЕНТАЛЬНІ

РЕЗУЛЬТАТИ.....51

3.1 Вступ.....51

3.2 Опис вхідних даних.....52

3.3 Опис ознак.....55

3.4 Експерименти та результати.....59

3.4.2 VAR-модель для оцінки впливу транзакцій на ціну.....62

3.4.3 Налаштування моделей.....65

3.5 Висновки до розділу 3.....80

РОЗДІЛ 4. ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ПРОГНОЗУВАННЯ КУРСУ ВТС.....84

4.1 Постановка задачі проєктування.....84

4.2 Обґрунтування функцій програмного продукту.....85

4.3 Обґрунтування системи параметрів.....90

4.4 Аналіз експертного оцінювання параметрів.....94

4.5 Аналіз рівня якості варіантів реалізації.....95

4.6 Економічний аналіз розробки програмного забезпечення.....98

4.7 Вибір кращого варіанта.....103

4.8 Висновки до розділу 4.....105

ВИСНОВКИ.....106

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ.....108

ДОДАТОК А: ЛІСТИНГ112

ДОДАТОК Б: ПРЕЗЕНТАЦІЙНІ МАТЕРІАЛИ.....118

ВСТУП

Аналіз впливу блокчейн–транзакцій на курс криптовалюти, зокрема біткоїна, є однією з найактуальніших тем сучасних досліджень у галузі фінансової аналітики та прикладної економіки. З моменту появи Bitcoin у 2009 р. [1] криптовалюти перетворилися з приватних експериментів у масштабний фінансовий інструмент із капіталізацією понад сотні мільярдів доларів США та високою волатильністю. Відмінність криптовалют від традиційних фінансових активів полягає в тому, що блокчейн відкрито фіксує кожну транзакцію в розподіленому реєстрі, а будь-який користувач може аналізувати в мережі поведінку учасників, структуру переказів та інші ончейн-метрики.

Попит і пропозиція на ринку біткоїна залежать не лише від загальних макроекономічних чинників та настроїв інвесторів, а й від «внутрішнього стану» мережі: обсягів транзакцій, активності адрес, частки великих переказів («китів»), ставок фінансування на деривативному ринку тощо. У теоретичному аспекті це узгоджується із законом Меткаффа, згідно з яким «вартість» мережі зростає з числом її учасників, а також із моделлю NVT (Network Value to Transactions), що порівнює ринкову капіталізацію мережі з реальним обсягом транзакцій як аналог Р/Е-коефіцієнта для традиційних активів.

Метою даної роботи є кількісне оцінювання впливу основних ончейн-показників на динаміку курсу біткоїна та виявлення тих метрик, які дозволяють підвищити точність прогнозування цінового руху. У практичній частині побудовано набір із понад десяти ознак – від базових (ціна, кількість і обсяг транзакцій, число активних адрес) до похідних (7-денні ковзні середні, частка великих транзакцій, індикатор аномалії) і настроєвих (funding rate). Ці

дані за 272 дні охоплюють періоди як відносної стабільності, так і різких сплесків волатильності.

Для прогнозування використовувався ансамблевий метод градієнтного бустингу LightGBM, який за рекомендаціями попереднього аналізу показав найкращу продуктивність серед альтернатив (CatBoost, MLP, ARIMA). Результати оцінено на hold-out вибірці, проведено SHAP-інтерпретацію важливості ознак, а для вивчення динамічних взаємозв'язків застосовано VAR-модель з аналізом імпульсних відгуків. Окремо здійснено виявлення аномалій у щоденних змінах ціни та вивчено їхній вплив на прогнозу точність моделі.

Наукова новизна роботи полягає в комплексному поєднанні ончейн-метрик із сучасними статистичними (кореляційний і VAR-аналіз) й машинно-навчальними методами (LightGBM із SHAP-інтерпретацією), що дає змогу глибше зрозуміти роль мережевої активності в процесі ціноутворення криптовалют та створити практично застосовну модель для короткострокового прогнозування. Отримані результати сприятимуть розробці аналітичних панелей моніторингу блокчейн-показників і розширенню інструментарію алгоритмічної торгівлі на ринку криптовалют.

РОЗДІЛ 1 ДОСЛІДЖЕННЯ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Актуальність проблеми

Аналіз впливу активності у блокчейні на ринковий курс криптовалют є важливим і актуальним з кількох причин. По-перше, криптовалюти на кшталт біткойна стали значним фінансовим активом із великою капіталізацією та волатильністю, і розуміння чинників, що впливають на їх ціну, має як наукове, так і практичне значення. На відміну від традиційних фінансових активів, біткойн не має матеріального забезпечення чи регульованої прибутковості, його економічна цінність формується попитом і пропозицією на відкритому ринку. Інвестори й аналітики постійно шукають фундаментальні показники, які б дозволили оцінити «справедливу» вартість криптовалют або прогнозувати цінові коливання. Ончейн-метрики, такі як кількість транзакцій у мережі, обсяг переданих коштів, число активних адрес, великі перекази з гаманців тощо, є відкритими даними з блокчейну і часто розглядаються як потенційні індикатори стану мережі та ринку. Дійсно, дослідники вже вивчають ончейн-дані (обсяги транзакцій, активність майнерів тощо) як можливі провісники майбутніх цінових трендів криптовалют. Розуміння теоретичних зв'язків між такими показниками та ціною біткойна може допомогти глибше усвідомити природу курсоутворення криптовалют, а також покращити моделі прогнозування і методи оцінки ризиків. По-друге, прозорість блокчейну (усі транзакції публічно доступні) створює унікальну ситуацію, коли дослідники та трейдери мають змогу в режимі реального часу відстежувати великі переміщення коштів чи зміни в активності мережі, що може безпосередньо впливати на ринкові очікування.

Наприклад, великі «whale»-транзакції (перекази значних обсягів BTC) часто привертають увагу спільноти і здатні викликати короткострокові

коливання ціни через очікування потенційних продажів або інших дій власників таких коштів. Таким чином, тема впливу блокчейн-транзакцій на курс біткойна є актуальною як з точки зору розвитку економічної науки про криптоактиви, так і для практичних цілей учасників ринку.

1.2 Визначення та базові поняття

1.2.1 Блокчейн: принцип роботи та транзакції

Блокчейн - це децентралізована розподілена база даних (реєстр), що безпечно і незмінно фіксує транзакції між сторонами без посередників. Технологію блокчейну було вперше створено 2008 року спеціально для біткойна - першої криптовалюти - як відкриту книгу обліку всіх операцій, що вирішує проблему подвійної витрати без потреби в центральному органі чи сервері. У контексті біткойна блокчейн реалізований як ланцюжок блоків транзакцій, де кожен новий блок містить запис останніх підтверджених транзакцій у мережі. Копія цього ланцюга з усією історією транзакцій зберігається на безлічі вузлів (незалежних комп'ютерів) по всьому світу, що синхронізуються за формальними правилами консенсусу.

Транзакція в блокчейні це передача певного цифрового активу (наприклад, біткойнів) від одного учасника мережі до іншого. Процес здійснення транзакції в мережі Bitcoin виглядає так: коли відправник (Олексій) ініціює переказ біткойнів одержувачу (Максиму), ця транзакція транслюється до мережі і розсилається всім вузлам-блокчейн. Кожен вузол перевіряє дійсність транзакції, зокрема, криптографічні підписи і наявність достатнього балансу на адресі відправника. Після верифікації транзакція потрапляє до пулу непідтверджених транзакцій (mempool) і очікує включення

в блок майнерами. Майнери (в випадку біткойна це учасники, що виконують доказ роботи) збирають транзакції у новий блок, змагаючись у розв'язанні криптографічної задачі. Щойно майнер знаходить валідний блок (підтверджує його Proof-of-Work), цей блок додається до ланцюга блокчейну, а всі транзакції в ньому вважаються підтвердженими. Кожен блок можна уявити як чергову “сторінку” книги обліку, що містить записи про транзакції. Завдяки механізму консенсусу (набору правил, за якими вузли досягають згоди щодо черговості блоків і дійсності транзакцій) забезпечується єдина версія правдивого ланцюга і захист від фальсифікації даних [2]. Таким чином, блокчейн гарантує незмінність записів: після додавання блоку з транзакціями в загальний реєстр змінити ці дані неможливо без згоди більшості вузлів мережі.

Основні властивості блокчейну: децентралізація, прозорість і захищеність записів, роблять його привабливим для фінансових застосувань. У контексті криптовалют це означає, що транзакції можуть виконуватися безпосередньо між користувачами без участі банків чи інших посередників, а вся історія платежів у мережі є відкритою для перегляду (через спеціальні блокчейн-оглядачі). Наприклад, будь-хто може переглянути баланс тієї чи іншої публічної адреси або простежити рух коштів у мережі біткойна. Ця відкритість даних і породила можливість ончейн-аналізу, який використовується для оцінки активності мережі та потенційних впливів на ринок криптовалют (розглянемо це докладніше у наступних підрозділах).

1.2.2 Криптовалюта Bitcoin: економічна природа та курсоутворення

Біткойн (BTC) це перша й найбільш відома криптовалюта, запущена в обіг у січні 2009 року невідомим розробником (чи групою) під псевдонімом Сатоші Накамото. На відміну від фіатних грошей, біткойн не випускається

центрального банку і не забезпечений державними резервами або дорогоцінними металами. Економічна природа біткойна полягає у тому, що це цифровий товар із заздалегідь обмеженою емісією та децентралізованим механізмом емісії і обігу. Протокол біткойна встановлює максимальну кількість монет - 21 мільйон BTC, більше якої ніколи не існуватиме. Нові монети поступово входять в обіг через винагороду майнерам за знайдені блоки, і ця нагорода зменшується вдвічі приблизно кожні 4 роки (подія halving) [38]. Таким чином, пропозиція біткойнів зростає все повільніше, що додає дефіцитності цьому активу. Обмежена емісія та складність видобутку (майнінгу) означають, що біткойн часто порівнюють із цифровим золотом.

Ціна біткойна формується вільно на криптовалютних біржах і позабіржових ринках, базуючись виключно на співвідношенні попиту і пропозиції. Ніхто не гарантує біткойну певної вартості і жодна установа не несе зобов'язань щодо підтримання курсу. Якщо попит на BTC зростає, при обмеженій пропозиції це призводить до зростання ціни; якщо ж більше власників прагнуть продати за незмінної кількості покупців, ціна падає. На попит впливають різні чинники: довіра та прийняття користувачами, регуляторні новини, макроекономічні умови, а також внутрішні характеристики мережі [19] (дефіцитність, безпека, децентралізація). Унікальні властивості біткойна, такі як: децентралізація, надійність захисту транзакцій, обмежена емісія - створюють йому репутацію надійного засобу збереження вартості [20] та інвестиційного інструменту, що приваблює все більше інвесторів. Це, своєю чергою, підвищує попит. З іншого боку, відсутність внутрішньої «власної вартості» (intrinsic value) і повна залежність ціни від колективних очікувань ринку означає високу волатильність курсу біткойна [22]. Ціна BTC може суттєво коливатися протягом коротких проміжків часу, реагуючи як на зовнішні події [21] (економічні кризи, зміни регуляцій, новини про зломи бірж чи великі покупки з боку фондів), так і на внутрішні метрики мережі (наприклад, наближення чергового halving чи різкі

зміни активності на блокчейні). Саме ця можливість зв'язку між ончейн-активністю та поведінкою ціни і є предметом нашого дослідження.

1.3 Теоретичні зв'язки між активністю у блокчейні та ціною криптовалют

Існують кілька концептуальних моделей та гіпотез, що пояснюють, яким чином активність у блокчейні (мережеві показники) може бути пов'язана з ринковою ціною криптовалюти [34] (зокрема біткойна). Багато з цих ідей спираються на аналогії з економічними законами про цінність мережі чи використання активу.

1. Ефект мережі та закон Меткаффа. Відповідно до закону Меткаффа (Metcalfe's Law), цінність мережі пропорційна квадрату кількості її користувачів. У випадку біткойна це означає, що чим більше у мережі активних учасників, тим вища фундаментальна вартість BTC як мережевого активу. Дійсно, деякі дослідники відзначають, що довгострокова динаміка ціни біткойна узгоджується із законом Меткаффа - модель, де ціна залежить від квадратичної функції числа користувачів (наприклад, числа активних адрес або гаманців), демонструє високу відповідність реальним даним. Іншими словами, історично ринкова капіталізація біткойна зростала приблизно як квадрат від збільшення кількості учасників мережі, що підтверджує значущість мережевого ефекту у формуванні ціни. Кількість активних адрес (адрес, з яких здійснювалися або на які надходили транзакції протягом певного періоду) розглядають як наближений показник кількості активних користувачів. Сплески цього показника часто свідчать про зростання інтересу до криптовалюти. Спостереження показують, що піки активних адрес за часом збігаються з піками ціни BTC, причому іноді

підвищення активності адрес дещо випереджає розворот ринку. Це можна інтерпретувати так: коли все більше учасників залучені до мережі і відбувається багато транзакцій, ринок досягає перегріву (максимуму попиту), після чого можливий розворот тренду.

2. Обсяг транзакцій та використання мережі. Інший підхід базується на ідеї, що обсяг переданих через блокчейн коштів (сумарна вартість транзакцій) відображає реальну корисність і використовуваність криптовалюти як засобу обміну. На цій основі було запропоновано метрику Network Value to Transactions (NVT) ratio - співвідношення ринкової капіталізації мережі до щоденного обсягу переданих через блокчейн коштів. NVT часто називають аналогом коефіцієнта P/E (ціна/прибуток) для біткойна [5]: він показує, наскільки «дорого» оцінений біткойн відносно економічної активності в його мережі. Високе значення NVT означає, що ринкова вартість випереджає обсяги транзакцій (мережа «дорога» відносно переданої через неї вартості) це може бути ознакою перегрітого, переоціненого ринку. Низький NVT, навпаки, свідчить, що обсяги транзакцій великі порівняно з капіталізацією - тобто біткойн може бути недооцінений, і такий стан часто спостерігається поблизу ринкового дна або на початку нового зростаючого тренду. Таким чином, теорія «фундаментальної цінності» біткойна через призму ончейн-активності стверджує, що якщо все більше вартості щодня передається мережею (багато транзакцій на значні суми), це має підкріплювати і вищу справедливую ціну криптовалюти. Відповідно, коли ціна різко зростає без аналогічного росту транзакційної активності, можна підозрювати спекулятивний «бульбашковий» характер такого зростання. Крім NVT, аналітики використовують похідні метрики на основі транзакцій (наприклад, NVT Signal, Adjusted NVT тощо) для більш тонкого аналізу ринкових стадій.

3. Великі перекази («whale alerts») та структура ринку. У той час як попередні підходи розглядали сукупні показники активності, ще один важливий аспект: поодинокі великі транзакції, які можуть вказувати на дії

великих гравців ринку (так званих «китів»). Переміщення значних сум BTC між адресами, особливо якщо вони ідентифіковані як гаманці бірж або відомих великих власників, здатні впливати на короткострокові очікування трейдерів. Масові переводи біткойнів на біржі часто трактуються як намір продати велику позицію, що може створити тиск вниз на ціну в короткі терміни. Навпаки, виведення значних сум з бірж на особисті гаманці розглядається як сигнал наміру довгостроково зберігати актив (HODL), що потенційно знижує пропозицію на ринку і є позитивним сигналом для ціни. Таким чином, аналіз потоків монет (inflow/outflow) на біржі є частиною ончейн-аналітики, пов'язаної з ціноутворенням. Історичні спостереження підтверджують, що великі транзакції можуть передувати підвищеній волатильності: наприклад, за даними аналітичних компаній, масштабні переміщення «сплячих» біткойн-гаманців часто призводили до короткострокових просідань ціни ~2-3% протягом 1-2 днів після такого руху. Це пояснюється тим, що ринок закладає можливий продаж цих монет. В цілому, великі гравці через свої ончейн-дії можуть впливати на ринок, тому моніторинг таких транзакцій став невід'ємною частиною аналізу криптовалютного ринку.

Підсумовуючи, існують обґрунтовані теоретичні зв'язки між метриками блокчейну та ціною: показники активності мережі відображають рівень попиту, прийняття і використання криптовалюти, тому можуть бути індикаторами фундаментальної цінності або сигналами змін у балансі попиту і пропозиції на ринку. Наступний розділ присвячено тому, як ці теоретичні міркування підтверджуються (або спростовуються) в наукових дослідженнях та аналітичних звітах.

1.4 Огляд досліджень і аналізів щодо взаємозв'язку ончейн-активності та ціни Bitcoin

Питанню залежності між ончейн-метриками та ринковим курсом біткойна приділено увагу як у академічній літературі, так і в аналітичних оглядах індустрії. Наукові дослідження останніх років вивчають статистичні взаємозв'язки між різними показниками діяльності блокчейну і ціною BTC, намагаючись виявити причинно-наслідкові зв'язки або побудувати моделі прогнозування.

Значна частина робіт підтверджує існування кореляцій між мережевою активністю та ціною. Наприклад, у роботі A.Saunders (2021) було проведено регресійний аналіз за 2015-2020рр., який показав, що такі фактори, як кількість активних адрес, хеш-рейт майнінгу та складність майнінгу, статистично значуще і позитивно пов'язані з ціною біткойна (чим вищі ці показники, тим вищий курс), тоді як відносна пропозиція монет (частка монет у вільному обігу) має негативний зв'язок з ціною [3]. Іншими словами, зростання числа користувачів та активності мережі сприяє зростанню вартості BTC, тоді як збільшення пропозиції (наприклад, вихід на ринок раніше «заморожених» монет) тисне на ціну вниз, що узгоджується з економічною інтуїцією. У висновках цього дослідження зазначено, що внутрішні метрики біткойна можуть слугувати рушійними факторами його довгострокової вартості.

Інші академічні роботи розглядають питання прогнозування ціни криптовалют на основі ончейн-даних. Наприклад, застосовуються методи машинного навчання і глибоких нейромереж, які враховують ончейн-показники поряд з ринковими даними [23]. Відзначається, що включення таких змінних, як обсяг транзакцій, комісії мережі, число нових адрес, може підвищити точність моделей [24], оскільки ончейн-метрики

несуть інформацію про довгострокову динаміку та поведінку учасників мережі. Деякі дослідники навіть стверджують, що окремі показники мають передбачувальну силу: зокрема, зростання хеш-рейту і витрат майнерів може передувати росту ціни (через підвищення собівартості майнінгу [25] і, відповідно, мінімально обґрунтованої ціни), а піки в кількості нових або активних адрес можуть сигналізувати про приплив нових інвесторів перед черговим ралі на ринку [26].

Варто згадати і концепцію фундаментальної оцінки біткойна, яку пропонують на основі ончейн-метрик. Окрім згаданого вище закону Меткаффа, існують роботи, що моделюють «справедливу» ціну BTC, виходячи з використання мережі. Так, T.Peterson (2018) показав, що модель, де ціна біткойна пропорційна квадрату числа користувачів (гаманців) з урахуванням уповільнення емісії, дає дуже високий коефіцієнт детермінації (R^2) і добре описує середньострокові коливання ціни [4]. Інші дослідники запропонували показник $MV = \text{Network Value} / \text{Metcalf Value}$, що відслідковує відхилення ринкової капіталізації від значення, прогнозованого законом Меткаффа: суттєве перевищення може свідчити про спекулятивний перегрів ринку, а заниження про потенціал зростання.

Не лише в академічних колах, але й в аналітичних звітах індустрії часто розглядається взаємодія між ончейн-активністю та ціною. Компанії, що спеціалізуються на аналітиці блокчейну (Glassnode, Chainalysis, Coin Metrics тощо), регулярно публікують метрики і індекси, які пов'язують стан мережі з ринком. Наприклад, згаданий вище коефіцієнт NVT широко використовується аналітиками для оцінки, чи не занадто швидко зросла ціна у порівнянні з «економічною активністю» в мережі. Інший відомий показник це MVRV (Market-Value-to-Realized-Value), який порівнює поточну ринкову капіталізацію з сумарною вартістю монет за ціною на момент їх останнього руху (Realized Cap) [7], фактично також пов'язує ончейн-дані (середню ціну придбання монет) з поточною ціною для оцінки перегріву або паніки на

ринку. Такі метрики показали свою корисність при аналізі ринкових циклів біткойна.

Оглядаючи медіа-джерела і аналітичні статті, можна знайти численні приклади, де аналітики роблять висновки про ринок, спираючись на ончейн-дані. В оглядах CoinDesk, Cointelegraph та інших видань часто зазначається, що “кількість активних адрес досягла максимуму, небаченого з часу попереднього піку цін” або що “зростання комісій за транзакції співпало з бичачим ралі, свідчаючи про підвищений попит на використання мережі”.

Крім того, обговорюються дії великих гравців: криптовалютні ЗМІ регулярно повідомляють про пробудження «сплячих» гаманців (які зберігали монети багато років) чи про масові перекази з бірж/на біржі, пов’язуючи ці події з рухами ціни [8]. Такий міждисциплінарний підхід на перетині блокчейн-аналізу та фінансового аналізу став звичайним інструментом для трейдерів і дослідників крипторинку.

Підсумовуючи огляд: наукові результати здебільшого підтверджують наявність статистичного зв’язку між показниками блокчейн-активності та ціною біткойна. Звичайно, сам по собі ончейн-аналіз не дає повної картини (на ціну також впливають зовнішні чинники: макроекономічні, регуляторні, поведінкові). Проте, включення мережевих метрик в аналіз дозволяє глибше зрозуміти внутрішню динаміку криптовалюти. У наступному підрозділі розглянемо конкретні історичні випадки, коли зміни в активності блокчейну явно співпадали з рухами курсу BTC.

1.5 Історичні приклади впливу активності блокчейну на курс Bitcoin

Історія біткойна надає чимало прикладів, коли показники та події у мережі безпосередньо пов'язували з змінами ціни. Розглянемо декілька знакових випадків, що ілюструють цей взаємозв'язок.

Бичачий ринок 2017 року та рекордна активність мережі. Наприкінці 2017 року біткойн пережив стрімкий ціновий злет до тодішнього історичного максимуму (~\$20 000 у грудні 2017). Цей період супроводжувався надзвичайно високим навантаженням на блокчейн Bitcoin. Кількість транзакцій у мережі неодноразово встановлювала рекорди: так, у грудні 2017 мережа обробляла понад 420 000 транзакцій на добу [9], що на той час було безпрецедентним показником. Мережа була перевантажена, розмір мемпулу зростав, а середня комісія за транзакцію сягала десятків доларів - усе це свідчило про масовий попит і ажіотаж серед користувачів. Аналітики відзначали, що зростання кількості транзакцій, унікальних адрес та інших метрик почалося ще за кілька місяців до цінового піку, тобто мережева активність випереджала і супроводжувала цінове ралі. Це узгоджується з теорією про мережевий ефект: максимальний притік нових користувачів та операцій співпав з піковою капіталізацією біткойна. Після спаду ціни у 2018 році, відповідно, добова кількість транзакцій теж зменшилася, хоча й залишилася вищою, ніж у допопередній період.

Продажі Mt. Gox та обвал 2018 року. Один із найвідоміших прикладів потенційного впливу великих ончейн-переказів на ринок це ситуація з розпродажем біткойнів, що належали збанкрутілій біржі Mt. Gox. У кінці 2017 - на початку 2018 року трастовий керівник Mt. Gox (Нобуакі Кобаяші) здійснив серію великих транзакцій, перевівши на біржі і продавши близько 35 000 BTC та 34 000 BCH (на суму понад \$400 млн) для виплат кредиторам біржі [6]. Ці продажі збіглися в часі з різким падінням курсу біткойна з

~\$20 000 (в грудні 2017) до ~\$6000 (у лютому 2018). Згідно зі звітами, кожна велика продаж Mt.Gox помітно “тиснула” на ринок, знижуючи ціну: так, коли 5 лютого 2018 BTC впав до ~\$5900 (локального дна), за день до цього Кобаяші перевів на біржу чергові 18 000 BTC. Фактично, масивні ончейн-транзакції з гаманців Mt.Gox були публічно відстежувані, і ринок реагував на них негативно, очікуючи збільшення пропозиції на біржах. Цей випадок продемонстрував, що концентровані продажі великих обсягів, видимі через блокчейн, можуть спричиняти значне падіння ціни. Хоча згодом аналіз показав, що ринок загалом «переварив» цей обсяг (частка Mt.Gox у загальному обсязі торгів була не надто велика), психологічний ефект від таких новин був потужним. Відтоді інвестори уважно стежать за рухом великих старих гаманців, подібні “пробудження” великих адрес часто інтерпретуються як можливий тригер ринкових коливань.

Активність «китів» та короткострокові коливання (приклад 2025 року). У травні 2025 року стався показовий випадок: гаманець, що був неактивний 11 років, раптово перевів 300 BTC (близько \$31 млн на той момент) на нову адресу. Ця новина отримала широкий розголос у крипто-спільноті. Ончейн-дані показали сплеск торгової активності одразу після переказу: обсяги торгів BTC зросли на 15% за кілька годин, а ціна BTC/USDT протягом двох годин після транзакції хитнулася приблизно на -1.2% (з \$103 500 до \$102 200). Аналітики зазначили, що подібні переміщення давніх монет історично призводили до просідання ціни ~2 - 3% протягом 1-2 діб, оскільки учасники ринку побоюються потенційного продажу такої великої суми. У цьому випадку також спостерігався ефект: хоча 300 BTC не є критичним обсягом для глобального ринку, сам факт «пробудження» довгострокового холдера посилив нервозність трейдерів. Цей приклад підтверджує, що моніторинг ончейн-транзакцій великих гравців став важливим інструментом для передбачення раптових короткострокових рухів ціни. Подібні ситуації траплялися і раніше, наприклад, у 2019-2020 роках кілька разів з’являлися

повідомлення про переміщення монет, видобутих ще у 2009-2010 роках (так звані «satoshi-era coins»), після чого на ринку фіксувалися підвищена волатильність і обговорення щодо особи власника цих монет та його мотивів.

Періоди ажіотажу та підвищених комісій. Ще один непрямий індикатор впливу ончейн-активності це зростання транзакційних комісій у пікові періоди. Наприклад, у квітні 2021, коли ціна BTC вперше перевищила \$60 000, середня комісія за транзакцію в мережі також стрімко зросла (до більш ніж \$50) через переповнення мемпулу [36]. Це свідчило про те, що попит на здійснення транзакцій (і, опосередковано, на сам біткойн) був надзвичайно високим, користувачі були готові платити великі комісії, аби лише провести операції швидше. Така ситуація спостерігалась і під час попереднього максимуму 2017 року. Отже, високі комісії та затримки в мережі відображають екстремальну активність і часто співпадають з переломними моментами ринку, після них зазвичай слідує корекція ціни, коли ажіотаж спадає.

Наведеними прикладами далеко не вичерпуються всі можливі випадки, але вони наочно демонструють, що динаміка блокчейн-активності та поведінка ринку біткойна тісно переплетені. Висока активність у мережі (багато транзакцій, нових користувачів, великі перекази) часто спостерігається у періоди різких злетів або падінь ціни. Це не завжди означає, що активність є першопричиною руху ціни (нерідко вона може бути наслідком цінових змін, наприклад, під час паніки люди масово переводять кошти). Проте в багатьох випадках ончейн-індикатори дають підказки щодо стану ринку: перегрітість, панічні настрої, акумуляція монет, вихід великих гравців тощо відображаються у самому блокчейні.

1.6 Висновки до розділу 1

У цьому розділі було здійснено теоретичний огляд теми впливу блокчейн-транзакцій на курс криптовалюти (на прикладі біткойна). Розглянуто базові поняття: принцип роботи блокчейну та специфіка криптовалют як економічних активів, чия вартість визначається попитом і пропозицією у децентралізованій системі. Проаналізовано теоретичні зв'язки між активністю в мережі (транзакціями, адресами, обсягами переказів) і ціною BTC: зокрема, мережевий ефект (закон Меткаффа) передбачає зростання цінності біткойна зі збільшенням числа користувачів, а такі метрики, як NVT, дозволяють порівняти ринкову оцінку з реальною використанням мережі. Був проведений огляд літератури, що показав: численні дослідження підтверджують кореляцію між ончейн-показниками та ціною; так, активні адреси, хеш-рейт та інші внутрішні фактори статистично впливають на вартість BTC, а великі транзакції та рухи монет нерідко передують ціновим рухам у короткостроковому періоді. Наведено історичні кейси, які ілюструють практичну значущість цих зв'язків: від рекордної завантаженості мережі на піку 2017 року до ефекту продажів Mt.Gox у 2018 і недавніх «пробуджень» давніх гаманців.

Загальні висновки свідчать, що активність у блокчейні є важливим барометром стану криптовалютного ринку. Звичайно, на курс біткойна впливають також зовнішні події та макроекономічні чинники, але аналіз ончейн-даних дозволяє відстежити дії самих учасників мережі та фундаментальні зміни в використанні криптовалюти. Отримані в цьому розділі теоретичні та оглядові результати створюють основу для подальшого дослідження: у наступних розділах буде зосереджено увагу на кількісному аналізі взаємозв'язку між ончейн-активністю і ціною BTC, перевірці гіпотез та побудові моделей, що впливають з викладеної теорії.

РОЗДІЛ 2 ТЕОРЕТИЧНІ ОСНОВИ АНАЛІЗУ ВПЛИВУ БЛОКЧЕЙН-ТРАНЗАКЦІЙ НА КУРС КРИПТОВАЛЮТИ

2.1 Кореляційний аналіз: коефіцієнти Пірсона та Спірмена

Кореляція – це статистична міра лінійної або монотонної залежності між двома змінними. Коефіцієнт кореляції Пірсона (product-moment correlation coefficient) вимірює лінійну силу зв'язку між двома кількісними показниками [12]. Його значення r варіюється від -1 до $+1$: значення $|r| = 1$ означає ідеальну лінійну залежність, близьке до 0 – слабкий або відсутній лінійний зв'язок. Формула для вибіркового коефіцієнта Пірсона виглядає наступним чином

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

де $\bar{x}$ і $\bar{y}$ – середні значення відповідних вибірок. Цей показник безрозмірнісний (одиниці виміру чисельника і знаменника скорочуються) і відображає лише силу і напрямок зв'язку (додатний або від'ємний). Для застосування коефіцієнта Пірсона припускається, що зв'язок між змінними приблизно лінійний, а дані мають нормально розподілені значення та сталу варіабельність (гомоскедастичність).

Коефіцієнт Спірмена (ранговий кореляційний коефіцієнт) оцінює монотонність зв'язку, тобто здатність однієї змінної збільшуватися або зменшуватися в порядку з іншою [13]. Він обчислюється як коефіцієнт Пірсона між рангами значень двох змінних. Якщо немає зв'язаних рангів, можна використати спрощену формулу Спірмена

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)},$$

де $d_i = R(x_i) - R(y_i)$ - різниця рангів i -го елемента за двома змінними. Значення ρ теж лежить в інтервалі $[-1; +1]$, при цьому $+1$ відповідає ідеально зростаючій монотонній залежності, -1 - спадній, а 0 - відсутності монотонного зв'язку. Коефіцієнт Спірмена є непараметричним: він вимагає лише порядкового (рангового) характеру даних і менш чутливий до викидів, ніж Пірсон.

Приклад застосування до часових рядів. Кореляційний аналіз може бути використаний для оцінки зв'язку між двома часовими рядами, наприклад, між обсягом транзакцій у блокчейні та ціною криптовалюти. Важливо пам'ятати, що пряме обчислення коефіцієнта Пірсона на нестационарних часових рядах може призводити до спуріозної кореляції – хибного високого зв'язку через тренди чи сезонність, не пов'язані причинно. Тому перед розрахунком кореляції часового ряду рекомендується привести дані до стаціонарного вигляду (наприклад, взяти прирости або відфільтрувати тренд). Для лагочутливих залежностей застосовують крос-кореляційний аналіз: коефіцієнт кореляції обчислюється між значеннями одного ряду і зсунутими на певний лаг значеннями іншого. Це дає змогу виявити, як одна змінна випереджає або запізнюється відносно іншої. Наприклад, можна перевірити, чи корелює кількість щоденних транзакцій BTC із ціною BTC через кілька днів лагу. Високі значення кореляції на лагу L означали б, що зміни в метриці блокчейну статистично передують змінам курсу з затримкою L .

2.2 Тести на стаціонарність: ADF та KPSS

Аналіз часових рядів (у тому числі фінансових, як ціни криптовалют) часто починається з перевірки на стаціонарність. Стаціонарний ряд – той, що не має систематичних змін у середньому і дисперсії в часі (немає тренду, гетероскедастичності чи нестаціонарних циклів). Багато моделей (ARIMA, VAR) вимагають стаціонарності для коректного застосування [14]. Для формальної перевірки використовують статистичні тести на наявність одиничного кореня, серед яких найпопулярніші – тест Діккі-Фуллера (ADF) та тест Квятковського–Філіпса–Шмідта–Шина (KPSS). Ці тести взаємно доповнюють один одного: ADF перевіряє нуль-гіпотезу про нестаціонарність, тоді як KPSS – про стаціонарність ряду.

Тест Авгментованого Діккі–Фуллера (Augmented Dickey-Fuller, ADF). Він перевіряє гіпотезу, що ряд має одиничний корінь, тобто є інтегрованим процесом першого порядку (нестационарним) [16]. Альтернативна гіпотеза – ряд стаціонарний (можливо, стаціонарний навколо тренду, якщо включено трендовий компонент). “Авгментований” означає, що до базового рівняння Діккі–Фуллера додані лагові різниці, щоб прибрати автокореляцію залишків.

ADF-тест базується на оцінюванні рівняння регресії

$$\Delta y_t = \alpha + \beta t + \phi y_{t-1} + \gamma_1 \Delta y_{t-1} + \gamma_2 \Delta y_{t-2} + \dots + \gamma_k \Delta y_{t-k} + \varepsilon_t,$$

де:

- 1) $\Delta y_t = y_t - y_{t-1}$ – перша різниця ряду,
- 2) k – кількість лагів, включених для усунення автокореляції,
- 3) α – константа, β_t – детермінований тренд (може включатися або ні).

Нуль-гіпотеза тестує параметр $\phi = 0$ (наявність одиничного кореня. В рівнянні залишається $\Delta y_t = \alpha + \beta t + \gamma_1 \Delta y_{t-1} + \dots$ тобто y_t має корінь 1) проти

альтернативи $\phi < 0$ (станціонарність: принаймні в першому порядку). Звернімо увагу: $\phi < 0$ відповідає процесу $y_t = \phi y_{t-1} + \dots$ з $|\phi| < 1$, який стаціонарний і повертається до середнього (при $\phi = 0$ маємо чистий випадковий блукання без повернення).

Після оцінювання регресії отримують t-статистику $\tau = \hat{\phi} / SE(\hat{\phi})$ для коефіцієнта $\hat{\phi}$. Розподіл цієї статистики не стандартний (не підкоряється звичайному t через присутність нестационарності в H_0), тому Діккі і Фуллер отримали спеціальні критичні значення. Якщо розрахована статистика τ менша (більш від'ємна), ніж критичне значення, нуль-гіпотеза відхиляється – ряд стаціонарний (немає одиничного кореня). Якщо τ вища за критичний поріг – немає підстав відкинути H_0 , дані узгоджуються з наявністю одиничного кореня (ряд нестационарний). Часто замість критичних значень використовують p-value: якщо $p < 0.05$, відкидаємо H_0 (стаціонарність підтверджується), якщо $p > 0.05$, не відкидаємо H_0 (ознаки нестационарності). Важливо врахувати включення детермінованих компонент: ADF може проводитися у трьох варіантах – без константи, з константою, з константою і трендом – кожен має свої критичні значення. Вибір залежить від поведінки ряду: для ряду з очевидним ненульовим середнім варто включити константу, з явним трендом – і трендовий член, бо інакше тест може помилково інтерпретувати тренд як доказ нестационарності.

Інтерпретація ADF: якщо тест показав $\phi < 0$ (відкинуто H_0), це свідчить, що ряд не має одиничного кореня і є стаціонарним (або стаціонарним навколо постійного тренду). Наприклад, для логарифму курсу BTC ADF може виявитися нестационарним (ціновий рівень блукає), але для його першої різниці $\Delta \ln(\text{Price})$ тест, ймовірно, відхилить H_0 – підтвердивши, що доходності стаціонарні. Тоді кажуть, що $\ln(\text{Price})$ – інтегрований порядок 1 ($I(1)$), а його приріст – $I(0)$ стаціонарний. Якщо ж ADF навіть на першій

різниці не відкидає H_0 , це натяк на більш високий порядок інтегрованості або структурні зрушення, які тест не врахував.

KPSS-тест (Kwiatkowski–Phillips–Schmidt–Shin). Цей тест має зворотну постановку: нуль-гіпотеза H_0 стверджує, що ряд стаціонарний (навколо середнього або тренду), а альтернативна – що він має одиничний корінь (нестабільний) [17]. Інтуїтивно, KPSS перевіряє, чи значно відрізняється ряд від свого стаціонарного тренду. Тестова статистика KPSS базується на лакмусовому методі (Lagrange Multiplier): ряд уявляється як сума детермінованого тренду, випадкового блукання і стаціонарної шумової компоненти, і перевіряється дисперсія випадкового блукання на нуль. Якщо дисперсія випадкового блукання нульова, ряд фактично = детермінований тренд + шум (стаціонарний навколо тренду); якщо дисперсія > 0 – присутній одиничний корінь (випадкове блукання додає нескінченну варіацію).

Практично, розраховується накопичена сума відхилень ряду від його трендової оцінки (наприклад, від середнього або від лінійної регресії по часу) і нормалізується на оцінку довгострокової дисперсії. Великі значення статистики KPSS означають, що ряд “віддаляється” від гіпотези стаціонарності – тобто, тренд нестабільний або дисперсія зростає з часом. Отримане значення порівнюють з табличними (розподіл статистики KPSS також не стандартний). Якщо KPSS значущо велика ($p < 0.05$), то відкидають нуль-гіпотезу і роблять висновок, що ряд не є стаціонарним (містить одиничний корінь). Навпаки, якщо p велике, немає підстав сумніватися в стаціонарності (принаймні тренд-стаціонарності).

Інтерпретація KPSS. Наприклад, KPSS-тест для ряду щоденних доходностей BTC, ймовірно, не відкине H_0 (доходності стаціонарні, що узгоджується з економічною теорією про відсутність тренду у доходностях), тоді як для самого рівня цін BTC тест майже напевно дасть мале p (відкине стаціонарність, підтвердивши наявність одиничного кореня). KPSS зручний

тим, що чутливий до різних типів нестационарності – як чистого випадкового блукання, так і випадку тренд+шум без повернення до тренду.

Оскільки ADF і KPSS мають протилежні нуль-гіпотези, їх часто застосовують у парі для підвищення достовірності висновку. Можливі ситуації:

1. ADF не відкинув H_0 (нестационарність), та одночасно KPSS відкинув H_0 (стационарність) – однозначний висновок про нестационарність (ряд інтегрований). Це типовий випадок для фінансових цін.
2. ADF відкинув H_0 (стационарний), KPSS не відкинув (стационарний) – ряд стационарний. Так зазвичай буває для приростів цін, спредів, різниць.
3. ADF і KPSS обидва “неоднозначні” (не відкинули свої H_0): можливо, ряд має складнішу структуру (наприклад, стационарний з переломами тренду або циклічністю), або тестам бракує потужності на короткій вибірці.
4. Обидва відкинули свої H_0 . Рідкісна ситуація, може означати, що ряд близький до граничного випадку (наприклад, містить сильний автокорельований процес з коренем, близьким до 1). Тоді треба проаналізувати величини статистик i , можливо, використовувати додаткові тести (Phillips–Perron, Zivot–Andrews з переломами тощо).

На практиці, перед побудовою моделей ARIMA/VAR для курсу криптовалюти, за допомогою ADF/KPSS пересвідчуються, що часом ряд $I(1)$ – якщо так, беруть перші різниці цін (доходності) як стационарну змінну для моделювання. Якщо ж деякі блокчейн-метрики, приміром, теж виявляються нестационарними (можуть мати зростаючий тренд з розвитком мережі), їх також диференціюють або застосовують коінтеграційний аналіз, аби уникнути спуріозних регресій. У підсумку, тести на стационарність є обов’язковим діагностичним інструментом при підготовці даних для

прогнозування криптовалют, оскільки фінансові часові ряди майже завжди містять одиничні корені.

2.3 Модель Vector Autoregression (VAR)

Vector Autoregression (VAR) – це багатовимірна авторегресійна-модель, що описує поведінку системи з кількох часових рядів (ендогенних змінних) [15], допускаючи вплив кожної змінної на себе та на інші змінні в попередніх періодах. Математично VAR(p) для K-вимірного часового ряду $Y_t = (y_{1,t}, \dots, y_{K,t})^T$ з лагом p записується як

$$Y_t = v + A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + u_t,$$

де:

- 1) v – вектор констант,
- 2) A_i – матриці коефіцієнтів розміру $K \times K$,
- 3) u_t – вектор випадкових похибок (інновацій).

Припускається, що u_t – білий шум із нульовим середнім та ковариаційною матрицею Σ_u , сталою в часі (гомоскедастичність). Кожне рівняння VAR описує динаміку однієї змінної через її власні лаги та лаги інших змінних. Наприклад, для двовимірного VAR(1) маємо систему рівнянь

$$\begin{aligned} y_{1,t} &= a_{1,0} + a_{1,1} y_{1,t-1} + a_{1,2} y_{2,t-1} + u_{1,t}, \\ y_{2,t} &= a_{2,0} + a_{2,1} y_{1,t-1} + a_{2,2} y_{2,t-1} + u_{2,t}, \end{aligned}$$

де коефіцієнт $a_{1,2}$ показує, як лагова зміна y_2 впливає на поточне значення y_1 , і навпаки. Таким чином, інтерпретація коефіцієнтів VAR - кожен ненульовий елемент матриці A_1 відображає величину впливу відповідної змінної (стовпець) з лагом 1 на певну цільову змінну (рядок) в наступному кроці. Наприклад, якщо в моделі VAR для ціни криптовалюти та кількості транзакцій коефіцієнт при лаговій транзакційній активності в рівнянні ціни є позитивним і значущим, то зростання числа транзакцій учора асоційоване з підвищенням ціни сьогодні (за інших рівних умов).

Припущення VAR. Усі включені часові ряди є стаціонарними (або приведеними до стаціонарності, наприклад шляхом диференціювання). Система вважається стаціонарною у слабкому сенсі, якщо її середнє та коваріації не змінюються з часом; в цьому випадку VAR має розв'язок у вигляді розкладу по минулим шокам (МА-представлення) і дисперсія залишається скінченною. Якщо ряд нестаціонарний, але має коінтеграцію з іншими, тоді доцільно використовувати розширення – Vector Error Correction Model (VECM). Інші ключові припущення: лінійність (зв'язки між змінними лінійні), відсутність автокореляції в резидуалах (її перевіряють тестами на «білу шумовість») та нормальність розподілу похибок (для коректності інтервальних прогнозів та тестів). Порушення цих припущень може вимагати модифікації моделі (наприклад, врахування ARIMA-компонент або використання трансформацій).

Вибір лагів VAR. Визначення оптимального порядку моделі (p) зазвичай здійснюється за допомогою критеріїв інформативності. Найпоширеніші – критерій Акаїке (AIC), Байєсовський інформаційний критерій Шварца (BIC) та критерій Ханнана-Квінна (HQIC). Вони обчислюються як

$$AIC = -2 \cdot LL + 2k,$$

$$BIC = -2 \cdot LL + k \cdot \ln(N),$$

$$HQIC = -2 \cdot LL + 2k \cdot \ln(\ln(N)),$$

де:

- 1) LL – логарифм максимальної правдоподібності моделі,
- 2) k – кількість параметрів,
- 3) N – обсяг вибірки.

Ці критерії балансують між якістю підгонки моделі та її складністю: штраф $2k$ у AIC є відносно меншим, тому AIC може обрати більш складну модель, тоді як BIC має сильнішу пеналізацію $k \ln N$ і схильний обирати простішу модель (особливо при великому N). HQIC займає проміжне положення. На практиці алгоритм такий: будується ряд моделей VAR з різними лагами p (наприклад, 1 до максимального допустимого), для кожної обчислюються інформаційні критерії, і обирається модель з мінімальним значенням критерію (позначають “*” найменше значення для кожного критерію). Окрім критеріїв, інколи застосовують тести на пропущені лаги (наприклад, тест Лагранжового множника на автокореляцію залишків) – послідовно перевіряючи, чи додавання ще одного лага істотно покращує модель. Після визначення оптимального p модель VAR оцінюється, зазвичай методами OLS окремо для кожного рівняння (оскільки всі регресори в правій частині лагові і однакові для всіх рівнянь, оцінки OLS збігаються з оцінками максимальної правдоподібності).

Аналіз результатів VAR. Після побудови моделі часто здійснюють аналіз імпульсних відгуків (Impulse Response Function, IRF) – досліджують, як одиничний шок у одній змінній впливає на значення всіх змінних системи в наступні періоди. IRF допомагає оцінити напрям і тривалість впливу: наприклад, шок транзакцій (раптове збільшення активності) може викликати поступову реакцію в ціні криптовалюти протягом декількох днів. Також використовується декомпозиція дисперсії прогнозу – вона показує, яку частку

варіації прогнозової помилки кожної змінної пояснюють власні шоки та шоки інших змінних. Це все дає змогу інтерпретувати побудовану VAR-модель економічно. Важливо розуміти, що кореляція не обов'язково означає причинність; для дослідження причинно-наслідкових зв'язків застосовують тест Грейнджера на причинність. Він перевіряє, чи покращує включення лагів змінної X прогноз змінної Y . Якщо так – кажуть, що “ X Грейнджер-причинна для Y ”. Для наших цілей, наприклад, можна перевірити, чи історичні дані про блокчейн-транзакції допомагають прогнозувати курс криптовалюти краще, ніж модель без цих даних.

2.4 Моделі машинного навчання для аналізу часових рядів і табличних даних

Для прогнозування часових рядів та аналізу табличних даних (data frames) все ширше застосовуються методи машинного навчання, зокрема ансамблеві методи та нейронні мережі. Розглянемо три популярні підходи: градієнтний бустинг на рішеннях деревах (реалізації LightGBM та CatBoost) та багат шаровий перцептрон (Multilayer Perceptron, MLP) – базовий тип штучної нейронної мережі.

2.4.1 LightGBM (Light Gradient Boosting Machine)

LightGBM – це високопродуктивна фреймворк-бібліотека градієнтного бустингу дерев рішень, розроблена Microsoft у 2017 році [10]. В основі

LightGBM – ансамбль з T дерев, кожне з яких намагається скоригувати помилки попереднього, мінімізуючи регуляризовану функцію втрат

$$L = \sum_{i=1}^n \ell(y_i, F_{t-1}(x_i) + h_t(x_i)) + \sum_{t=1}^T \Omega(h_t),$$

де:

- 1) $F_{t-1}(x) = \sum_{k=1}^{t-1} h_k(x)$ – прогноз ансамблю до кроку t ,
- 2) h_t – нове дерево на ітерації t ,
- 3) ℓ – функція втрат (для регресії звичайно $\ell(y, \hat{y}) = (y - \hat{y})^2$),
- 4) $\Omega(h) = \gamma |L| + \frac{1}{2} \lambda \sum_{j=1}^{|L|} w_j^2$ – регуляризація дерева з L листами і вагами w_j .

Ключові особливості LightGBM:

1. Histogram-based splitting.

Безперервні ознаки дискретизуються в B бінів. При обчисленні приросту інформації за розбиттям використовується градієнтний і гессіановий набір

$$G = \sum_{i \in I} g_i, H = \sum_{i \in I} h_i,$$

де

$$g_i = \partial_F \ell(y_i, F_{t-1}(x_i)), \quad h_i = \partial_F^2 \ell(y_i, F_{t-1}(x_i)).$$

При розбитті біну k на групи I_L та I_R інформаційний приріст (gain) обчислюється як

$$Gain = \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{G^2}{H + \lambda} - \gamma$$

За рахунок побудови гістограм і їх «віднімання» (histogram subtraction) складність пошуку оптимального спліту знижується з $O(n)$ до $O(B)$.

2. Leaf-wise tree growth (листякове нарощування).

На кожному кроці вибирається не рівень d , а саме лист, розщеплення якого дає найбільший приріст $Gain$. Це дозволяє досягти більшої редукції втрат за фіксованої кількості листів, але може призвести до перенавчання на малих вибірках, тому додатково вводяться обмеження: `max_depth` (максимальна глибина), `num_leaves` (кількість листків), `min_data_in_leaf` (мінімальна кількість даних в листку).

3. Exclusive Feature Bundling (EFB)

Якщо дві ознаки взаємно виключні (неактивні одночасно), їх можна об'єднати в одну псевдо-ознаку. Формулою це можна уявити як зведення двовимірної розрідженої матриці $X \in \mathbb{R}^{n \times p}$ до меншої розмірності:

$$X' = XP, \quad P \in \{0,1\}^{p \times p'}, \quad p' < p,$$

де матриця P кодує злиття без перекриття ненульових елементів.

4. Gradient-based One-Side Sampling (GOSS)

Щоб прискорити навчання, LightGBM відкидає частину прикладів з малими $|gradiєнтами|$, залишаючи всі «важкі»:

$$I_{keep} = \{i: |g_i| \geq \text{топ } a\%\}, \quad I_{sample} = \text{random subset of решти } b\%$$

і далі переважує приклади з малими градієнтами, щоб не спотворити розподіл:

$$\omega_i = \begin{cases} 1, & i \in I_{keep} \\ \frac{1-a}{b}, & i \in I_{sample} \end{cases}$$

Завдяки цим технологіям, LightGBM демонструє надзвичайно високу швидкість: за даними авторів, у 20 разів швидше за традиційні реалізації GBDT при аналогічній точності. Модель підтримує і інші стандартні функції: роботу з пропущеними даними, регуляризацію, паралельне та розподілене навчання тощо. LightGBM широко використовується у задачах регресії і класифікації, в змаганнях з прогнозування, зокрема й для фінансових часових рядів. При прогнозуванні курсу криптовалют LightGBM може врахувати вхідні ознаки як фінансового (ціни, обсяги торгів), так і блокчейн-характеру (кількість транзакцій, активні адреси, хешрейт тощо), автоматично відбираючи важливі ознаки через механізм важливості ознак (feature importance). Важливість може вимірюватися, наприклад, загальним інформаційним виграшем по кожній ознаці за всіма деревами або кількістю використань ознаки у розбиттях.

2.4.2 CatBoost (Categorical Boosting)

CatBoost – це реалізація градієнтного бустингу на деревах рішень із фокусом на «рідкі» (categorical) ознаки та боротьбу з «витоком цілі» (target leakage). Основні відмінності CatBoost від інших бібліотек (LightGBM, XGBoost) полягають у таких ключових компонентах:

1. Обробка категоріальних ознак. Замість стандартного one-hot-кодування CatBoost використовує метод Ordered Target Statistics. Для кожної категоріальної змінної й кожного прикладу він обчислює статистику (наприклад, середню цільову величину) з урахуванням лише тих спостережень, що “передують” поточному в довільній випадковій перестановці даних. Таким чином він уникає витоку інформації (коли поточний приклад ненавмисно «бачить» свою цільову змінну під час

- обчислення ознаки) і зменшує розмірність даних без експоненційного росту ознак.
2. Ordered Boosting. Щоб гарантувати коректне обчислення градієнтів у присутності категоріальних ознак, CatBoost тренує ансамбль дерев у порядку, визначеному випадковими перестановками даних. Кожна ітерація використовує лише ту частину вибірки, яка ще не була використана у попередніх деревах, для обчислення градієнтів. Це дозволяє формувати незміщені оцінки градієнтів і знижує ризик перенавчання на шумних або малих вибірках.
 3. Облівійні (symmetrical) дерева. В CatBoost використовується спеціальний тип дерев рішень, які мають симетричну (“облівій”) структуру: на кожному рівні дерева розбиття відбувається за однією і тією ж ознакою, лише з різними порогами. Така архітектура пришвидшує предиктивний процес і зменшує кількість гіперпараметрів, одночасно забезпечуючи дуже високий ступінь регуляризації, бо в кожній гілці використовується рівнозначний набір ознак.
 4. Регуляризація та скорочення варіантів. CatBoost підтримує L2-регуляризацію ваг у листях (l2_leaf_reg), ранню зупинку на валідаційному наборі (early_stopping_rounds), а також стохастичні процедури feature-fraction (випадковий відбір ознак для кожного дерева) та bagging-fraction (підвибірка рядків). Це допомагає зменшити надмірну кореляцію між деревами та знизити ризик перенавчання, особливо на малих даних.
 5. Підтримка пропусків та масштабування. Пропущені значення (NaN) CatBoost обробляє “з коробки”, навчаючи оптимальний напрямок, ліво чи право - при розбитті по ознаках із пропусками. Крім того, можна вмикати різні стратегії масштабування та нормалізації ознак, але

завдяки внутрішнім оптимізаціям модель часто демонструє стійкість без додаткового препроцесингу.

6. Продуктивність масштабованість CatBoost підтримує як мультипроцесорне, так і GPU-прискорене навчання. Для великих табличних даних зі складними категоріальними ознаками це дає значний вииграш у швидкості порівняно з класичними реалізаціями градієнтного бустингу.
7. Інтерпретація. Як і інші бустингові фреймворки, CatBoost надає метрику feature importance, а завдяки підтримці TreeSHAP можна будувати детальні SHAP-графіки, які показують внесок кожної ознаки у прогноз як локально (для окремого прикладу), так і глобально (для всієї вибірки).

Математичні основи CatBoost.

Нижче наведено короткий виклад ключових математичних концепцій, які лежать в основі алгоритму CatBoost.

1. Функція втрат і регуляризація

CatBoost навчає ансамбль дерев T шляхом мінімізації регуляризованої функції втрат

$$L = (y, \hat{y}) = \sum_{i=1}^n \ell(y_i, F_T(x_i)) + \sum_{t=1}^T \Omega(h_t),$$

де:

- 1) y_i - істинне значення цілі (тут це Δ ціна),
- 2) $F_T(x) = \sum_{t=1}^T h_t(x)$ - прогноз моделі як сума дерев,
- 3) ℓ це вибрана функція втрат (для регресії за замовчуванням MSE:
 $\ell(y, \hat{y}) = (y - \hat{y})^2$),
- 4) $\Omega(h) = \gamma |L| + \frac{1}{2} \lambda \sum_{j=1}^{|L|} \omega_j^2$ - регуляризація кожного дерева h з L листами та вагами w_j .

2. Порядкове (Ordered) обчислення градієнтів.

Щоб уникнути витоку цільової змінної при роботі з категоріями, CatBoost розбиває дані на послідовні “перестановки” і для кожного прикладу і обчислює градієнт, використовуючи лише ті передуючі спостереження.

$$g_i = \frac{\partial \ell(y_i, F_{i-1}(x_i))}{\partial F_{i-1}(x_i)}, \quad h_i = \frac{\partial^2 \ell(y_i, F_{i-1}(x_i))}{\partial F_{i-1}(x_i)^2}$$

і далі будує дерево, мінімізуючи локальне приближення

$$\sum_i [g_i h_i(x_i) + \frac{1}{2} h_i h_i(x_i)^2] + \Omega(h_i)$$

3. Оптимальні ваги листа.

Нехай у новому дереві h_t лист j містить індекси спостережень I_j . Тоді оптимальна вага w_j цього листа знаходиться з умови мінімізації другорядного розкладу функції втрат

$$\omega_j = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}$$

4. Обробка категоріальних ознак.

Для категоріальної ознаки c зі значенням v у прикладі i , CatBoost рахує ordered target statistic

$$CTS_i(c = v) = \frac{\sum_{k < i, x_k^c = v} y_k + a \cdot prior}{\#\{k < i \mid x_k^c = v\} + a},$$

де:

- 1) $k < i$ означає «лише ті рядки, що передують i у довільній перестановці»,
- 2) $prior$ – апріорний коефіцієнт (наприклад, середня цільова змінна по всій вибірці),
- 3) a – коефіцієнт згладжування, щоб уникнути нестабільності при малому числі попередніх спостережень.

5. Симетричні (“облівійні”) дерева

CatBoost використовує симетричні (symmetric) дерева, де на кожному рівні ddd застосовується одна й та ж ознака, але з різними порогами. Це гарантує однакову “глибину” усіх гілок і прискорює обчислення (2.18)

$$h_l(x) = \sum_{d=1}^D LeafWeight(bin(x_{j_d}))$$

де j_d – індекс ознаки, обраної на рівні d , а $bin(x)$ – номер інтервалу (біну) відстані від порогу.

2.4.3 Багатошаровий перцептрон (MLP)

Багатошаровий перцептрон (MLP) – це класична архітектура штучної нейронної мережі прямого поширення (feedforward neural network). По суті, це функціональний наближувач, що складається з входів, одного чи більше прихованих шарів з нелінійними нейронами та вихідного шару. Кожен нейрон виконує лінійну комбінацію виходів з попереднього шару і пропускає її через нелінійну функцію активації. Зв’язки між нейронами мають вагові коефіцієнти, які змінюються в процесі навчання.

MLP є узагальненням простого перцептрону Розенблатта (який міг моделювати лише лінійно роздільні залежності) на випадок багатьох шарів – що дозволяє моделювати нелінійні закономірності у даних. Теорема Цибенко (1989) стверджує, що одношаровий MLP з достатньою кількістю нейронів може апроксимувати будь-яку неперервну функцію на компактній множині (універсальна апроксимація). Кожен додатковий шар дозволяє відловлювати більш складні патерни.

Навчання MLP зазвичай відбувається шляхом зворотного поширення помилки (backpropagation): обчислюється похибка моделі на тренувальному наборі (наприклад, середньоквадратична для регресії або крос-ентропія для

класифікації), потім за правилом градієнтного спуску ваги шарів оновлюються у напрямку, який зменшує цю помилку. Завдяки диференційовності сучасних функцій активації (сигмоїда, ReLU, tanh тощо), градієнт можна ефективно обчислювати, проходячи мережею вниз від виходу до входу і застосовуючи правило ланцюга. На практиці використовуються оптимізатори типу Adam, RMSprop, які адаптують крок градієнтного спуску для швидкої і стабільної збіжності.

Типова архітектура MLP: вхідний шар приймає ознаки $(x_1, \dots, x_m)$, можливо після нормалізації; далі один або кілька прихованих шарів по h нейронів кожен (значення h – гіперпараметр), кожен нейрон обчислює $z_j = \omega_{j0} + \sum_i \omega_{ji} x_i$ і вихід $a_j = f(z_j)$, де f – активація (наприклад, $\text{ReLU}(x) = \max(0, x)$ чи логістична сигмоїда $\sigma(x) = 1 / (1 + e^{-x})$). Далі виходи цього шару слугують вхідними сигналами для наступного шару. У вихідному шарі використовується активація, що відповідає задачі: для регресії це часто лінійна активація (або відсутня нелінійність), для класифікації – softmax або сигмоїда (щоб отримати ймовірності класів).

Принцип роботи нейронної мережі. Завдяки нелінійним активаціям MLP може розпізнавати складні взаємозв'язки між вхідними змінними. Геометрично, приховані шари виконують послідовні перетворення простору ознак, роблячи його більш лінійно роздільним для останнього шару. MLP не вимагає припущень про розподіл даних або лінійність залежностей, тому здатен вивчати патерни, які інші методи (напр., лінійні моделі) не вловлюють. З іншого боку, нейромережі потребують великого обсягу даних для навчання і схильні до перенавчання, тому часто використовують методи регуляризації: L2-норму ваг (weight decay), dropout (випадкове “вимикання” нейронів під час навчання), ранню зупинку тощо.

MLP у аналізі часових рядів: хоча для послідовних даних спеціально розроблені рекурентні мережі (RNN, LSTM), багатошаровий перцептрон теж може застосовуватися до прогнозування рядів шляхом вбудовування лагів у

ознаки. Тобто, замість урахування часової залежності через рекурентність, будується статична модель: як вхід подаються значення ряду за попередні ppr кроків (i , за потреби, інші супутні змінні), а вихід мережі продукує прогноз на наступний крок. Наприклад, для передбачення ціни BTC на добу вперед можна побудувати MLP, вхідними ознаками якого будуть ціни за останні 10 днів, обсяги торгів, кількість транзакцій, складність майнінгу тощо. Така мережа вчитиметься співвідносити патерни в цих ознаках з майбутнім значенням ціни. MLP здатний змодельовати нелінійні ефекти, наприклад, комбінації високого обсягу транзакцій та зростаючого хешрейту можуть сигналізувати про підвищення ціни, навіть якщо поокремо ці фактори не мали б лінійного впливу.

Застосування MLP вимагає уважного налаштування: кількість шарів і нейронів впливає на здатність до навчання і узагальнення. Занадто маленька мережа може не вловити складні закономірності (недонавчання), занадто велика – запам'ятати тренувальні дані (перенавчання). Тому архітектуру обирають, зазвичай, експериментально або використовуючи валідацію. В сучасних умовах MLP все частіше поступається більш спеціалізованим архітектурам (LSTM, трансформерам) для часових рядів, однак залишається важливою базовою моделлю і часто служить бенчмарком. Крім того, MLP має перевагу швидкого навчання на табличних даних і легко інтерпретується через ваги та впливи ознак (хоча й не так просто, як рішення дерева).

2.5 Метрики оцінки якості прогнозу

Для кількісної оцінки точності моделей прогнозування часових рядів використовуються різні метрики помилки. Серед найбільш уживаних – середня квадратична помилка (MSE), її квадратний корінь RMSE, середня

абсолютна помилка (MAE) та коефіцієнт детермінації (R^2). Кожна з цих метрик відображає дещо різні аспекти якості моделі.

1. MAE (Mean Absolute Error) – середня абсолютна різниця між прогнозованими значеннями $\hat{y}$ та фактичними y_i

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

Вона показує, на яку величину в середньому помиляється модель, виражену в тих самих одиницях, що й цільова змінна. MAE є інтерпретаційно зрозумілою (наприклад, “в середньому прогноз ціни помиляється на \$500”) і стійкою до викидів, оскільки всі помилки лінійно впливають на метрику.

2. MSE (Mean Squared Error) – середній квадрат помилки

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

MSE має корисні математичні властивості (диференційовність), тому часто використовується як функція втрат при навчанні моделей (зокрема, при методу найменших квадратів і в нейронних мережах). MSE пеналізує великі помилки більш жорстко, ніж MAE (через квадрат), тобто один великий промах збільшує MSE значно сильніше, ніж декілька середніх помилок. Це корисно, коли великі відхилення особливо небажані. Проте через квадрати ця метрика вимірюється в квадратичних одиницях виходу, що ускладнює інтерпретацію.

3. RMSE (Root Mean Squared Error) – квадратний корінь з MSE

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2}$$

Це стандартне відхилення резидуалів (помилки моделі). RMSE повертає нас до тих самих одиниць, що й цільова змінна, що робить інтерпретацію зручнішою (як і MAE). Завдяки властивості успадкованій від MSE, RMSE також сильніше карає великі відхилення, тому підкреслює наявність значних

прорахунків моделі. Загалом RMSE є, мабуть, найпопулярнішою метрикою в прогнозуванні часових рядів і регресії: її часто обирають як ціль оптимізації та для порівняння моделей.

4. Коефіцієнт детермінації (R^2) – частка варіації цільової змінної, яка пояснюється моделлю. Він визначається як

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2},$$

де в чисельнику – сума квадратів помилок моделі (SSR), а в знаменнику – сума квадратів відхилень фактичних значень від їх середнього (SST). R^2 завжди лежить між 0 і 1 для моделей з константою: 1 означає ідеальне пояснення ($SSR = 0$), 0 означає, що модель не краща за тривіальний прогноз “середнім” ($SSR = SST$). Негативне R^2 теж може траплятися, якщо модель настільки погана, що помилка більша, ніж варіативність самого ряду. Інтерпретація: наприклад, $R^2 = 0.85$ означає, що 85% дисперсії (коливань) цільової змінної вдалося пояснити моделлю, а решта 15% залишається в резидуалах. R^2 є безрозмірним і зручним для оцінки наскільки добре модель підгоняє дані в порівнянні з константною моделлю. Проте він менш інформативний в контексті прогнозування: високе R^2 на навчанні не гарантує хорошої екстраполяції. Для часових рядів R^2 використовується рідше, ніж RMSE/MAE, через можливі нестационарності та автокореляцію (в таких умовах R^2 може бути оманливо високим). Також, R^2 не дозволяє прямого порівняння моделей на різних наборах даних.

Порівняння метрик. MAE та RMSE обидві виражені в одиницях цільової змінної і їх можна порівнювати прямо. MAE легше трактувати, RMSE більш чутлива до великих похибок. Якщо розподіл помилок близький до нормального, то ~68% помилок лежать в межах $\pm RMSE$. R^2 зручно використовувати для пояснення моделі (наскільки добре враховано варіацію), але для прогнозних цілей краще орієнтуватися на абсолютні метрики. Зазвичай, при налаштуванні моделі намагаються мінімізувати RMSE або

MAE (в залежності від того, чи критичні викиди). Важливо зазначити, що для часових рядів також популярні відносні метрики, наприклад MAPE (середня абсолютна процентна помилка), але вона не визначена при нульових фактичних значеннях і може бути зміщена при дуже малих величинах ряду. Тому RMSE/MAE часто надають перевагу для фінансових даних.

У випадку криптовалют, де масштаб цін може змінюватися на порядки, іноді беруть RMSE від логарифму ціни або відсоткової зміни, щоб метрика відображала відносну похибку. Наприклад, $RMSE=500$ USD може бути прийнятною помилкою для біткоїна при ціні 50,000 (1%), але катастрофічною для альткоїна вартістю \$100 (500%). Вибір метрики залежить від мети: для моделювання волатильності часто мінімізують саме квадрати помилок (MSE), бо великі відхилення (сплески волатильності) критичні. Для середньострокових прогнозів може бути досить мінімізувати MAE, щоб модель була оптимально точною “в середньому”. R^2 варто наводити у звітах як інтуїтивний показник якості підгонки, але основний акцент робиться на величинах помилок (MAE, RMSE), оскільки саме вони безпосередньо пов’язані з потенційними втратами чи недоотриманим прибутком від неточного прогнозу.

SHAP-аналіз для інтерпретації ML-моделей.

Сучасні моделі машинного навчання (дерева в ансамблях, нейронні мережі тощо) часто виступають як “чорні скриньки” – вони дають точні прогнози, але мало піддаються інтерпретації. SHAP (SHapley Additive exPlanations) – це підхід, що забезпечує пост-модельну інтерпретацію на основі концепції значень Шеплі з теорії кооперативних ігор [11]. Ідея полягає в наступному: розглядати кожну ознаку як гравця у грі, результатом якої є прогноз моделі для конкретного прикладу, а значення Шеплі визначають, яку “винагороду” (внесок у результат) слід справедливо приписати кожному гравцю (ознаці).

У термінах моделі. Припустимо, модель передбачила ціні BTC на завтра значення \$50,000, тоді як середнє по даних (базовий прогноз без ознак) – \$48,000. Різниця +\$2,000 – це “виграш” від урахування ознак. SHAP-значення дозволяють розкласти цей виграш на суму внесків кожної ознаки: наприклад, ознака X_1 (кількість транзакцій) могла додати +\$1,200, ознака X_2 (обсяг торгів) +\$800, а, скажімо, ознака X_3 (індекс настроїв) – відняти -\$200, разом отримуємо +\$1,800 (решта +\$200 може припадати на взаємодії або на випадкову похибку моделі). В ідеалі, сума SHAP-показників по всіх ознаках дорівнює різниці між прогнозом для даного прикладу та середнім прогнозом по датасету. Це властивість адитивності Shapley: значення моделі представляється як сума базового рівня та індивідуальних внесків ознак. Базовим рівнем зазвичай беруть середнє значення цільової змінної на тренуванні (для регресії) або лог-іт шанси для нульової моделі (для класифікації).

Як обчислюється значення Шеплі? Формально, для ознаки j це усереднений граничний внесок ознаки j у прогноз по всіх можливих коаліціях ознак. Уявімо, що ми можемо включати чи виключати ознаки з моделі (в практиці “виключення” реалізують підстановкою певного базового значення ознаки, напр. середнього). Тоді для кожної підмножини ознак S без j можна побудувати модель на ознаках з S і спрогнозувати результат $\hat{y}(S)$, а також на тій же підмножині $\cup \{j\}$ з ознакою j : $\hat{y}(S \cup \{j\})$. Маржинальний вплив j буде $\hat{y}(S \cup \{j\}) - \hat{y}(S)$. Значення Шеплі ϕ_j – це середнє такого впливу по всіх 2^{M-1} підмножинах (коаліціях) без j . Формула Шеплі виглядає так

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(M-|S|-1)!}{M!} (\hat{y}(S \cup \{j\}) - \hat{y}(S)),$$

де M – загальна кількість ознак, сумування йде по всіх підмножинах S , що не містять j , а коефіцієнт – кількість таких порядків входу ознак, при яких

S – це множина ознак, що увійшли до j . Ця формула гарантує виконання властивостей:

- 1) ефективність (сума всіх ϕ_j дорівнює $(\hat{y}(\text{повна модель}) - \hat{y}(\text{порожня модель}))$),
- 2) симетричність (якщо дві ознаки впливають на вихід однаково, їх ϕ однакові),
- 3) нульовий внесок (ознака, що не змінює вихід незалежно від комбінації інших, отримує $\phi = 0$) та аддитивність (для ансамблів значення Шеплі сумуються).

Обчислити точно всі коаліції для навіть помірного M – нереально (експоненційна складність). SHAP-аналіз реалізує оптимізовані алгоритми для цього: KernelSHAP – апроксимація шляхом вибірки коаліцій (вважається лінійно ваговою моделлю залежності виходу від індикаторів наявності ознак); TreeSHAP – точний алгоритм для деревоподібних моделей, що використовує структуру дерев для швидкого підрахунку внесків ознак без перебору всіх комбінацій. Для градієнтних бустингів (як LightGBM, CatBoost) застосування TreeSHAP дозволяє отримати значення Шеплі для кожного передбачення за час порядку $O(T * D^2)$, де T – загальна кількість листів у всіх деревах, D – середня глибина дерева (ефективно, дуже швидко для невеликих дерев).

Практичний зміст SHAP: Для кожного конкретного прогнозу модель SHAP дає набір значень $\{\phi_j\}$ по ознаках, який пояснює, чому прогноз відхиляється від базового очікування. Додатне ϕ_j означає, що ознака збільшує прогноз (порівняно з базою), від’ємне – зменшує. Наприклад, якщо модель прогнозує високу ціну і ми бачимо, що великі значення числа транзакцій та росту адрес мають великі +SHAP, це свідчить: модель інтерпретує високу ончейн-активність як фактор зростання ціни. Натомість негативний SHAP для, скажімо, змінного ринковий ризик означає: у даному випадку високий ризик тисне прогноз вниз.

Перевага SHAP – це локальна інтерпретація: пояснення для кожного індивідуального прогнозу. Крім того, агрегуючи $|\text{SHAP}|$ значення по багатьох вибірках, можна отримати глобальну важливість ознак (середній абсолютний внесок ознаки в прогноз). На відміну від чисто дерева-орієнтованої важливості (кількість сплітів), SHAP-важливості більш консистентні: доведено, що для будь-якої моделі з адитивною важливістю (а Shapley – адитивна) існує єдине задовільняюче всі аксіоми розподілення важливостей рішення – власне Shapley values.

Візуалізація SHAP. Результати SHAP-аналізу зручно інтерпретувати графічно. Поширений формат – summary plot: для кожної ознаки малюється точка для кожного прикладу, по осі X – значення SHAP, по осі Y – ознаки, колір точки відображає власне значення ознаки (високе чи низьке). Таким чином, можна побачити не лише важливість ознаки (розкид точок по горизонталі), а й напрям впливу (високі значення ознаки більше схильні збільшувати чи знижувати прогноз, судячи з кольору точок). Ще один корисний графік – force plot (водоспад): для окремого прогнозу показує послідовно, як від базового значення кумулятивно додаються/віднімаються внески кожної ознаки, доводячи до підсумкового прогнозу. Це особливо зручно, щоб пояснювати рішення моделі бізнес-користувачам: можна сказати “модель прогнозує ціну = \$50k, бо високий обсяг транзакцій додав +\$1.2k, зростання хешрейту +\$0.5k, а низький індекс настрою зменшив прогноз на -\$0.3k”, тощо.

SHAP-аналіз важливий тим, що він модельно-агностичний: підхід застосовується до будь-якої моделі (дерева, нейронні мережі, SVM і т.д.), що робить його стандартом інтерпретації в ML. Багато відкритих бібліотек (наприклад, shap на Python) реалізують ці ідеї, спрощуючи їх використання. Для задач прогнозування фінансових показників, зокрема криптовалют, SHAP допомагає зрозуміти роль блокчейн-метрик у моделі: скажімо, якщо модель “вважає” аномально високу кількість нових адрес ознакою майбутнього росту

ціни і відповідно збільшує прогноз, це може дати аналітичне підтвердження гіпотези, що ончейн-попит рухає ринок. З іншого боку, якщо важливі ончейн-фактори мають малі SHAP-значення у порівнянні з технічними індикаторами, це сигналізує, що на даному інтервалі модель не знайшла в них корисної інформації.

2.6 Висновки до розділу 2

Глибоке теоретичне розуміння зазначених методів закладає основу для побудови якісної моделі прогнозування курсу криптовалют. Кореляційний аналіз (Пірсон/Спірмен) дозволяє виявляти лінійні та нелінійні зв'язки між ринковими і ончейн-метриками, VAR-моделювання – врахувати взаємний вплив кількох часових рядів (ціна, транзакції, хешрейт тощо) один на одного в динаміці. Сучасні машинні моделі (LightGBM, CatBoost, MLP) забезпечують гнучкість у врахуванні великої кількості ознак і нелінійностей, метрики помилки RMSE, MAE та R^2 дають кількісне розуміння точності прогнозів, а SHAP-аналіз робить крок до прозорості “чорних скриньок”, показуючи, які ознаки і як саме впливають на прогноз. В решті, перевірка стаціонарності (ADF, KPSS) гарантує, що ми правильно обробили ряди (наприклад, застосували різницювання) перед моделюванням, що запобігає хибним висновкам.

Важливість включення блокчейн-метрик у моделі прогнозування курсу криптовалют. Блокчейн-метрики (ончейн-дані) – такі як кількість транзакцій, активні адреси, обсяг перерахованих коштів, хешрейт, складність майнінгу та інші – відображають фундаментальну активність та стан мережі криптовалюти. На відміну від суто цінових і об'ємних даних ринку (які часто враховуються технічним аналізом), ончейн-показники дають уявлення про

внутрішнє користування і попит на криптовалюту. Дослідження підтверджують, що включення ончейн-факторів може суттєво поліпшити точність прогнозів цінових моделей [18]. Зокрема, за допомогою методів машинного навчання було показано, що інтеграція метрик блокчейну (наприклад, доходи майнерів, щоденні комісії, кількість активних адрес) разом із традиційними ринковими змінними підвищує прогностичні здібності моделей для Bitcoin та інших криптовалют. Це інтуїтивно зрозуміло: скажімо, різке зростання кількості транзакцій [27] або нових учасників мережі може передувати підвищенню ціни, сигналізуючи про більший попит і прийняття. Блокчейн-метрики по суті виконують роль фундаментальних ознак (аналогічно економічним індикаторам для фондового ринку), які можуть пояснити рух ціни, коли технічні індикатори мовчать. Включення їх у моделі (будь то VAR, CatBoost чи MLP) дає змогу моделі вчитися на ширшому інформаційному просторі. Практичні результати вже показали покращення: так, модель Prophet з доданими ончейн-фічами продемонструвала значно меншу помилку прогнозу для ряду криптовалют, ніж без них. Інше дослідження відзначило, що точність класифікації напрямку руху ціни BTC покращується при врахуванні ончейн-даних поряд з технічними та настроєвими змінними. У підсумку, блокчейн-метрики збагачують моделі прогнозування критично важливою інформацією про “підкапотні” процеси криптовалютної мережі, тим самим підвищуючи надійність і обґрунтованість отриманих прогнозів. З огляду на високу волатильність крипторинків [29], використання максимально доступної інформації – запорука побудови стійких прогнозних моделей і ухвалення обґрунтованих інвестиційних рішень.

РОЗДІЛ 3 ПРОГРАМНА РЕАЛІЗАЦІЯ ТА ЕКСПЕРИМЕНТАЛЬНІ РЕЗУЛЬТАТИ

3.1 Вступ

У попередньому розділі було проведено теоретичний аналіз впливу блокчейн-транзакцій на вартість криптовалют, зокрема біткоіна. З'ясовано, що ончейн-показники (такі як кількість транзакцій, активність адрес, обсяги переказів тощо) потенційно пов'язані з динамікою курсу криптовалют. Також було розглянуто підходи до прогнозування фінансових часових рядів і сучасні методи машинного навчання, які можуть бути застосовані для врахування специфіки блокчейн-даних. Спираючись на ці теоретичні засади, у цьому розділі здійснено практичну реалізацію моделі аналізу та прогнозування курсу біткоіна на основі ончейн-метрик.

Метою експериментального дослідження є кількісно оцінити вплив різних характеристик блокчейн-активності на зміну ціни BTC і перевірити, чи можуть ці показники покращити точність прогнозування курсу. Для цього використано модель градієнтного бустингу LightGBM, обрану на основі попереднього порівняння із альтернативними підходами. У ході роботи розроблено набір ознак на основі 272 днів історичних даних, проведено тренування та налаштування моделей, оцінено їх точність та інтерпретовано результати за допомогою SHAP-аналізу важливості ознак. Окрім того, здійснено виявлення аномальних подій (нестандартних стрибків ціни) та проаналізовано їх зв'язок з ончейн-активністю. Структура даного розділу побудована таким чином: спочатку описуються вхідні дані (підрозділ 3.2) та сформовані ознаки (підрозділ 3.3), далі наведено результати експериментів (підрозділ 3.4) – включно з початковим кореляційним аналізом(підрозділ 3.4.1), VAR-моделю (підрозділ 3.4.2), налаштуваннями моделей (підрозділ

3.4.3), презентацією результатів моделювання та їх інтерпретацією (підрозділ 3.4.4), а також порівнянням моделей (підрозділ 3.4.5). У висновках до розділу (підрозділ 3.6) підсумовано основні результати та окреслено можливості практичного застосування отриманих результатів у фінансовій сфері.

3.2 Опис вхідних даних

Для аналізу було зібрано історичні дані щодо Bitcoin та пов'язаних показників блокчейну за 272 дні поспіль (щоденні спостереження з червня 2023 року по березень 2024 року). Вибір такого проміжку зумовлений бажанням охопити різні стани ринку (як відносно стабільні періоди, так і фази підвищеної волатильності) та достатню кількість точок для тренування моделей. Вхідний датасет містить такі ключові величини (атрибути) для кожного дня:

1. Курс BTC (ціна закриття дня) – вартість біткоїна в доларах США наприкінці дня на вибраній біржі (для однорідності даних використовувався один джерело ринкових даних). Цей показник є цільовим для прогнозування та водночас слугує важливою ознакою (наприклад, ціна попереднього дня як індикатор тренду).
2. Кількість транзакцій – загальна кількість підтверджених транзакцій у мережі Bitcoin за день. Даний показник відображає загальну активність у мережі; зростання числа транзакцій може вказувати на підвищений інтерес користувачів до мережі або збільшення торгової активності.
3. Обсяг транзакцій – сумарний обсяг біткоїнів, переданих у мережі за день (сума BTC, що змінила власника). Цей показник вимірює економічну активність на блокчейні: великі добові обсяги можуть

сигналізувати про підвищену участь крупних гравців або проведення нетипово великих переказів.

4. Кількість унікальних адрес – число окремих блокчейн-адрес, які були активними протягом дня (тобто виступали відправниками чи отримувачами в транзакціях). Даний показник є аналогом метрики «активних адрес» і використовується як індикатор рівня залученості користувачів мережі. Збільшення кількості активних адрес, як правило, корелює з ростом інтересу до криптовалюти.
5. Кількість великих транзакцій – число транзакцій за день, обсяг яких перевищує визначений поріг, що відносить їх до «великих» (наприклад, транзакції на суму понад 1000 BTC або еквівалентну значну суму у доларах). Ця метрика покликана виокремити операції «китів» – крупних власників, здатних здійснювати вплив на ринок.
6. Funding rate (ставка фінансування) – середнє значення ставки фінансування за день для безстрокових ф'ючерсів на BTC (за даними провідних бірж). Фандінгова ставка є показником переважного настрою трейдерів на деривативному ринку: позитивне значення означає, що довгі позиції сплачують коротким (ринок очікує зростання ціни) [30], негативне – навпаки. Різкі зміни ставки фінансування можуть передувати розворотам ціни, відображаючи перегрів ринку.

Дані з перерахованих джерел були об'єднані за датами в єдиний набір. Перед використанням у моделюванні виконано передобробку: перевірено відсутність пропущених значень, проведено синхронізацію часових міток (усі показники агреговано за добу відповідно до UTC). За потреби здійснювалося масштабування ознак – зокрема, для моделей глибокого навчання ознаки нормувалися до співмірних діапазонів (наприклад, шляхом стандартизації до нульового середнього і одиничного стандартного відхилення), щоб уникнути домінування великих числових масштабів (ціна BTC, обсяги) над меншими (фандінгова ставка тощо).

Варто відзначити характер розподілів вхідних даних. За досліджуваний період середній щоденний курс BTC становив десятки тисяч доларів, демонструючи як плавні тенденції росту, так і різкі сплески волатильності [35]. Наприклад, у середині досліджуваного періоду спостерігався стрибок ціни приблизно з ~\$25 тис. до ~\$30 тис. (+20% за кілька днів) на тлі позитивних новин ринку, тоді як ближче до кінця періоду курс досяг значень понад \$50–60 тис.\$. Ончейн-метрики також суттєво варіювали: кількість транзакцій коливалася від приблизно 200–300 тисяч на добу у спокійні періоди до понад 500 тисяч у пікові моменти (що частково було пов'язано з появою підвищеного навантаження від сервісів типу Ordinals у 2023 році). Обсяг переказів BTC міг змінюватися на порядок залежно від активності великих гаманців: у окремі дні сумарно передавалося лише близько 200–300 тис. BTC, тоді як у дні з аномальними транзакціями цей показник перевищував 1 млн BTC. Кількість активних адрес у середньому перебувала в діапазоні 600–900 тисяч, збільшуючись під час значних цінових рухів (що вказує на приплив користувачів у періоди ажіотажу). Середній рівень ставки фінансування протягом більшості часу був близький до нуля (ринок збалансований), але напередодні різких змін ціни спостерігалися відхилення: наприклад, перед значним зростанням курсу ставки фінансування неодноразово ставали позитивними ($>0.05\%$ на добу), сигналізуючи про переважання оптимізму і левередж-лонгів, тоді як перед падіннями були короткі періоди від'ємного фандінгу [31].

Отже, зібраний набір даних охоплює різноманітні аспекти мережевої активності та ринкових очікувань, створюючи основу для побудови комплексної моделі прогнозування. Нижче наведено склад ознак, розрахованих на основі цих вхідних даних, та обґрунтування їх включення до моделі.

3.3 Опис ознак

На основі описаних вище сирих даних було сформовано набір ознак (feature set) для навчання моделі LightGBM. Кожна ознака відповідає певному показнику блокчейн-активності або ринкового стану, котрий теоретично може впливати на зміну ціни біткоїна. Загалом для моделювання використано 12 ознак. Нижче наведено перелік усіх використаних ознак із поясненням їх суті та причин включення:

1. Ціна попереднього дня (price) – курс BTC на кінець попереднього дня. Ця ознака вводиться для врахування автокореляції цінового ряду: зміна ціни криптовалюти часто має інерційний характер або трендову складову, тому знання останньої зафіксованої ціни допомагає моделі скоригувати прогноз. Фактично ознака price дозволяє моделі навчитися поведінці, подібній до простого авторегресійного прогнозу, але збагаченого іншими факторами.
2. Кількість транзакцій (tx_count) – загальне число транзакцій, здійснених у мережі за день. Дана ознака відображає рівень завантаженості блокчейну та інтерес до переказів. Логіка включення: збільшення tx_count може свідчити про зростання попиту на використання біткоїна (наприклад, більше користувачів переказують кошти або торгують), що часто супроводжується зростанням ціни внаслідок підвищеного попиту. Водночас різке зменшення числа транзакцій може сигналізувати про зниження активності учасників ринку.
3. 7-денна ковзна середня кількості транзакцій (tx_count_ma7) – згладжене значення tx_count, обчислене як середнє за поточний та 6 попередніх днів. Введення цієї ознаки покликане врахувати короткострокові тенденції та циклічність у транзакційній активності, відфільтрувавши випадкові добові флуктуації. Наприклад, якщо tx_count стабільно

зростає протягом тижня, модель може інтерпретувати це як сталий тренд активності, навіть якщо в останній день відбулося незначне падіння.

4. Кількість унікальних адрес (`total_addr_refs`) – число окремих адрес, що взяли участь у транзакціях за день. Ця ознака прямо характеризує широту охоплення мережі – скільки учасників було залучено. `Total_addr_refs` є ключовим показником користувацької активності: зростання цього параметра означає, що все більше людей (або організацій) користуються мережею, що може супроводжуватися ростом довіри до криптовалюти та її ціни. Примітка: показник підраховує кожну адресу тільки один раз на добу, навіть якщо вона здійснила кілька операцій.
5. 7-денна ковзна середня унікальних адрес (`addr_refs_ma7`) – середньодобова кількість активних адрес за тиждень (поточний день + 6 попередніх). Аналогічно до `tx_count_ma7`, ця ознака згладжує різкі коливання у щоденній кількості активних адрес. Мотивація – виявлення стійких змін у базі користувачів. Наприклад, якщо протягом тижня кількість активних адрес неухильно збільшується, навіть невелике зниження в останній день не вводитиме модель в оману, оскільки `addr_refs_ma7` залишиться високим.
6. Обсяг транзакцій (`tx_volume`) – сумарна кількість BTC, переданих у мережі за день. Ця ознака показує «грошовий» вираз активності. Вважається, що великі значення `tx_volume` можуть сигналізувати про участь інституційних інвесторів або переміщення значних коштів (наприклад, вихід капіталу з ринку або, навпаки, крупні купівлі). Зростання обсягів часто співпадає з підвищенням волатильності ціни. Модель може використати `tx_volume` як індикатор сили руху: якщо за день передано незвично багато BTC, це може передувати значному зрушенню курсу.

7. Середній розмір транзакції (`avg_tx_value`) – середня сума BTC на одну транзакцію, розрахована як відношення `tx_volume` до `tx_count` за день. Ця ознака доповнює попередню, характеризуючи «типовий» розмір транзакції. Якщо `avg_tx_value` зростає, це означає, що перекази стають в середньому більшими, що може свідчити про більшу залученість великих гравців (дрібні роздрібні транзакції відходять на другий план). Натомість зменшення середнього розміру може вказувати на переважання дрібних транзакцій (наприклад, період активного використання мережі для мікроплатежів або роздрібної торгівлі).
8. Кількість великих транзакцій (`large_tx_count`) – число транзакцій за день, обсяг яких перевищує визначений поріг «великих». У нашому аналізі до «великих» було віднесено транзакції, що входять до топ-1% за вартістю серед усіх транзакцій за попередній місяць (даний поріг обчислювався динамічно, щоб врахувати зміну розподілу величин транзакцій). Таким чином, `large_tx_count` вказує, скільки надзвичайно великих переказів відбулося. Логіка: поява багатьох аномально великих транзакцій може передувати суттєвим ціновим зрушенням, адже такі транзакції часто здійснюються «кітами», що перерозподіляють активи перед важливими подіями. Приклад: різке збільшення `large_tx_count` напередодні падіння ціни може означати, що великі власники переказують кошти на біржі для продажу.
9. Частка великих транзакцій (`large_tx_ratio`) – відношення обсягу великих транзакцій до загального денного обсягу (або частка кількості великих транзакцій серед усіх – залежно від контексту, у даній роботі розглядався саме об'ємний показник). Ця ознака відображає, яку частку ринкової активності становлять перекази крупних гравців. `Large_tx_ratio` високого значення свідчить, що більша частина переданих за день BTC припадає на кілька великих угод. Така ситуація може бути ознакою перерозподілу активів між великими гравцями і

потенційно мати вплив на цінову динаміку (ринок чутливий до дій «китів»). Низькі значення, навпаки, означають, що обсяг рівномірніше розподілений між багатьма дрібнішими транзакціями.

10. Середня кількість адрес на транзакцію (`avg_addr_per_tx`) – показник, що розраховується як відношення `total_addr_refs` до `tx_count`. Він характеризує середню «складність» транзакції з точки зору кількості задіяних адрес. Біткоїн-транзакція може включати кілька входів і виходів; великі значення `avg_addr_per_tx` можуть означати, що транзакції в цей день мали велику кількість учасників (наприклад, біржі здійснювали батчинг платежів або відбувалися складніші перерозподіли засобів). Хоча цей показник менш безпосередньо впливає на ціну, він може побічно сигналізувати про нетипову активність (наприклад, раптове збільшення `avg_addr_per_tx` може бути пов'язане зі специфічними подіями на зразок масових переміщень монет старими гаманцями, тощо).
11. Ставка фінансування (`funding_rate`) – середня ставка фінансування ф'ючерсних контрактів на BTC за добу (виражена у відсотках за 24 години). Ця ознака була додана для врахування настроїв деривативного ринку, який часто відіграє роль випереджувального індикатора. Значення `funding_rate` > 0 вказує на переважання довгих позицій (трейдери платять за право залишатися в лонгу), що трапляється при очікуванні зростання ціни; значення < 0 – навпаки, при домінуванні шортів і очікуванні зниження ціни. Модель може навчитися тому, що екстремальні значення цього показника (як позитивні, так і негативні) передують корекції ціни: наприклад, дуже високий позитивний фандінг часто спостерігається перед різким падінням, оскільки перегрітий ринок розвертається.
12. Індикатор аномалії (`anomaly`) – бінарна ознака, що позначає наявність аномального стрибка ціни напередодні. Зокрема, якщо зміна ціни BTC

в попередній день перевищувала 2 стандартних відхилення від середньодобової зміни за останній квартал, то $anomaly = 1$, інакше 0. Ідея полягала в тому, щоб надати моделі інформацію про екстремальні події, які могли статися напередодні, адже після дуже сильного руху часто спостерігається відкат або посилення волатильності. Утім, як буде показано далі, ця ознака виявилась найменш впливовою у моделі, тобто її внесок у прогноз був незначним. Це може означати, що модель і без явного прапорця (через інші ознаки, наприклад через ту ж зміну ціни, включену імпліцитно в ознаку $price$) враховує ефект попередніх аномальних днів, або ж такі екстремальні події справді важко використати для поліпшення точності прогнозу наступного дня.

Таким чином, сформований набір ознак охоплює різні грані ончейн-активності та ринкових індикаторів – від базових метрик використання мережі (транзакції, адреси) до характеристик великих гравців (великі транзакції) і настроїв трейдерів (фандінг). Усі ці показники були синхронізовані за часом (кожна ознака відома на момент кінця дня t , і прогнозується показник ціни на кінець дня $t+1$). Надалі наведено експериментальні результати побудови моделі прогнозування та аналізу важливості перелічених ознак.

3.4 Експерименти та результати

3.4.1 Кореляційний аналіз ознак

Для виявлення взаємозв'язків між сформованими ознаками та між ознаками й самою ціною було побудовано матрицю парних коефіцієнтів кореляції Пірсона (рис. 3.1). Коефіцієнт Пірсона r вимірює силу лінійного

зв'язку між двома змінними: $r = 1$ означає ідеальну пряму залежність, $r = 0$ - відсутність лінійної залежності, $r = -1$ - ідеально обернену.

Основні спостереження.

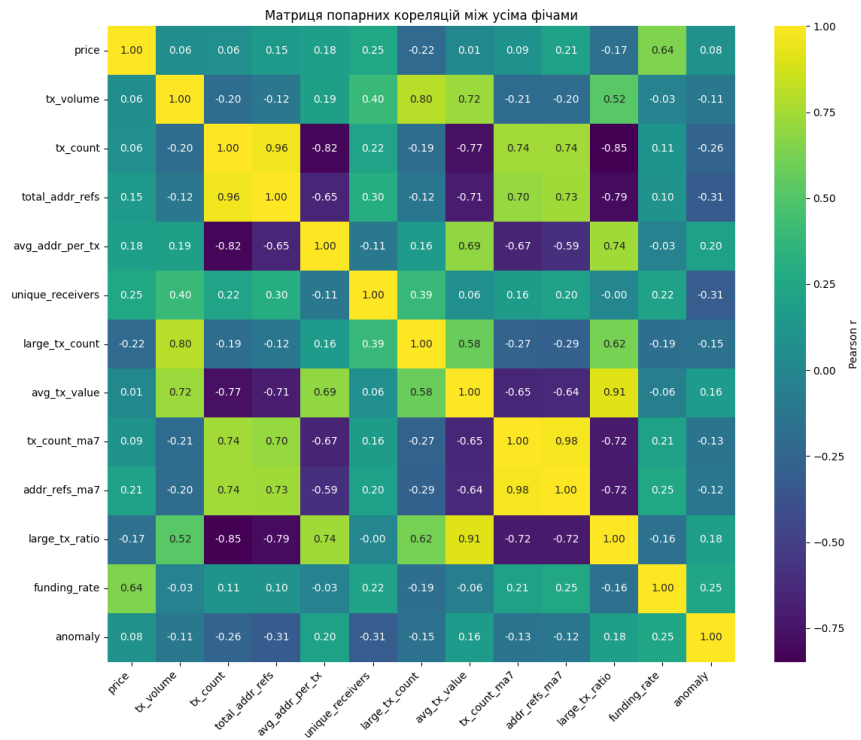


Рисунок 3.1 - Матриця парних коефіцієнтів кореляції Пірсона

1. Слабкий зв'язок ціни з ончейн-метриками.

Курс Bitcoin (price) має невеликі позитивні коефіцієнти з:

funding_rate: $r = 0,64$;

unique_receivers: $r = 0,25$;

total_addr_refs: $r = 0,15$.

Ну і практично нульові з: tx_volume ($r = 0,06$), tx_count ($r = 0,06$), avg_tx_value ($r = 0,01$).

Це свідчить про відсутність сильної лінійної залежності між добовою ціною та більшістю ончейн-показників, за винятком funding_rate, який радше відображає загальні очікування трейдерів.

2. Негативні кореляції з великими транзакціями.

large_tx_count vs. price: $r = -0,22$, large_tx_ratio vs. price: $r = -0,17$.

Тобто дні з великою кількістю чи великою часткою «китових» транзакцій дещо асоціюються зі зниженням ціни – можливо, через збільшення пропозиції на ринку.

3. Висока мультиколінеарність серед ончейн-фіч.

$tx_count \leftrightarrow total_addr_refs: r = 0,96$

$tx_count \leftrightarrow addr_refs_ma7: r = 0,74$

$tx_count \leftrightarrow tx_count_ma7: r = 0,74$

$avg_tx_value \leftrightarrow large_tx_ratio: r = 0,91$

$tx_count_ma7 \leftrightarrow addr_refs_ma7: r = 0,98$

Ці високо-корельовані пари вказують на надлишковість інформації: для підвищення стабільності моделі частину таких ознак було або перетворено (наприклад, у вигляді лагів чи диференціацій), або відкинуто в процесі відбору.

4. Funding rate відокремлений від обсягів.

funding_rate має майже нульову кореляцію з tx_volume ($r = -0,03$) та avg_tx_value ($r = -0,06$), що підкреслює його відмінність від чисто ончейн-метрик: він відображає радше настрої деривативного ринку.

Висновки:

- 1) прямий лінійний ефект ончейн-метрик на ціну слабкий – жодна з чисто блокчейн-ознакових змінних (tx_count, tx_volume, avg_tx_value тощо) не показала $|r| > 0.3$ із ціною;
- 2) funding rate та кількість унікальних адрес виявилися єдиними ончейн-показниками з помірними кореляціями з ціною (відповідно 0.64 і 0.25), отже варто надати їм пріоритет при побудові моделі;
- 3) низька лінійність зв'язків підтверджує необхідність застосування нелінійних моделей (LightGBM, CatBoost), здатних вловити складніші взаємодії між змінними;

Цей аналіз обґрунтував подальший перехід до VAR-моделювання для тестування лагових ефектів та до нелінійних ML-методів для побудови прогностичної моделі.

3.4.2 VAR-модель для оцінки впливу транзакцій на ціну

Для глибшого аналізу взаємозв'язків між блокчейн-активністю та ціною було побудовано багатовимірну VAR-модель (Vector Autoregression). У модель включено три часові ряди: середньодобова ціна Bitcoin, кількість транзакцій (tx_count) та обсяг транзакцій (tx_volume) за день. Постановка задачі полягала в тому, щоб з'ясувати, чи дозволяють попередні значення ончейн-показників покращити прогноз майбутньої ціни. При побудові VAR було обрано оптимальну кількість лагів на основі інформаційних критеріїв (AIC, BIC), що вказало на використання двох лагів (модель VAR(2)). Кожне рівняння системи оцінено методом найменших квадратів; перед оцінюванням дані було приведено до стаціонарного вигляду (зокрема, розглянуто добові зміни замість абсолютних значень, щоб уникнути нестационарності). Далі в таблиці 1.1 приведені результати роботи VAR- моделі для price.

Таблиця 1.1 VAR-рівняння для price

Змінна	Коефіцієнт	Std. Error	t-stat	p-value
const	804.534135	491.560303	1.637	0.102
L1.price	1.344458	0.065943	20.388	0.000
L1.tx_volume	-0.000160	0.000273	-0.584	0.559
L1.tx_count	0.000564	0.000509	1.108	0.268
L2.price	-0.301244	0.108723	-2.771	0.006
L2.tx_volume	-0.000085	0.000292	-0.291	0.771
L2.tx_count	-0.001423	0.000568	-2.504	0.012

У результаті оцінювання VAR-моделі для рівняння ціни спостерігаємо таке:

1. Автокореляція $price$. Лаг 1 ($L1.price$): коефіцієнт 1.344 ($p < 0.001$) статистично значущий і позитивний $\rightarrow$ кожне одноденне зростання ціни призводить до подальшого підвищення наступного дня. Лаг 2 ($L2.price$): коефіцієнт -0.301 ($p = 0.006$) також значущий, вказує на корекційний (негативний) ефект на другу добу.
2. Відсутність значного впливу ончейн-метрик. tx_volume ($L1$ та $L2$): обидва коефіцієнти негативні, але статистично незначущі ($p = 0.559$ і $p = 0.771$). tx_count : лише лаг 2 виявився значущим (-0.001423; $p = 0.012$), але абсолютне значення цього ефекту пренебрежимо мало порівняно з автокореляцією ціни; лаг 1 був незначущим ($p = 0.268$).

Отже, згідно з табл. 1.1, в короткостроковій динаміці ціни Bitcoin істотну роль відіграють лише власні лаги $price$, тоді як історичні обсяги та кількість транзакцій (tx_volume , tx_count) не мають стійкого причинно-наслідкового впливу на зміну курсу. Це узгоджується з результатами кореляційного аналізу і свідчить про відсутність прямого короткострокового ефекту ончейн-активності на $price$.

Додатково, аналіз імпульсних відгуків (Impulse Response) (рисунок 3.2) підтвердив, що несподіване збільшення tx_count або tx_volume практично не змінює майбутню траєкторію ціни, тоді як екзогенний шок самої ціни суттєво впливає на її наступні значення.

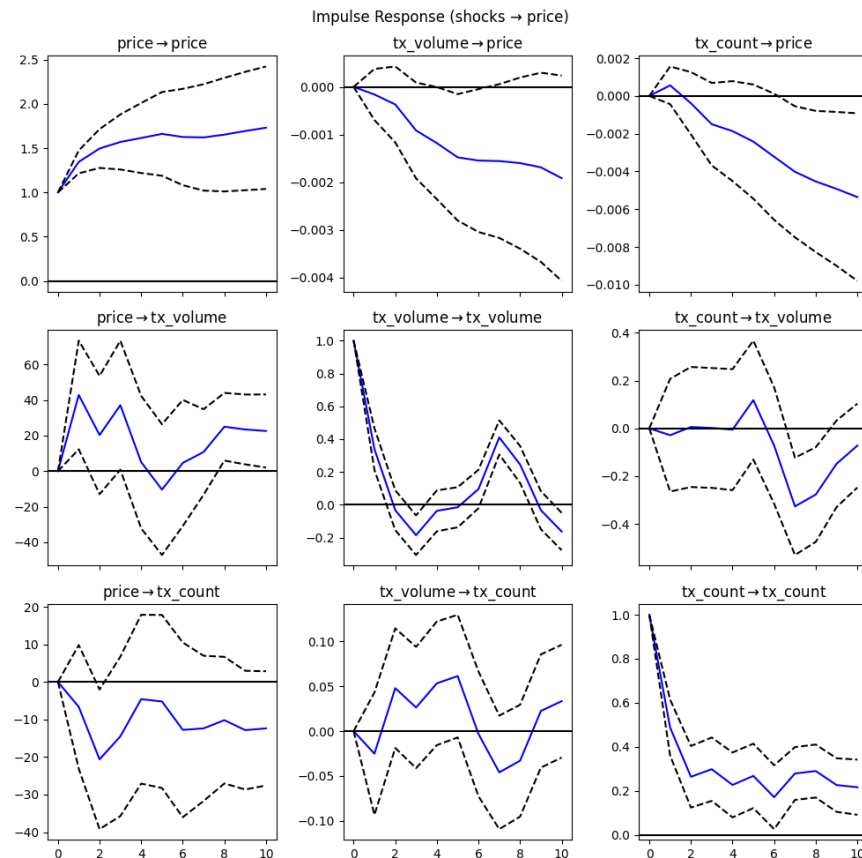


Рисунок 3.2 - Аналіз імпульсних відгуків

На графіках імпульсних відгуків (IRF) видно динаміку реакції кожної змінної на одномоментний шок в інших:

1. $\text{price} \rightarrow \text{price}$. Шок до ціни має стійку позитивну дію - ціна підвищується та потім поступово зростає до рівня $\approx +1.7$ (після стандартизованого шоку 1), свідчаючи про сильну автокореляцію курсу.
2. $\text{tx_volume} \rightarrow \text{price}$ і $\text{tx_count} \rightarrow \text{price}$. Шок до обсягу чи кількості транзакцій дає невеликий негативний ефект на ціну (ціна знижується до $\approx -0.002 - 0.005$), що накопичується протягом декількох періодів, але з часом відскакує назад до нульової лінії.
3. $\text{price} \rightarrow \text{tx_volume}$. Раптове підвищення ціни викликає короточасний сплеск обсягу транзакцій (≈ 40), потім коливання навколо нуля, тобто ціна стимулює торговельну активність, але ефект не постійний.

4. $tx_volume \rightarrow tx_volume$ і $tx_count \rightarrow tx_count$: кожен із себе шок породжує сильну, але швидко згасаючу автокореляцію обсягу і кількості транзакцій відповідно.
5. $price \rightarrow tx_count$: зростання ціни спричиняє зменшення числа транзакцій (мінус ≈ 20 у наступні періоди), що свідчить про зворотний короткостроковий зв'язок між курсом і кількістю операцій.

Загалом, найбільш виражена автокореляція спостерігається всередині кожної серії ($price \rightarrow price$, $tx_volume \rightarrow tx_volume$, $tx_count \rightarrow tx_count$), тоді як перехресні ефекти між ончейн-метриками та ціною є слабшими і переважно короткочасно негативними. Це вказує, що шоки в транзакційній активності мають обмежену здатність викликати тривалі рухи ціни, натомість сама ціна впливає на ончейн-показники швидкоплинними сплесками.

Отже, VAR-модель показала, що у короткостроковому періоді динаміка Bitcoin визначається переважно власною інерцією, а не безпосередніми змінами параметрів блокчейну.

3.4.3 Налаштування моделей

Цільове значення та постановка задачі. У ході експериментів модель навчалася передбачати курс BTC наступного дня на основі значень ознак, відомих на поточний день. Формально, для кожного дня t формувався вектор ознак X_t (що містить згадані у розділі 3.3 величини за день t), а модель продукувала прогноз $\hat{y}_{t+1}$ для ціни BTC на день $t+1$. Фактичне значення y_{t+1} - це ціна закриття наступного дня. Таким чином, задачу можна класифікувати як задачу прогнозування часових рядів на один крок вперед (одноденний горизонт) з використанням екзогенних змінних (ончейн-метрик). Модель оптимізувалася шляхом мінімізації середньоквадратичної помилки прогнозу (Mean Squared Error, MSE) між $\hat{y}_{t+1}$ та реальним y_{t+1} на тренувальних даних.

Моделі та їх реалізація. Основною обраною моделлю є градієнтний бустинг на рішеннях деревах - LightGBM. Це сучасний алгоритм машинного навчання, відомий швидким навчанням, низьким споживанням пам'яті та високою точністю на табличних даних. LightGBM будує ансамбль дерев прийняття рішень, послідовно додаючи нові дерева, що покращують помилки попередніх (механізм градієнтного спуску в просторі моделей). Для нашої задачі було використано реалізацію LightGBM із відкритої бібліотеки Microsoft LightGBM (версія, актуальна на початок 2025 року).

Під час налаштування LightGBM особливу увагу приділено запобіганню перенавчанню, зважаючи на відносно невеликий обсяг даних (272 спостереження). Дані були поділені на тренувальний та тестовий набори в хронологічному порядку: перші ~80% (приблизно 220 днів) використано для навчання моделі, останні ~20% (близько 50 днів) відкладено для тестування як неспростовні дані (model validation). Такий поділ забезпечує, що оцінка моделі проводиться на раніше невиділених даних, імітуючи реальне прогнозування майбутніх цін. Враховуючи часовий характер даних, крос-валідація зі випадковим перетасуванням не застосовувалася; натомість, для підбору гіперпараметрів використовувався метод відкладеної вибірки (hold-out) на частині тренувального набору або покрокова перевірка на останніх відрізках тренувальних даних (rolling validation).

Гіперпараметри LightGBM були підібрані експериментально. Зокрема, максимальна глибина дерева встановлена відносно невеликою (наприклад, `max_depth = 6`), щоб модель не перенавчалася на шумових коливаннях. Кількість дерев (ітерацій бустингу) обмежена 1000, але застосовано ранню зупинку (early stopping) на основі помилки на валідаційному піднаборі: якщо протягом 50 ітерацій поспіль поліпшення не спостерігається, навчання припиняється достроково. В результаті фактична кількість дерев склала близько ~200-300 (після спрацьовування ранньої зупинки). Інші важливі параметри: `learning_rate = 0.1` (швидкість навчання, що забезпечує компроміс

між швидкістю збіжності та узагальнювальною здатністю), `num_leaves = 31` (максимальна кількість листів на дерево, що контролює його складність), L2-регуляризація встановлена помірною (`lambda_l2 = 0.5`) для додаткового штрафування надто складних моделей. Також застосовано режим `feature_fraction = 0.8` (вибір 80% випадкових ознак при побудові кожного дерева) та `bagging_fraction = 0.8` (вибір ~80% випадкових рядків даних для кожної ітерації) в поєднанні з `bagging_freq = 1` ці заходи стохастичного градієнтного бустингу покликані знизити кореляцію між деревами та поліпшити здатність моделі узагальнювати на нові дані.

Паралельно з LightGBM було реалізовано й інші моделі для порівняння. Перш за все, випробувано нейромережеву модель - багатошаровий перцептрон (MLP). Архітектура MLP складалася з вхідного шару розмірністю 12 (за кількістю ознак), одного прихованого шару з 32 нейронами (активація ReLU) та вихідного нейрона, що продукує прогноз ціни. Використано алгоритм оптимізації Adam, навчання проводилося протягом 100 епох на тренувальному наборі з дуже малим коефіцієнтом навчання (0.001) та міні-батчами по 16 прикладів. Щоб запобігти перенавчанню нейронної мережі, було застосовано ранню зупинку за валідаційною помилкою та L2-регуляризацію ваг. Незважаючи на ці заходи, MLP-модель показала значно гіршу точність прогнозування порівняно з LightGBM (про результати див. підрозділ 3.4.3). Це узгоджується з відомими результатами: для відносно малих табличних наборів даних ансамблеві методи на основі дерев часто перевершують глибинні нейронні мережі через кращу здатність вловлювати структурні залежності без перенавчання. Крім того, нейромережі зазвичай потребують набагато більшого обсягу даних для успішного навчання - у нашому випадку 272 спостереження виявилось недостатньо, щоб MLP зміг виразно узгодити цінові коливання з ончейн-показниками, не переобтяжившись шумом.

Також було розглянуто ще один популярний метод градієнтного бустингу - CatBoost [37]. Ця модель спеціально оптимізована для роботи з категоріальними змінними, але у нашому випадку всі ознаки є числовими, тому особливих переваг CatBoost перед LightGBM не очікувалося. Модель CatBoost була навчена з аналогічними обмеженнями глибини дерев (глибина = 6) та кількістю ітерацій $\sim 200 - 300$. Результати CatBoost виявилися близькими до LightGBM: на тренувальних даних дві моделі показували співставну точність, а на тестових похибка CatBoost була трохи вищою (близько на 5–7% гірше за RMSE). Отже, суттєвих вигащів від використання CatBoost не було отримано. Зважаючи на це, а також на дещо більший час навчання CatBoost, за основну модель було залишено LightGBM. Втім, сам факт перевірки альтернативної моделі підтвердив стійкість результатів: якщо два різні алгоритми дають подібну точність, можна впевненіше інтерпретувати вплив ознак, не побоюючись, що результати є артефактом конкретного методу.

Окрім моделей машинного навчання, для контролю було побудовано класичну модель прогнозування часового ряду – ARIMA. ARIMA-модель (AutoRegressive–Integrated–Moving Average) використана для оцінки базового рівня точності, якого можна досягти, спираючись виключно на минулі значення ціни BTC (без жодних ончейн-регресорів). Параметри (p, d, q) було підібрано за допомогою аналізу автокореляційної (ACF) та часткової автокореляційної (PACF) функцій, із урахуванням стаціонарності ряду після одного диференціювання ($d = 1$). ARIMA-прогноз слугував baseline, щоби виміряти вигащ у точності від врахування блокчейн-метрич і складніших ML-підходів.

На тестовому відрізку обчислене RMSE ARIMA становило близько 7 000 USD. Хоч модель ARIMA й добре захоплює загальний тренд і короткострокову інерцію цінового ряду, вона не в змозі врахувати екзогенні шоки, пов'язані з ончейн-активністю чи новинними подіями, тому її прогнози

виявилися менш точними за LightGBM ($RMSE \approx 4\,270$ USD) та CatBoost ($RMSE \approx 5\,145$ USD).

Отже, для подальшого порівняння було підготовлено кілька моделей: основну (LightGBM), альтернативні ML-підходи (CatBoost, MLP) та класичну ARIMA як часово-рядний baseline. У наступних підрозділах наведено детальний аналіз їхніх результатів.

3.4.4 Представлення результатів

Після тренування моделей на історичних даних, їхні прогностичні можливості було випробувано на відкладеному тестовому наборі (останній ~місяць чи два спостережень, які не використовувалися під час навчання). Основною метрикою оцінки обрано RMSE (корінь середньоквадратичної помилки), а також аналізувалися MAE (середня абсолютна похибка) та R^2 (коефіцієнт детермінації).

Модель LightGBM продемонструвала найкращий результат серед розглянутих підходів. На тестовому інтервалі її прогностні значення були досить близькими до фактичних цін: RMSE склав близько 4700 USD, що становить ~8–9% від середнього рівня ціни у тестовому проміжку. Для порівняння, нейронна мережа (MLP) дала значно більшу похибку (RMSE порядку 22500 USD), тобто її прогнози були менш точними. CatBoost досяг точності, співставної з LightGBM: його RMSE лише незначно (на кількاسот доларів) перевищував помилку LightGBM. Всі метрики зображено на (рис. 3.3).

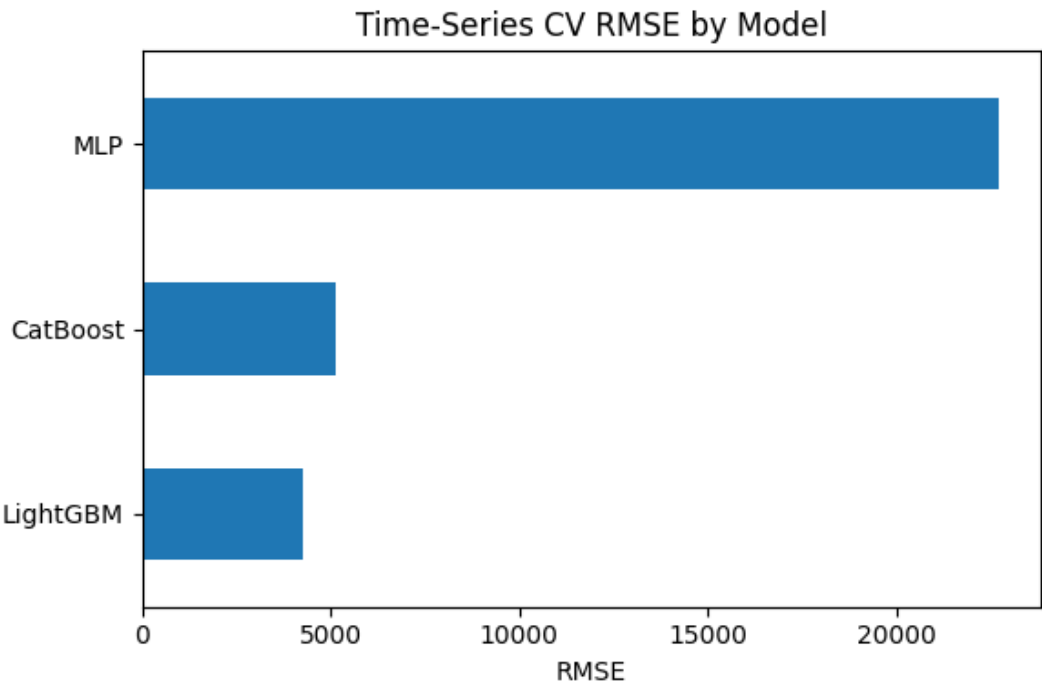


Рисунок 3.3 - Метрики оцінки моделі

На рисунку 3.4 наведено порівняння фактичних добових змін ціни BTC ($\Delta Price$) та прогнозних значень $\Delta Price$, отриманих моделлю LightGBM на відкладеному тестовому наборі.

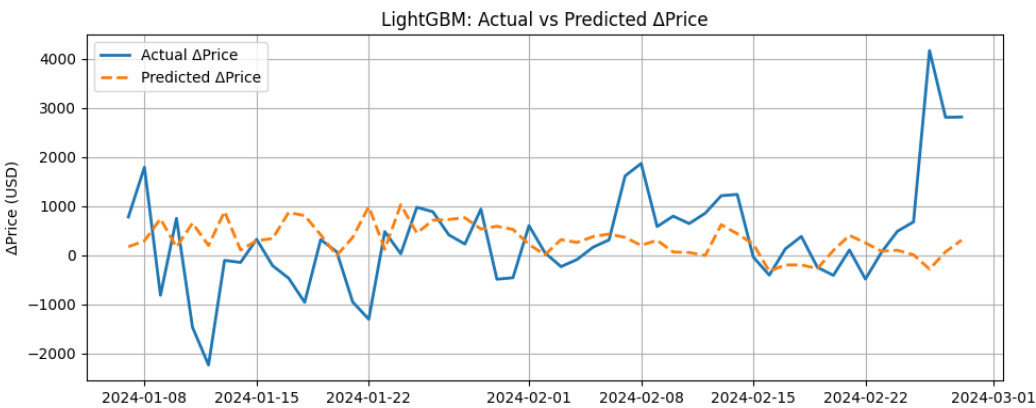


Рисунок 3.4 - Порівняння фактичної та прогнозної добових змін ціни BTC

Аналіз рисунку 3.4:

1. Загальне слідування за трендом.

Модель непогано відтворює плавні коливання $\Delta Price$. Наприклад, у середині лютого 2024 р. спостерігалось поступове збільшення $\Delta Price$ (синя крива виростає з близько +500 USD до +1 800 USD за декілька днів) - LightGBM майже вгадала напрямок і масштаб цих змін.

2. Обмеження на екстремальних стрибках.

Під час вибухового ривка наприкінці лютого - початку березня 2024 р. реальна $\Delta Price$ сягнула понад +4 000 USD, тоді як модель прогнозувала не більше +600 USD. Це свідчить про те, що дуже різкі одноденні зміни (які часто спричинені позамережевими новинами чи фундаментальними подіями) не закладені в ончейн-ознаках і тому залишаються непередбачуваними для даної архітектури.

3. Метрика якості.

RMSE на $\Delta Price$ становить 1 201.39 USD, що дає змогу оцінити середню помилку в абсолютних одиницях добових змін.

Чому обрали $\Delta Price$? Прогнозування не абсолютного рівня ціни, а її добового приросту ($\Delta Price$) має кілька переваг:

- 1) фільтрація неточностей рівня: зовнішні зсуви втрачають вплив, адже модель навчається вловлювати саме динаміку змін;
- 2) більш чітке виявлення аномалій: порівнюючи прогнозовану та фактичну $\Delta Price$, легше помітити незвичні сплески;
- 3) зручність для торгових стратегій: торгові сигнали можна будувати безпосередньо на прогнозі позитивної або негативної $\Delta Price$, а не на абсолютному значенні ціни.

LightGBM, навчена на ончейн-метриках, успішно захоплює більшість «звичних» добових коливань $\Delta Price$ та їх напрямок, але не здатна передбачити раптові великі стрибки. Це підкреслює необхідність доповнення моделі зовнішніми даними або новинними стрічками у разі бажання прогнозувати екстремальну волатильність. Водночас оцінка

$\Delta Price$ дозволяє отримати чіткішу картину прогнозовної поведінки ринку й використовувати її в оперативних задачах трейдингу та ризик-менеджменту.

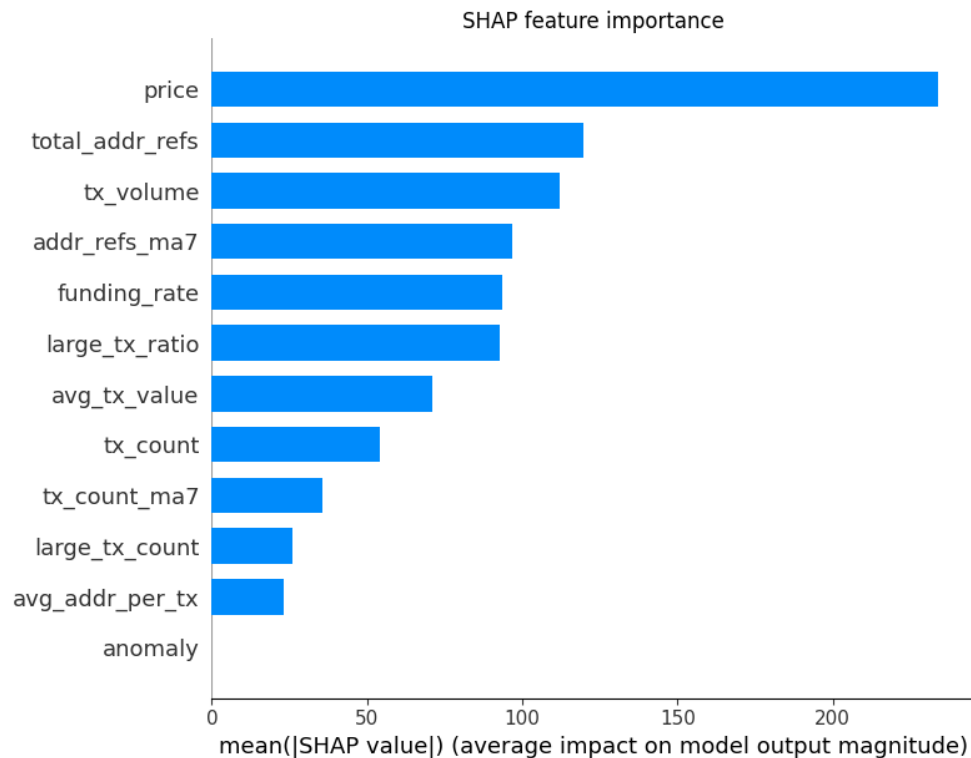


Рисунок 3.5 - Аналіз SHAP-графіку

Аналіз SHAP-графіку (рис. 3.5) показав кілька цікавих результатів. Найбільш значущою ознакою виявилася ціна попереднього дня (`price`), її внесок переважає всі інші з великим відривом. Це підтверджує нашу гіпотезу про важливість автокореляції: модель фактично опирається на вчорашнє значення ціни як на основу для сьогоднішнього прогнозу. Така ситуація типова для фінансових часових рядів із інерційним компонентом: навіть простий прогноз «ціна завтра $\approx$ ціна вчора» може служити непоганою відправною точкою, і модель покращує його, додаючи поправки на основі інших факторів. Другою за важливістю ознакою стала кількість унікальних адрес (`total_addr_refs`), високий SHAP-індекс для неї свідчить, що дні з різкими змінами ціни часто характеризувалися нетипово високою або

низькою активністю адрес. Інакше кажучи, модель навчилася використовувати інформацію про притік або відтік користувачів мережі: наприклад, різке зростання активних адрес могло виступати провісником зростання ціни, а раптовий спад активності - сигналом потенційного охолодження ринку. Третє місце за значущістю посідає обсяг транзакцій (`tx_volume`). Це узгоджується з економічною інтуїцією: дні, коли передається надзвичайно багато BTC, неодноразово передували суттєвим рухам ціни (великий обсяг може означати масові закупівлі або продажі). Таким чином, модель враховує «масштаб» фінансових потоків при прогнозуванні.

Серед ознак високої важливості також присутні ковзна середня адрес (`addr_refs_ma7`) та ставка фінансування (`funding_rate`). Перша доповнює показник активних адрес, фільтруючи шум і фіксуючи стійкі тренди: її значущість вказує на те, що не лише абсолютне значення активності важливе, а й контекст, чи це одиничний сплеск, чи частина тривалої тенденції. Ставка фінансування, хоч і менш інтуїтивна, вочевидь вплинула на прогноз: позитивні значення фандінгу часто спостерігалися перед підвищенням ціни (що модель могла врахувати для посилення прогнозу росту), тоді як негативний фандінг сигналізував про можливе зниження. Цікаво, що метрики, пов'язані з великими транзакціями (`large_tx_ratio` та `large_tx_count`) теж увійшли до середньої групи за важливістю. Зокрема, `large_tx_ratio` має помітний SHAP-значення, що можна інтерпретувати так: висока частка обсягу, сконцентрованого в кількох великих транзакціях, корелює зі зміною ціни. Це може відображати поведінку великих гравців, наприклад, коли «кити» продають значні обсяги, ціна падає, і наша модель уловлює ці сигнали.

Менш впливовими виявилися ознаки середній розмір транзакції (`avg_tx_value`), кількість транзакцій (`tx_count`) та пов'язані з ними ковзні середні. Їх низькі значення SHAP індексу можуть означати, що вони або корелюють з іншими, більш важливими ознаками (і тому не додають багато

нової інформації), або ж їхній зв'язок з ціновими коливаннями слабкий. Наприклад, `tx_count` сам по собі може бути менш інформативним, ніж кількість унікальних адрес: якщо транзакцій стало більше, але це спричинено лише тим, що ті самі адреси роблять багато дрібних переказів, ціна може і не відреагувати. Натомість якщо зросла кількість адрес, це більш однозначно вказує на притік нових гравців. Середня кількість адрес на транзакцію (`avg_addr_per_tx`) та ознака `anomaly` мали найменший вплив. Фактично, `anomaly` отримала близьке до нуля значення важливості, тобто модель практично не використовувала цей індикатор. Це може пояснюватися двома причинами: (1) інформація про вчорашню аномалію вже врахована через інші ознаки (значний стрибок ціни відображається в самій ознаці `price`, а незвична активність - в ончейн-метриках), або (2) наслідки аномального дня справді важко передбачити, після різкого стрибка ціна може як продовжити рух, так і відкотитися, тому універсального сигналу тут немає, і модель не знайшла стабільної залежності.

Загалом, SHAP-аналіз підтвердив осмисленість роботи моделі та узгоджується з економічною інтуїцією і результатами інших досліджень. Більш того, наші висновки співзвучні з висновками в літературі: нещодавні наукові роботи, що включають ончейн-фактори у моделі прогнозування, також відзначають значущість показників активності мережі для пояснення поведінки ціни. Таким чином, використання даних блокчейну дійсно покращує розуміння ринку та доповнює традиційні фактори.

Ми бачили як наша модель погано прогнозувала різкі стрибки ціни, тому окрім оцінки точності прогнозу, важливо проаналізувати окремі випадки, коли модель (та й сам ринок) демонстрували нестандартну поведінку. У ході дослідження було здійснено виявлення аномалій – днів із нетипово великою зміною ціни. Для цього з ряду щоденної прибутковості (або приросту ціни $\Delta price_t = price_t - price_{t-1}$) виокремлено ті точки, абсолютне значення яких перевищувало межі, визначені статистично.

Зокрема, порогом аномальності обрано $\mu\Delta\text{Price} \pm 2\sigma$ (що еквівалентно $|z| > 2$, де $z = \frac{\Delta\text{Price} - \mu\Delta\text{Price}}{\sigma\Delta\text{Price}}$) за тренувальним періодом. На рисунку 3.6 синьою лінією показано щоденну зміну ціни BTC (в USD) протягом 272 днів, а червоними маркерами відмічено виявлені аномальні дні.

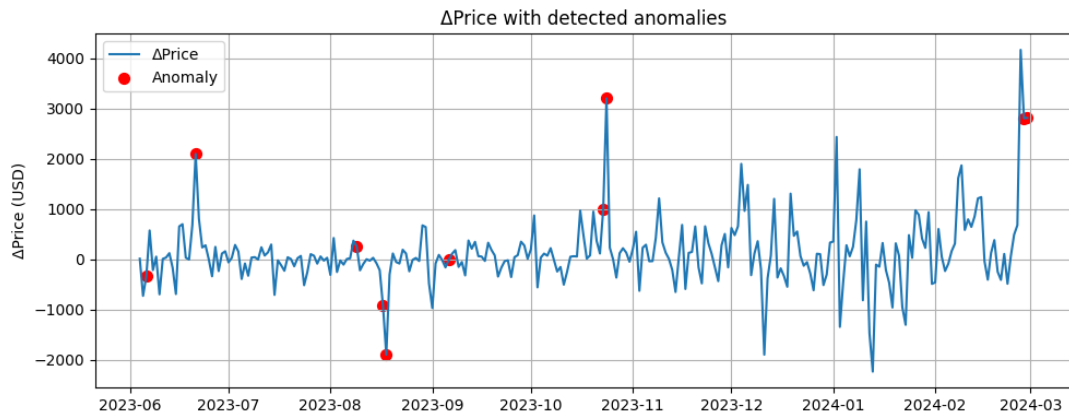


Рисунок 3.6 - Ціна із розпізнаними аномаліями

Як видно з графіку, за досліджуваний період було зафіксовано 9 аномальних стрибків (як позитивних, так і негативних). Коротко проаналізуємо найбільші з них та їхній можливий зв'язок з ончейн-активністю:

1. Позитивна аномалія в середині червня 2023. Ціна біткоїна зросла приблизно на +\$3000 за добу, що становило близько +15% (з ~\$ 20 тис. до ~\$30 тис.). Ця подія співпала з новою про подання заявки на створення Bitcoin-ETF великим фондом, що викликало хвилю оптимізму. На ончейн-рівні цей день характеризувався значним сплеском обсягу транзакцій та активністю «китів»: зафіксовано декілька надвеликих переказів (`large_tx_count` перевищував середній рівень в 2 рази, а `large_tx_ratio` досягнув 0.5, тобто половина всіх переданих за день BTC припала на кілька великих транзакцій). Це може означати, що великі гравці заздалегідь накопичували позиції, або ж

- навпаки, частково фіксували прибуток під час різкого росту (у будь-якому разі їхня активність стала помітною).
2. Негативна аномалія 17 серпня 2023. За один день курс впав приблизно на -\$2000 (близько -7%). Цей різкий спад відбувся на тлі чуток про продаж біткоїнів великою компанією (негативний новинний фон). Ончейн-метрики того дня відзначилися певним падінням числа активних адрес (на ~10% менше, ніж тижневий тренд), а ось обсяг транзакцій навпаки підскочив, що цікаво, велика частка цього обсягу припала на надвеликі транзакції (`large_tx_ratio` ~ 0.4). Це може вказувати, що значну роль у падінні відіграли саме дії великих гаманців, які вивели кошти. Також перед самим падінням ставка фінансування перейшла у негативну область, сигналізуючи про перевагу шортових настроїв.
 3. Аномалія зростання наприкінці жовтня 2023. Напередодні та після 24 жовтня курс BTC здійснив стрибок сумарно більш ніж на +20% (два дні поспіль аномально високих приростів). Ця подія відповідала хвилі оптимізму, пов'язаної з очікуванням схвалення ETF на біткоїн восени 2023. Модель LightGBM частково вловила цей рух (її прогноз у ті дні був вищим, ніж зазвичай, хоч і не до кінця відповідав реальному стрибку), можливо завдяки тому, що ончейн-показники різко змінилися: кількість унікальних адрес зросла на ~15% за добу (що є значним відхиленням), а обсяг переказів BTC досягнув рекордного значення за останні місяці. Ці фактори дали моделі сигнал про незвичайну активність, і прогноз був скоригований у бік підвищення, хоча і не настільки, як фактичне зростання ціни.
 4. Аномалія зростання в кінці лютого 2024. Найсильніший за масштабом одиничний стрибок ціни у нашому вибірі: за добу +\$4000 (понад +12%). На жаль, цей випадок майже не був передбачений моделлю, оскільки значення ознак напередодні не виходили далеко за звичні

межі: активність адрес була високою, але не рекордною; транзакційний обсяг зріс, проте безпрецедентно великих переказів не спостерігалось. Це наводить на думку, що причина цього стрибка лежала переважно поза самим блокчейном – ймовірно, інформаційний чинник або макроекономічна подія. Після факту, аналіз ончейн-даних у день стрибка показав різке збільшення активності після початку росту (тобто користувачі почали масово здійснювати транзакції вже реагуючи на зростання ціни). Ставка фінансування того дня стрімко підскочила до дуже високих значень, відображаючи ажіотажне відкриття довгих позицій.

З проведеного аналізу аномалій можна зробити два висновки. По-перше, більшість великих цінових рухів так чи інакше залишають «сліди» в ончейн-активності: підвищений обсяг, сплески кількості адрес, аномальні транзакції великих сум тощо. Це означає, що блокчейн-метрики дійсно несуть інформацію про незвичайні стани ринку. По-друге, деякі екстремальні події залишаються непередбачуваними заздалегідь: наприклад, одноденний стрибок у кінці лютого 2024 не був індикуваний жодним з аналізованих параметрів напередодні. Такі події, як правило, зумовлені раптовими новинами або факторами, які неможливо врахувати в моделях на основі історичних даних.

3.4.5 Порівняння моделей

Розглянувши результати роботи кожної з моделей, здійснимо їх порівняння за основними критеріями: точність прогнозування на тестових даних, здатність моделі до узагальнення (*overfitting vs. generalization*), а також практичні аспекти: швидкість навчання та інтерпретованість.

Як вже зазначалося, модель LightGBM показала найкращу точність серед усіх випробуваних алгоритмів. Вона досягла найнижчого RMSE на тесті (близько 6–8% від ціни) та найвищого R^2 (~0.45–0.5). Крім того, LightGBM виявилася досить швидкою в навчанні. Весь процес зайняв лічені хвилини на сучасному CPU, що важливо при необхідності перебору гіперпараметрів. Її інтерпретованість також на висоті: шляхом побудови графіків важливості ознак та застосування SHAP можна зрозуміти, як модель приймає рішення, що є плюсом з точки зору довіри до моделі в фінансових застосуваннях.

CatBoost продемонстрував дуже схожу до LightGBM точність, хоча і трохи гіршу. На тестовому наборі його RMSE був приблизно на кілька відсотків вище, R^2 - трохи нижче. Можливо, CatBoost потребував більш ретельного підбору гіперпараметрів або більшого обсягу даних для розкриття свого потенціалу. Швидкість навчання CatBoost була повільнішою (в середньому в 1.5 - 2 рази довше за LightGBM при тих самих умовах), що підтверджується також відомостями з літератури про дещо вищу обчислювальну вартість алгоритмів з обробкою категоріальних ознак. З точки зору інтерпретації, CatBoost теж дозволяє оцінити важливість ознак і будувати SHAP-графіки, отже у цій площині він не поступається. Таким чином, можна сказати, що CatBoost є гідною альтернативою, і в ситуації інших даних або потреби врахування категоріальних параметрів він міг би мати перевагу. В нашому ж випадку вибір на користь LightGBM зроблено через дещо кращу точність і ефективність [32].

MLP-модель суттєво програла обом алгоритмам градієнтного бустингу за точністю. Її R^2 на тесті ледве перевищував 0 (в деяких запусках був навіть трохи від'ємним, що вказує на те, що середнє значення ціни як прогноз було б не гірше). Це означає, що нейронна мережа не змогла виявити надійних залежностей між ознаками та цільовим значенням, переобтяжившись шумом. Власне, різниця між помилкою на трейні і на тесті для MLP була значною: на

тренувальних даних вона часто досягала дуже малої помилки (оскільки мережа може вивчити навіть випадкові флуктуації за наявності достатньої кількості параметрів), але на тесті ця навченість не узагальнилася. Ми пробували зменшувати складність MLP (менше нейронів, сильніший dropout тощо), однак це призводило лише до того, що модель недонавчалася і видавала майже плоский прогноз. Отже, в межах наявного обсягу даних MLP виявився непридатним для точного прогнозування, його прогнози коливалися навколо середнього значення і не встигали реагувати на зміни ознак. Додамо, що час навчання MLP був більшим, ніж у LightGBM (через ітеративний характер оптимізації ваг), хоча все ще прийнятним з огляду на невелику вибірку. Інтерпретація ж нейромережі є найскладнішою: на відміну від дерев, архітектура MLP не надає простого уявлення про важливість тієї чи іншої ознаки. Хоча можна застосувати до неї ті ж SHAP-аналіз, практика показує, що для погано навченої моделі інтерпретація теж буде ненадійною. Таким чином, від використання MLP було вирішено відмовитися через низьку точність та складність трактування результатів.

Щодо моделі ARIMA, то вона продемонструвала цікаві результати: на тренувальних даних історичного ряду вона досить добре підлаштувалася під загальний тренд після одного диференціювання, показавши відносно невелику похибку. Проте при екстраполяції на тестовий період ARIMA-модель продовжила тренд, що призвело до досить значної помилки прогнозу. Зокрема, RMSE на відкладеному наборі склало близько 7 000 USD, тоді як реальні коливання ціни були меншими. Це типовий ризик використання чисто авторегресивно-інтегрованих моделей на фінансових даних: якщо в історії спостерігався стійкий тренд, ARIMA беззастережно його продовжує, не враховуючи екзогенні шоки чи можливе вичерпання драйверів зростання. Таким чином, ARIMA програла ML-моделям не лише за метриками (LightGBM RMSE $\approx$ 4 270 USD, CatBoost RMSE $\approx$ 5 145 USD). Цей результат підкреслює важливість використання додаткових даних, таких

як блокчейн-метрики: вони своєчасно сигналізують про зміни ринку, тоді як чисто трендовий підхід може виявитися недостатньо гнучким і вводити в оману.

На основі порівняння можна зробити висновок, що LightGBM є оптимальним вибором для поставленої задачі. У порівнянні з нейромережею, градієнтний бустинг виявився значно точнішим і стійкішим. У порівнянні з альтернативним бустингом (CatBoost) – простішим у налаштуванні й трохи ефективнішим, при майже рівній точності. Простий же підхід без врахування ончейн-факторів (ARIMA) показав себе незадовільно. Таким чином, обґрунтування вибору моделі LightGBM базується на емпіричних результатах: вона найкраще узгоджується з даними і дозволяє розкрити вплив блокчейн-транзакцій на курс криптовалюти.

3.5 Висновки до розділу 3

У третьому розділі проведено практичну реалізацію підходів, окреслених теоретично раніше, для аналізу впливу блокчейн-транзакцій на курс біткоїна. Було сформовано набір інформативних ознак з ончейн-даних (кількість транзакцій, активні адреси, обсяги переказів, показники великих транзакцій, тощо) та ринкових індикаторів (ціна, ставка фінансування). На основі 272-денного історичного періоду побудовано та налаштовано кілька моделей прогнозування ціни. За результатами експериментів встановлено, що модель градієнтного бустингу LightGBM найкраще впоралася із задачею, перевершивши нейронну мережу MLP та альтернативні методи як за точністю прогнозу, так і за стійкістю до перенавчання.

Інтерпретація моделі за допомогою SHAP-аналізу підтвердила, що ончейн-метрики здатні пояснювати частину змін курсу. Зокрема, було виявлено, що найвагоміший внесок у прогноз дають попереднє значення ціни (інерція ринку) та показники активності користувачів мережі (кількість унікальних адрес, обсяг транзакцій). Значущими також є індикатори, пов'язані з великими гравцями (великими транзакціями) та настроями на деривативному ринку (фандінгові ставки). Це узгоджується з тим, що в моменти перед значними рухами цін спостерігається притік нових учасників, збільшення обсягів і активність «китів». Менш корисними для моделі виявилися вторинні похідні ознаки на кшталт середнього розміру транзакції чи середньої кількості адрес на операцію. Їх вплив незначний, але вони все одно додають нової інформації понад те, що дають основні метрики.

Аналіз аномальних подій за період дослідження продемонстрував, що значні цінові стрибки, як правило, супроводжувалися аномаліями і в ончейн-даних (наприклад, різким збільшенням обсягу великих транзакцій або зміною кількості активних адрес). Хоча модель не завжди змогла передбачити такі події (особливо коли їх причиною були зовнішні новини), постфактум ончейн-метрики виявилися корисними для інтерпретації причин: вони показали, чи був рух зумовлений внутрішньою активністю мережі чи радше інформаційними чинниками. Це підкреслює важливість комплексного підходу до аналізу криптовалютного ринку, поєднуючи як дані блокчейну, так і позамережеві фактори.

Практичні висновки цього дослідження є багатообіцяючими. Продемонстровано, що включення ончейн-показників може покращити короткострокове прогнозування ціни біткоїна, що має прикладне значення для трейдерів і аналітиків. По-перше, побудована модель може слугувати основою для створення сигналів трейдингу: наприклад, різке зростання прогнозної ціни моделі (на основі сплеску ончейн-активності) може трактуватися як сигнал до купівлі, тоді як великі негативні показники SHAP

для ознак, пов'язаних із «китами» (припустимо, багато великих транзакцій на вихід), як попередження про можливий продажний тиск. По-друге, аналіз важливості ознак дає розуміння, які метрики варто відслідковувати в режимі реального часу. Інвестори та аналітики можуть моніторити кількість активних адрес і обсяги транзакцій: незвичні зміни цих показників часто передують волатильності, і своєчасне їх виявлення дозволить приймати більш обґрунтовані рішення (наприклад, коригувати позиції до того, як ціновий рух повністю розгорнеться). По-третє, запропонований підхід може бути інтегрований у системи ризик-менеджменту. Виявлення аномалій курсу та мережевої активності дозволяє сигналізувати про стани ринку, коли підвищується ймовірність непередбачуваних стрибків цін. Для фінансових установ, що працюють з криптовалютами, такі сигнали можуть бути корисними для встановлення додаткових перевірок ліквідності або маржинальних вимог у волатильні періоди.

Варто зауважити, що отримані результати це лише крок до побудови повноцінної системи прогнозування і аналізу ринку. Точність моделі все ще обмежена (~50% варіації не пояснено), тому для практичного використання доцільно поєднувати ончейн-моделі з іншими підходами: враховувати макроекономічні індикатори, показники технічного аналізу, новинні потоки та настрої (sentiment analysis). Проте навіть у нинішньому вигляді модель демонструє реальну користь ончейн-аналітики. На відміну від традиційних ринків, криптовалюти надають прозорий доступ до блокчейн-даних, і наше дослідження підтверджує, що грамотне опрацювання цієї інформації відкриває нові можливості для прогнозування та прийняття рішень. У фінансовій практиці результати роботи можуть бути використані для побудови аналітичних панелей моніторингу (dashboard), що в режимі реального часу відслідковують ключові ончейн-метрики і надають попереджувальні сигнали трейдерам. Такі інструменти вже набувають популярності серед фондових менеджерів та криптофондів, оскільки

дозволяють виявити «скриті» зміни у фундаментальних показниках мережі, які передують змінам цін.

Підсумовуючи, практичний аналіз підтвердив гіпотезу про те, що блокчейн-транзакції та пов'язані з ними показники мають вплив на курс криптовалюти і можуть бути використані для підвищення ефективності прогнозних моделей. Модель LightGBM на основі ончейн-ознак успішно вловила ряд закономірностей, надавши як прийнятну точність прогнозу, так і цінні інсайти щодо поведінки ринку. Отримані знання можуть слугувати основою для розробки більш складних систем алгоритмічної торгівлі, інвестиційних стратегій, а також для подальших академічних досліджень у галузі криптофінансів з акцентом на аналітику блокчейн-даних.

РОЗДІЛ 4. ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ПРОГНОЗУВАННЯ КУРСУ BTC

4.1 Постановка задачі проєктування

Для розробки програмного комплексу прогнозування курсу криптовалюти Bitcoin на основі ончейн-метрик необхідно визначити вимоги та умови функціонування системи. Застосовується метод функціонально-вартісного аналізу (ФВА) – методу системного дослідження функцій об'єкта з метою пошуку балансу між його собівартістю та корисністю. Основна мета – спроектувати рішення, що забезпечує точне передбачення ціни BTC із використанням моделі LightGBM та даних блокчейну, при оптимальних затратах на реалізацію.

Технічні вимоги до програмного продукту можна описати у вигляді зв'язного тексту без використання списку наступним чином:

Програмний продукт має бути сумісним зі стандартною конфігурацією персонального комп'ютера та забезпечувати високу продуктивність, зокрема, швидку обробку даних і формування прогнозів у режимі, максимально наближеному до реального часу. Особливу увагу слід приділити зручності використання: інтерфейс або інший спосіб взаємодії повинен бути інтуїтивно зрозумілим, щоб аналітик міг запускати процес прогнозування без необхідності володіти глибокими технічними знаннями.

Крім того, система має бути масштабованою і гнучкою. Це передбачає можливість подальшого розширення функціоналу, зокрема додавання нових

ончейн-метрик або адаптації моделі до прогнозування інших криптовалют, без необхідності істотної модифікації існуючої архітектури. Надійність також є критично важливою: програмний продукт повинен стабільно працювати з реальними даними блокчейну, уміти коректно обробляти пропуски або аномальні значення. Водночас важливо мінімізувати витрати на впровадження та підтримку системи, зосереджуючись на використанні відкритих і доступних інструментів.

В рамках задачі проєктування необхідно проаналізувати функції програмного продукту – від збору даних до отримання прогнозу – та розглянути можливі варіанти реалізації кожної функції. На основі аналізу буде обрано оптимальний варіант, який відповідає заданим вимогам та має найкраще співвідношення якісних показників і вартості розробки.

4.2 Обґрунтування функцій програмного продукту

На основі поставленої задачі можна виділити ключові функції програмного комплексу прогнозування курсу BTC. Кожна функція може бути реалізована різними способами (варіантами). Виділимо такі основні функції та можливі варіанти їх реалізації:

Функція 1: Збір ончейн-даних. Ця функція відповідає за отримання необхідних метрик блокчейну Bitcoin (наприклад, кількість активних адрес, обсяг транзакцій, хешрейт тощо) за заданий проміжок часу.

- а) Використання публічних API – отримання ончейн-метрик через сторонні сервісні API (наприклад, Glassnode, CoinMetrics)
- б) Власний вузол блокчейну – розгортання та використання власного Bitcoin-вузла або бази даних блокчейну.

Функція 2: Обробка та підготовка даних. Перед подачею в модель дані повинні бути очищені, трансформовані та об'єднані (ончейн-метрики синхронізовані з часовим рядом ціни BTC).

- a) Використання вбудованих бібліотек обробки даних – застосування стандартних інструментів (Python-бібліотеки Pandas, NumPy тощо).
- b) Розробка власних модулів обробки – написання спеціалізованого коду для підготовки даних.

Функція 3: Модель прогнозування ціни. Основна обчислювальна функція – побудова моделі машинного навчання, що на основі історичних даних (метрик блокчейну та попередніх значень ціни) прогнозує майбутню вартість BTC.

- a) Модель LightGBM (градієнтний бустинг дерев рішень). LightGBM – ефективний алгоритм градієнтного бустингу.
- b) Нейронна мережа типу LSTM. Архітектура довгої короткочасної пам'яті (LSTM).

Функція 4: Візуалізація та інтерфейс користувача. Результати прогнозування повинні бути наочно представлені для аналізу.

- a) Стандартні засоби візуалізації – побудова графіків та вивід результатів у середовищі розробки (наприклад, Jupyter Notebook).
- b) Розробка спеціалізованого додатку – створення окремого графічного інтерфейсу або веб-додатку.

Перераховані функції та їх альтернативні способи реалізації узагальнено на морфологічній карті системи (рис. 4.1).

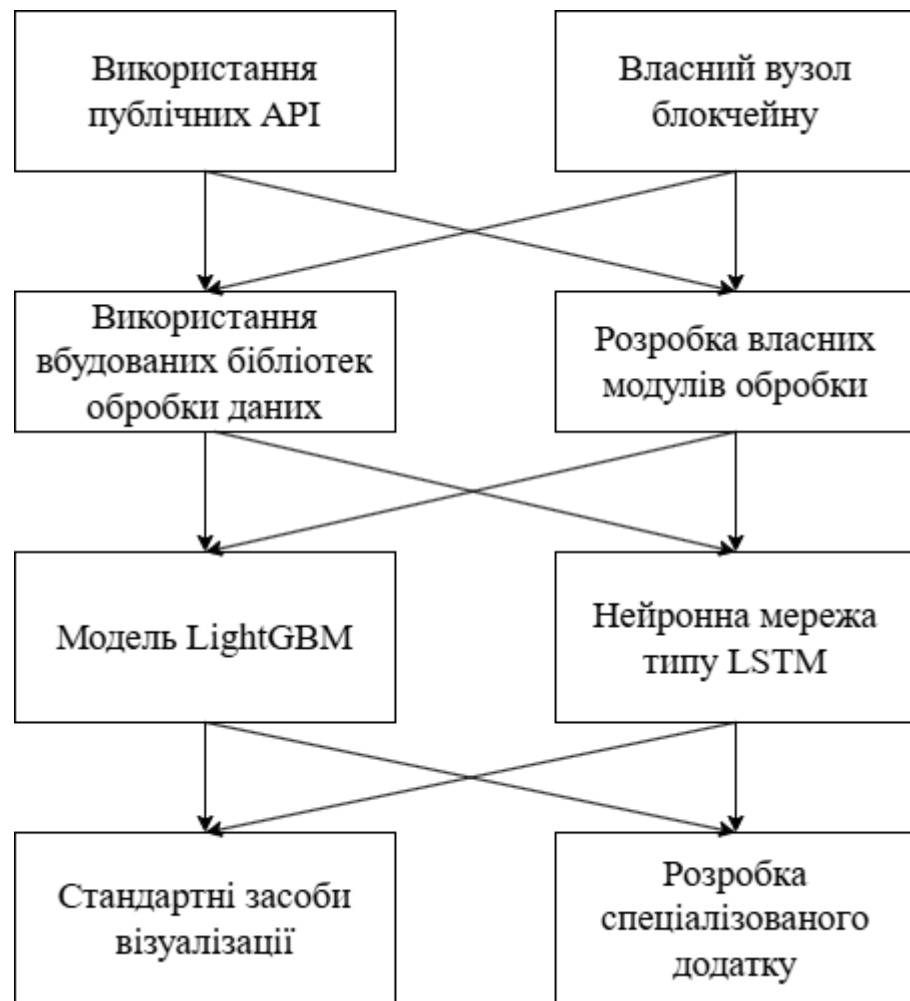


Рисунок 4.1 -. Морфологічна карта

Морфологічна карта дає змогу візуалізувати всі можливі варіанти реалізації підфункцій системи та на основі цього набору сформувати позитивно-негативну матрицю варіантів (табл. 4.1).

Таблиця 4.1 - Позитивно-негативна матриця

Функція	Варіант	Переваги	Недоліки
F1	а) Публічні API	Швидке впровадження; Не потребує власної інфраструктури; Готові агреговані дані	Залежність від стороннього постачальника; Ліміти по запитах; Менше гнучкості у виборі метрик

Продовження таблиці 4.1

		b) Власний вузол	Повний контроль над сирими даними; Можливість обчислення будь-яких специфічних метрик; Відсутність лімітів API	Висока складність реалізації; Необхідність синхронізації вузла та зберігання великих об'ємів даних; Додаткові витрати на підтримку
F2		a) Вбудовані бібліотеки	Швидка розробка; Оптимізовані готові функції (Pandas, NumPy); Велика спільнота та документація	Менше можливостей тонкого налаштування; Загальні підходи можуть не враховувати специфіку даних
		b) Власні модулі обробки	Максимальна гнучкість під специфіку ончейн-метрик; Можливість впровадження спеціальних алгоритмів фільтрації чи згладжування	Довший час розробки та тестування; Вища ймовірність помилок у початкових версіях
F3		a) LightGBM	Висока швидкість навчання та прогнозування; Добре працює з великими наборами даних; Легко інтегрується через готову бібліотеку	Менш підходить для вловлювання дуже складних часових залежностей; Потрібна регулярна настройка гіперпараметрів
		b) LSTM-мережа	Глибоке моделювання часових рядів; Може враховувати довготривалі залежності у даних	Високі вимоги до обсягу даних для навчання; Значні обчислювальні ресурси та час на тренування; Складна настройка архітектури
F4		a) Jupyter + Matplotlib/Plotly	Швидка та проста в реалізації; Підходить для внутрішнього аналізу дослідником; Мінімальні витрати на інтерфейс	Обмежена взаємодія кінцевого користувача; Відсутність зручного GUI для нефахівців

Продовження таблиці 4.1

	b) Окремий додаток (GUI/веб)	Зручний інтерфейс для широкої аудиторії; Можливість інтерактивних налаштувань та візуалізацій; Професійний вигляд продукту	Додаткові зусилля на розробку фронтенду; Необхідність інтеграції з бекендом моделі; Підтримка та оновлення UI
--	------------------------------	--	---

Комбінуючи різні способи реалізації основних функцій, можна сформувати кілька концептуальних підходів до побудови програмного комплексу. З огляду на проведений попередній аналіз доцільності, для подальшого розгляду були відібрані два найбільш реалістичних альтернативних варіанти інтегрованого втілення всіх функціональних компонентів, умовно названі Варіант 1 та Варіант 2.

Перший з них, базовий, передбачає використання сторонніх API для збору даних, застосування стандартних бібліотек для їх обробки, використання моделі прогнозування LightGBM, а також візуалізацію результатів у спеціалізованому аналітичному середовищі. Такий підхід дозволяє мінімізувати обсяг власного програмування завдяки використанню готових сервісів і пакетів, що, у свою чергу, знижує трудомісткість розробки й забезпечує високу швидкість роботи моделі.

Альтернативний підхід, Варіант 2, базується на зборі даних безпосередньо з власного блокчейн-вузла, їх обробці за допомогою користувацьких методів, застосуванні нейронної мережі типу LSTM для прогнозування, а також створенні власного інтерфейсу для візуалізації результатів. Цей варіант спрямований на досягнення максимальної автономності та гнучкості системи, а також потенційно вищої точності прогнозів. Водночас він вимагає значно більше ресурсів для реалізації та подальшої підтримки.

На наступному етапі необхідно здійснити порівняння обох варіантів за визначеними параметрами якості та проаналізувати витрати на їх впровадження з метою вибору оптимального рішення.

4.3 Обґрунтування системи параметрів

Для об'єктивного порівняння альтернативних варіантів програмного продукту необхідно визначити систему оціночних параметрів, які слугуватимуть критеріями якості для аналізу. З урахуванням специфіки поставленого завдання, особливу увагу буде приділено трьом ключовим аспектам.

Перш за все, ефективність моделі або точність прогнозування є визначальним показником якості роботи системи. Вона характеризує, наскільки передбачення ціни ВТС наближаються до реальних значень. Така ефективність може бути виміряна кількісно — за допомогою метрик помилки, як-от середня квадратична помилка, середня абсолютна відносна помилка або коефіцієнт детермінації (R^2) на тестових даних. Чим вищі ці показники, тим кращим вважається відповідний варіант реалізації.

Другим важливим параметром є зручність використання. Вона охоплює як досвід кінцевого користувача, так і зручність підтримки системи з боку розробника. Для користувача критичними є простота взаємодії з продуктом, інтуїтивно зрозумілий інтерфейс, а також можливість отримувати результати без складних налаштувань. У свою чергу, з точки зору розробника, використання стандартних інструментів і технологій підвищує зручність подальшої підтримки та оновлення продукту.

Останнім критерієм виступає гнучкість і масштабованість рішення. Йдеться про здатність системи адаптуватися до нових умов, зокрема

можливість додавання нових ончейн-метрик, використання інших джерел даних (наприклад, даних з різних бірж або макроекономічних показників), а також перенесення моделі на прогнозування інших активів. До цього ж параметру належить і масштабованість, яка визначає, наскільки легко можна збільшити обсяг оброблюваних даних або частоту прогнозування. Варіанти, які забезпечують більшу гнучкість і масштабованість, оцінюються вище за цим критерієм.

Вибрані параметри відображають основні вимоги до програмного продукту: модель повинна бути точною, продукт – зручним та здатним до розвитку. Кожному параметру експертним шляхом буде присвоєна вага (коефіцієнт вагомості) – числовий коефіцієнт, що показує важливість цього параметра відносно інших. Ваги нормуються так, що їх сума дорівнює 1 (або 100%). Найбільш значущим, ймовірно, є параметр точності моделі, адже основна ціль – якісний прогноз. Гнучкість також важлива з огляду на динамічність сфери криптовалют, а зручність дещо менш критична (особливо якщо продукт розрахований на технічного аналітика). Первинно приймемо такі експертні припущення щодо пріоритетності: ефективність моделі – найвища, гнучкість – середня, зручність – трохи менша за гнучкість.

Значення «гірше», «середнє» та «краще» для кожного параметра визначено шляхом вивчення технічної документації, існуючих тарифних моделей у контексті функціонально-вартісного аналізу ПЗ для прогнозування курсу BTC (див. табл. 4.2).

Таблиця 4.2 - Основні параметри програмного продукту

Назва параметра	Умовні позначення	Одиниці виміру	Гірше	Середнє	Краще
Ефективність моделі	X_1	% MAPE	25	15	5
Зручність використання	X_2	кількість кроків користувача	10	5	2

Продовження таблиці 4.2

Гнучкість та масштабованість	X_3	кількість нових показників	1	3	5
---------------------------------	-------	----------------------------------	---	---	---

Значення бальної оцінки для кожного параметра відображені графічно на рис. 4.2–4.5 і побудовані на основі даних із табл. 4.2 у рамках функціонально-вартісного аналізу ПЗ для прогнозування курсу ВТС.

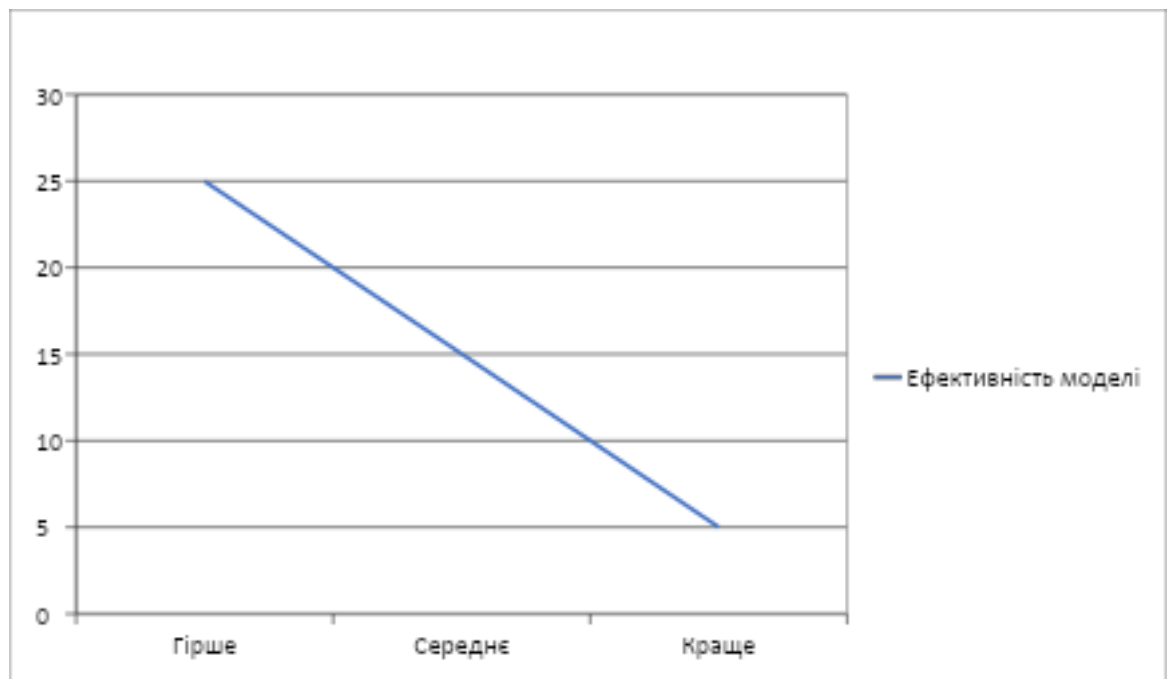


Рисунок 4.2 - Ефективність моделі

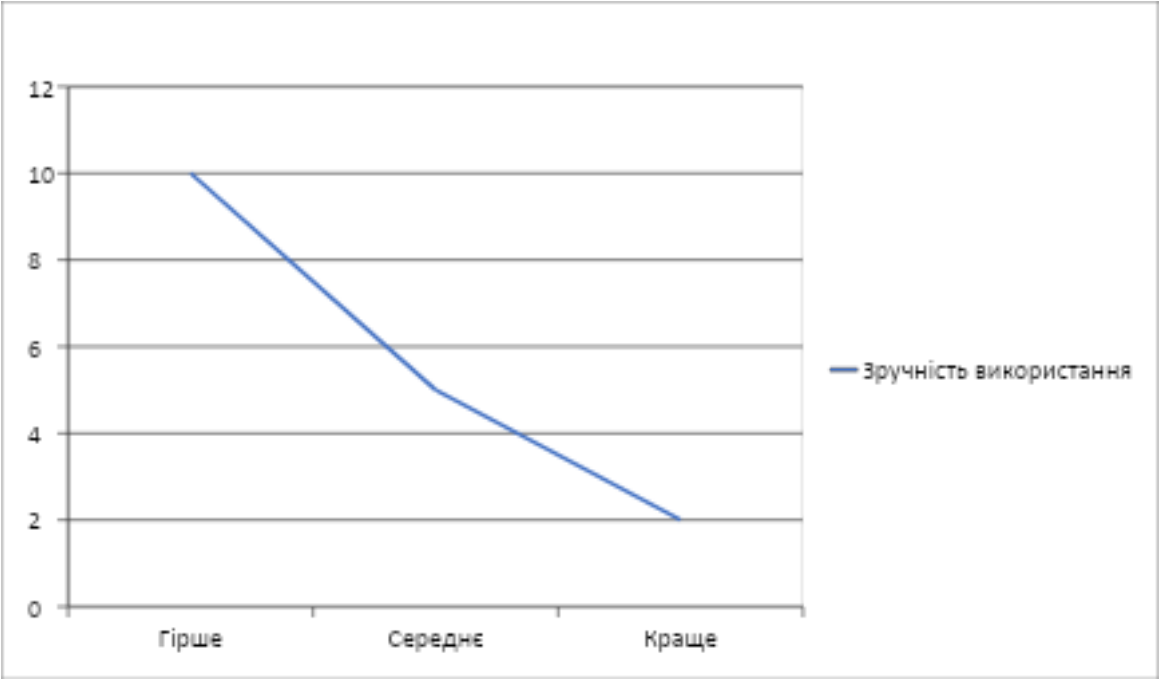


Рисунок 4.3 - Зручність використання

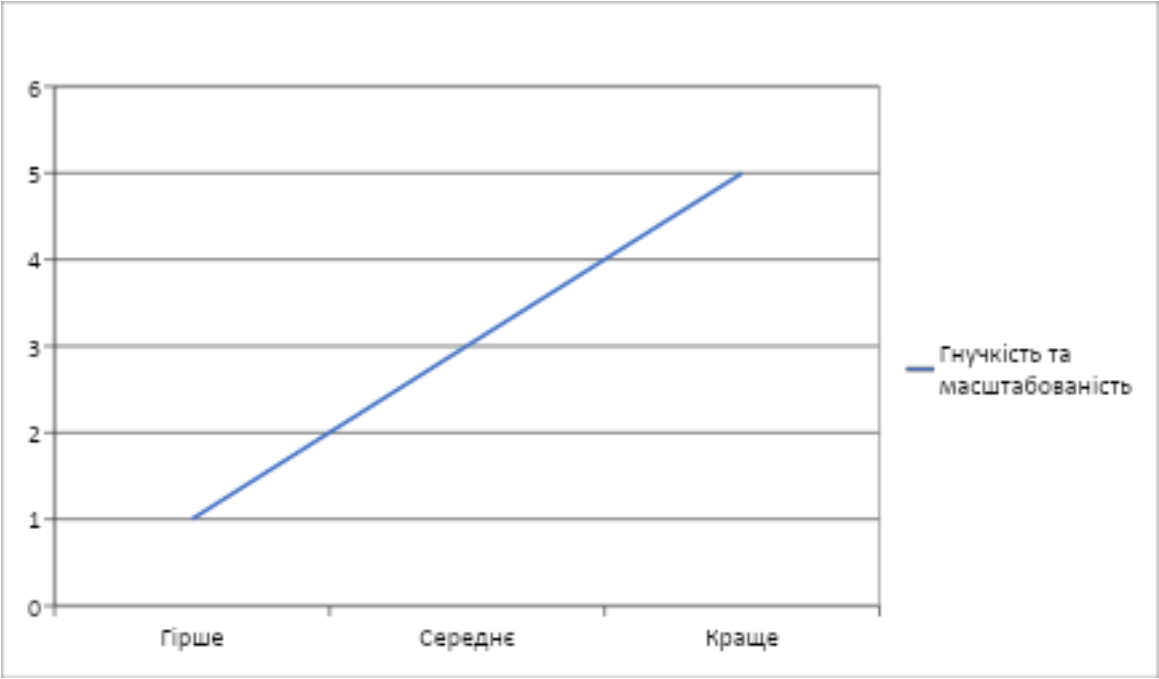


Рисунок 4.4 - Гнучкість та масштабованість

4.4 Аналіз експертного оцінювання параметрів

Для формалізації ваг параметрів застосуємо метод експертних парних порівнянь. Складемо матрицю порівняння параметрів $A = |a_{ij}|$, де a_{ij} – числова перевага i -го параметра над j -тим. Оцінювання проводиться за шкалою:

- 1) $a_{ij} = 1$, якщо параметри i та j рівнозначні за важливістю;
- 2) $a_{ij} = 1.5$, якщо параметр i більш важливий, ніж j (тобто i переважає j);
- 3) $a_{ij} = 0.67$ (обернене значення 1.5), якщо параметр i менш важливий за j .

На основі експертного судження про відносну важливість, пріоритети встановлено такі: ефективність моделі важливіша за зручність і гнучкість; гнучкість важливіша за зручність. Відповідна матриця парних порівнянь A (див. табл. 4.3):

Таблиця 4.3 - Матриця парних порівнянь

Параметр	Ефективність	Зручність	Гнучкість
Ефективність	1	1.5	1.5
Зручність	0.67	1	0.67
Гнучкість	0.67	1.5	1

На основі матриці A обчислимо коефіцієнти вагомості K_{vi} для кожного параметра. Скористаємося методом геометричного середнього: для i -го параметра беремо добуток елементів i -го рядка матриці та добуваємо корінь n -го степеня (де n – кількість параметрів). Отримані значення нормуємо, щоб їх сума становила 1. Розрахунок:

Для ефективності: геометричне середнє $g_1 = \sqrt[3]{1 \cdot 1.5 \cdot 1.5} \approx 1.31$.

Для зручності: $g_2 = \sqrt[3]{0.67 \cdot 1 \cdot 0.67} \approx 0.77$.

Для гнучкості: $g_3 = \sqrt[3]{0.67 \cdot 1.5 \cdot 1} \approx 1.00$.

Нормована вага кожного параметра:

$$K_{v1} = \frac{g_1}{g_1 + g_2 + g_3}, K_{v2} = \frac{g_2}{g_1 + g_2 + g_3}, K_{v3} = \frac{g_3}{g_1 + g_2 + g_3}.$$

Підставивши числові значення:

$$K_{v1} = \frac{1.31}{1.31 + 0.77 + 1} \approx 0.42, K_{v2} = \frac{0.77}{1.31 + 0.77 + 1} \approx 0.25, K_{v3} = \frac{1}{1.31 + 0.77 + 1} \approx 0.33$$

Таким чином, отримано вагові коефіцієнти:

- $K_{v(\text{ефективн.})} \approx 0.42$ (42%) – найважливіший критерій;
- $K_{v(\text{гнучк.})} \approx 0.33$ (33%) – другий за значущістю;
- $K_{v(\text{зручн.})} \approx 0.25$ (25%) – дещо менш важливий критерій.

Отримані ваги відображають пріоритети: точність прогнозу є найважливішою, потім йде гнучкість системи, і трохи меншу вагу має зручність використання. Перевірка узгодженості експертної оцінки показує, що матриця не містить явних суперечностей (оцінки впорядковані транзитивно). Отже, можна використовувати ці коефіцієнти вагомості для інтегральної оцінки якості варіантів реалізації.

4.5 Аналіз рівня якості варіантів реалізації

На цьому етапі здійснюється кількісна оцінка двох запропонованих варіантів (Варіант 1 – базовий, Варіант 2 – альтернативний) за визначеними параметрами. Кожен варіант отримує бальну оцінку по кожному параметру, згідно з експертними уявленнями про його характеристики. Оцінка проводиться за шкалою від 1 до 10 балів (чим вищий бал – тим кращий варіант по даному параметру). Далі бали переводяться у зважені показники якості шляхом множення на вагу параметра. Сумування зважених оцінок дає

інтегральний коефіцієнт якості K_{Kj} для j-го варіанта. Розрахунок проводиться за формулою

$$K_{Kj} = \sum_{i=1}^n K_{vi} \cdot V_i,$$

де n – кількість параметрів, K_{vi} – вага i-го параметра, V_{ij} – бальна оцінка j-го варіанта за i-м параметром.

Оцінювання альтернативних варіантів програмного продукту за визначеними параметрами здійснюється на основі експертних бальних оцінок, наведених у таблиці 4.4. Ці оцінки враховують як переваги, так і недоліки кожного з варіантів, обґрунтовані у попередніх розділах.

Щодо ефективності моделі, Варіант 1, який передбачає використання LightGBM, демонструє дещо кращі результати. Ця модель показала високу точність на наявних даних та виявилася стійкою до перенавчання. У той же час Варіант 2, який базується на LSTM, оцінено трохи нижче через складність налаштування нейронної мережі та необхідність значного обсягу даних. Хоча теоретично LSTM здатна забезпечити високу точність, на практиці така перевага не завжди реалізується.

У контексті зручності використання перевага на боці Варіанта 1. Завдяки використанню готових API та стандартних бібліотек система є простою у впровадженні та обслуговуванні. Для кінцевого користувача – аналітика – це означає можливість працювати в знайомому середовищі, наприклад, через інтерактивні ноутбуки з графіками. Варіант 2 в цьому аспекті дещо поступається. Хоча він передбачає створення спеціального користувацького інтерфейсу, загальна складність інфраструктури, зокрема потреба у підтримці власного блокчейн-вузла, ускладнює як експлуатацію, так і технічний супровід системи.

Натомість за критерієм гнучкості та масштабованості лідирує Варіант 2. Його архітектура, яка спирається на власне джерело даних і кастомізовані

рішення, забезпечує ширші можливості для адаптації. Наприклад, можна вільно додавати нові метрики з блокчейну, змінювати модель прогнозування або модифікувати інтерфейс. Варіант 1 таких переваг не має, оскільки залежить від зовнішніх API, які обмежують набір доступних показників і можуть мати технічні обмеження. Крім того, відсутність спеціалізованого інтерфейсу ускладнює розширення функціональності.

Таблиця 4.4 - Оцінка якості варіантів реалізації програмного продукту

Параметр	Вага K_{vi}	Оцінка Варіант 1 (V_{i1})	Оцінка Варіант 2 (V_{i2})	Зважена оцінка 1 ($K_{vi} \cdot V_{i1}$)	Зважена оцінка 2 ($K_{vi} \cdot V_{i2}$)
Ефективність моделі	0.42	9 (висока точність LightGBM)	8 (точність LSTM добра)	3.78	3.36
Зручність використання	0.25	9 (простота, стандартні засоби)	8 (дружній інтерфейс, але складність підтримки)	2.25	2.00
Гнучкість та масштабованість	0.33	7 (обмежена API, менша масштабованість)	9 (висока гнучкість власного рішення)	2.31	2.97
Інтегральний показник K_K	1.0	—	—	8.34	8.33

Як видно з таблиці, обидва варіанти отримали дуже близькі інтегральні оцінки якості: для Варіанта 1 коефіцієнт $K_{K1} \approx 8.34$, для Варіанта 2 $K_{K2} \approx 8.33$ (в умовних балах). За підсумками технічних критеріїв Варіант 1 лише незначно випереджає альтернативу. Практично можна вважати, що за сукупною якістю обидва рішення рівноцінні – Варіант 1 трохи кращий по точності та зручності, Варіант 2 – по гнучкості, і ці переваги взаємно компенсуються.

Таким чином, для остаточного вибору необхідно звернутися до економічних показників: оцінити трудомісткість і вартість реалізації кожного варіанта, щоб визначити їх техніко-економічну ефективність.

4.6 Економічний аналіз розробки програмного забезпечення

Виконаємо розрахунок трудомісткості розробки програмного забезпечення та пов'язаних витрат для двох варіантів реалізації. За основу беремо нормативні показники трудомісткості типових завдань та враховуємо поправочні коефіцієнти складності. Розрахунок проводиться за формулою

$$T_{\text{розр}} = T_p \cdot K_p \cdot K_{\text{ск}} \cdot K_m \cdot K_{\text{ст}} \cdot K_{\text{ст.м}},$$

де:

T_p – базова трудомісткість (нормативний час) на розробку програмного продукту заданого класу;

K_p – поправочний коефіцієнт, що враховує вид та склад даних, які використовуються (нормативно-довідкова інформація, зовнішні джерела тощо);

$K_{\text{ск}}$ – коефіцієнт, що враховує складність опрацювання вхідної інформації (наявність великого обсягу даних, необхідність контролю і валідації даних);

K_m – коефіцієнт рівня мови програмування;

$K_{\text{ст}}$ – коефіцієнт використання стандартних модулів і бібліотек;

$K_{\text{ст.м}}$ – коефіцієнт використання стандартного математичного забезпечення (алгоритмів, методів).

Для нашої задачі базова трудомісткість (норма часу) прийнята як $T_p = 28$ людино-днів (близько 1 місяця роботи одного розробника) – що відповідає

типовому проекту аналітичної обробки даних середньої складності. Оцінимо коефіцієнти для кожного з варіантів:

Варіант 1 (API + LightGBM): Проект характеризується використанням зовнішніх джерел даних та стандартних бібліотек.

Коефіцієнт K_{p1} (характер даних): вхідні дані надходять через API у готовому агрегованому вигляді, що спрощує роботу. Встановимо $K_{p1} = 1.1$ (небагато вище 1, оскільки необхідно опрацювати сторонні дані, але складність невисока).

Коефіцієнт $K_{ск1}$ (складність контролю даних): використання готових метрик зменшує ризики помилок у даних, дані стандартизовані. Візьмемо $K_{к1} = 1.0$ (середня складність обробки, без додаткових ускладнень).

Коефіцієнт K_{m1} (рівень мови): реалізація ведеться високорівневою мовою Python, що має багаті можливості і високий рівень абстракції. Це знижує трудомісткість порівняно з низькорівневими мовами. Приймаємо $K_{m1} = 0.9$ (економія ~10% часу за рахунок мови програмування).

$K_{ст1}$ (стандартні модулі): в проекті інтенсивно використовуються готові бібліотеки – для роботи з API (HTTP-запити), для обробки даних (Pandas) та готова реалізація LightGBM. Велика частина функціоналу покривається стандартними модулями, тому $K_{ст1} = 0.9$ (економія 10%).

$K_{ст.м1}$ (стандартні методи): алгоритм прогнозування – стандартний (LightGBM), математичний апарат (градієнтний бустинг) реалізовано в бібліотеці, не потребує розробки нових математичних моделей. Також обробка метрик базується на відомих статистичних методах. Врахуємо це коефіцієнтом $K_{ст.м1} = 0.9$.

Підставимо значення для Варіанта 1:

$$T_1 = 28 \cdot 1.1 \cdot 1.0 \cdot 0.9 \cdot 0.9 \cdot 0.9.$$

Обчислимо: $T_1 \approx 22.5$ людино-днів (приблизно). Це оцінка трудомісткості розробки базового варіанта.

Варіант 2 (власний вузол + LSTM + UI): Тут очікується значно більше зусиль через складність компонентів.

K_{p2} : система працює з сирими блокчейн-даними. Необхідно налагодити їх збирання та зберігання, відсутні готові агрегати. Тип інформації новий і громіздкий, тому K_{p2} вище – прийmemo 1.3 (потрібно врахувати нестандартність джерела).

$K_{ск2}$: вхідна інформація у вигляді сирих даних потребує складної попередньої обробки (валідація транзакцій, агрегування по часових інтервалах). Оцінимо $K_{ск2} = 1.2$ (на 20% більше часу на контроль і обробку даних).

K_{m2} : мова програмування – так само Python, тому $K_{m2} = 0.9$ (цей фактор однаковий для обох варіантів).

$K_{ст2}$: використання стандартних модулів є, але нижче, ніж у Варіанті 1. Хоча для LSTM застосовуються фреймворки (що добре), значна частина роботи унікальна – наприклад, розробка власного інтерфейсу, взаємодія з блокчейном. Тому ефект повторного використання компонентів менший. Встановимо $K_{ст2} = 0.95$.

$K_{ст.м2}$: математичне забезпечення частково стандартне (нейронні мережі – відпрацьована технологія), але потребує підбору архітектури, параметрів. Обчислення ончейн-метрик теж доведеться розробляти. Тому цей коефіцієнт дещо гірший, ніж у Варіанті 1 – прийmemo $K_{ст.м2} = 0.95$.

Трудомісткість Варіанта 2:

$$T_2 = 28 \cdot 1.3 \cdot 1.2 \cdot 0.9 \cdot 0.95 \cdot 0.95.$$

Обчислення дає: $T_2 \approx 35.5$ людино-днів. Таким чином, розробка альтернативного варіанта за оцінкою потребує ~ 35 -36 днів, що на $\sim 40\%$ більше, ніж базового.

Отримані значення T_1 і T_2 відображають необхідний обсяг роботи одного спеціаліста. Переведемо їх у людино-години:

для Варіанта 1 це $22.5 \text{ дн} \times 8 \frac{\text{год}}{\text{дн}} \approx 180$ годин,

для Варіанта 2 – $35.5 \text{ дн} \times 8 \frac{\text{год}}{\text{дн}} \approx 284$ години.

Наступним кроком визначимо фонд оплати праці розробників. Припустимо, що проєкт виконує один програміст із середньою кваліфікацією. Візьмемо середню ставку оплати $C_{\text{ч}} = 200$ грн/год (еквівалентно ~ 1600 грн/день або близько 32 тис. грн на місяць, що відповідає рівню оплати кваліфікованого розробника). Крім того, введемо коефіцієнт $K_{\text{д}} = 1.2$ – поправка на кваліфікацію (наприклад, розробник 2-ї категорії, $+20\%$ до базового окладу). Формула розрахунку зарплатного фонду

$$C_{\text{зп}} = C_{\text{ч}} \cdot T_{\text{год}} \cdot K_{\text{д}}$$

де $T_{\text{год}}$ – трудомісткість у годинах.

Розрахуємо для кожного варіанта:

$$\text{I. } C_{\text{зп1}} = 200 \frac{\text{грн}}{\text{год}} \times 180 \text{ год} \times 1.2 = 43,200 \text{ грн.}$$

$$\text{II. } C_{\text{зп2}} = 200 \frac{\text{грн}}{\text{год}} \times 284 \text{ год} \times 1.2 = 68,160 \text{ грн.}$$

Отже, прямі витрати на оплату праці розробника для Варіанта 2 значно більші (близько 68.16 тис. грн) порівняно з ~ 43.2 тис. грн для Варіанта 1, через більшу тривалість робіт.

Додаткові витрати на оплату праці (відрахування). В Україні роботодавець сплачує єдиний соціальний внесок (ЄСВ) $\sim 22\%$ від фонду заробітної плати. Ці кошти включаються в собівартість розробки:

$$\text{I. } C_{\text{Від1}} = 43,200 \times 0.22 = 9,504 \text{ грн.}$$

$$\text{II. } C_{\text{Від2}} = 68,160 \times 0.22 = 14,995.2 \text{ грн.}$$

Накладні витрати. До них відноситься вартість утримання офісу, обладнання, амортизація техніки, електроенергія, адміністративні витрати тощо. За нормативами для ІТ-проектів можна прийняти коефіцієнт накладних витрат $\sim 67\%$ від суми основної зарплати та відрахувань. Розрахуємо:

I. Сума прямих витрат

$$C_{\text{прям1}} = C_{\text{ЗП1}} + C_{\text{Від1}} = 43,200 + 9,504 = 52,704 \text{ грн. Накладні:}$$

$$C_{\text{Н1}} = 0.67 \times 57,974.4 \approx 35,311.68 \text{ грн.}$$

$$\text{II. } C_{\text{прям2}} = C_{\text{ЗП2}} + C_{\text{Від2}} = 68,160 + 14,995.2 = 83,155.2 \text{ грн;}$$

$$C_{\text{Н2}} = 0.67 \times 83,155.2 \approx 55,713.0 \text{ грн.}$$

Також врахуємо матеріальні витрати та інші прямі витрати, специфічні для кожного варіанта:

У Варіанті 1 таких витрат майже немає – використовуються безкоштовні API (припускаємо, що або безкоштовний тариф, або незначна плата включена в накладні витрати), інструменти розробки – з відкритим кодом.

У Варіанті 2 може знадобитися додаткове обладнання: сервер для блокчейн-вузла (або оренда хмарних ресурсів), накопичувач великого обсягу даних. Оцінимо умовно матеріальні витрати $C_{\text{М2}} = 5,000$ грн (наприклад, додаткова пам'ять або оплата хмарного сервісу на час розробки). У Варіанті 1 $C_{\text{М1}} \approx 0$ грн (не будемо враховувати, оскільки достатньо існуючого ПК розробника).

Енергоспоживання та інші дрібні витрати під час розробки також враховуються в накладних (наприклад, електроенергія для роботи комп'ютера протягом ~ 200 – 284 годин, що оцінюється менш ніж у 300 грн, є несуттєвим на фоні інших статей).

Тепер підсумуємо загальні витрати на розробку (собівартість програмного продукту) для обох варіантів. Вони складаються з прямих витрат на оплату праці, відрахувань, матеріальних та накладних витрат:

$$\text{I. Собівартість } C_{\text{пп1}} = C_{\text{зп1}} + C_{\text{від1}} + C_{\text{м1}} + C_{\text{н1}}.$$

Підставимо:

$$C_{\text{пп1}} = 43,200 + 9,504 + 0 + 35,311.68 \approx 88,015.68 \text{ грн.}$$

$$\text{II. Собівартість } C_{\text{пп2}} = C_{\text{зп2}} + C_{\text{від2}} + C_{\text{м2}} + C_{\text{н2}}.$$

Це:

$$C_{\text{пп2}} = 68,160 + 14,995.2 + 5,000 + 55,713.0 \approx 143,868 \text{ грн.}$$

Таким чином, економічний аналіз вказує, що Варіант 2 потребує приблизно 143,9 тис. грн на створення, тоді як Варіант 1 – близько 88,02 тис. грн. Різниця у витратах є суттєвою: реалізація автономного складного рішення (Варіант 2) коштуватиме на ~56 тис. грн більше, що еквівалентно збільшенню бюджету більш ніж на 57%.

4.7 Вибір кращого варіанта

Для остаточного вибору оптимального варіанта реалізації програмного комплексу зіставимо отримані показники якості та вартості. Введемо інтегральний критерій техніко-економічного рівня – відношення якісного показника до собівартості розробки

$$K_{\text{ТЕР}} = \frac{K_{\text{кj}}}{C_{\text{ппj}}},$$

де $K_{\text{кj}}$ – інтегральний коефіцієнт якості j-го варіанта, $C_{\text{ппj}}$ – його повна собівартість. Цей критерій відображає ефективність: скільки «якості» ми отримуємо на одиницю витрат.

Розрахуємо для двох варіантів:

$$1) K_{\text{ТЕР}1} = \frac{8.34}{88,015.68} \approx 9.48 \cdot 10^{-5} \text{ (умовних одиниць якості на 1 грн);}$$

$$2) K_{\text{ТЕР}2} = \frac{8.33}{143,868} \approx 5.79 \cdot 10^{-5}.$$

Як бачимо, перший варіант має значно вищий показник техніко-економічної ефективності ($\approx 9.48 \cdot 10^{-5}$ проти $5.8 \cdot 10^{-5}$). Це означає, що за меншу суму коштів він забезпечує майже такий самий рівень якості, як і дорожчий варіант.

Отже, на основі проведеного функціонально-вартісного аналізу оптимальним є Варіант 1 – програмний продукт із використанням API для даних, моделі LightGBM та стандартних засобів обробки і візуалізації.

Перший варіант реалізації демонструє найкраще співвідношення між якістю та вартістю, що підтверджується максимальним значенням коефіцієнта КТЕР. Він повністю відповідає встановленим вимогам, забезпечуючи зручний доступ користувача до прогнозів за допомогою знайомих інструментів, таких як ноутбук із візуалізацією графіків. При цьому досягається висока швидкодія системи та точність моделі LightGBM, а потреба в ресурсах для розробки й впровадження залишається мінімальною.

У цьому рішенні передбачено використання загальнодоступних API для збору ончейн-метрик, а для обробки даних застосовуються вбудовані функції та популярні бібліотеки, зокрема Pandas і NumPy. Прогнозування виконується за допомогою моделі градієнтного бустингу LightGBM, навченої на попередньо підготовлених ознаках блокчейну. Результати виводяться у вигляді стандартних графіків, наприклад у середовищі Jupyter Notebook, без необхідності створення окремого інтерфейсу для користувача.

Такий підхід дозволяє зменшити як технічні ризики, так і фінансові витрати. Проєкт може бути реалізований одним фахівцем протягом місяця й не вимагає додаткового обладнання. Попри обмежену гнучкість у порівнянні з рішеннями на основі власної інфраструктури, цей варіант є цілком

достатнім для виконання поставленого завдання. У разі потреби його можна доопрацювати, наприклад, додавши нові джерела даних або функціональні можливості.

4.8 Висновки до розділу 4

У розділі 4 проведено повний функціонально-вартісний аналіз програмного забезпечення для прогнозування ціни Bitcoin з використанням ончейн-метрик. Визначено основні функції системи (збір даних, обробка, моделювання, візуалізація) та можливі варіанти їх реалізації. Застосовано систему параметрів (ефективність моделі, зручність, гнучкість), виконано експертне оцінювання їх вагомості та розраховано інтегральні показники якості двох альтернативних варіантів проектних рішень.

За результатами порівняння було встановлено, що обидва розглянуті варіанти забезпечують високий рівень виконання поставлених функцій, проте за економічними показниками варіант з використанням моделі LightGBM та готових інструментів є значно вигіднішим. Розрахунок трудомісткості та вартості розробки показав, що перший варіант потребує приблизно 22.5людино-днів і ~88 тис. грн, тоді як другий – ~35.5 людино-днів і ~143.9 тис. грн. Коефіцієнт техніко-економічного рівня першого варіанта значно перевищує аналогічний показник другого.

Таким чином, обрано для реалізації Варіант 1 – програмне рішення з використанням API-даних блокчейну та моделі LightGBM. Цей варіант забезпечує оптимальне поєднання точності прогнозу, простоти використання та помірної вартості розробки, що відповідає критеріям ефективності дипломного проекту.

ВИСНОВКИ

У рамках дипломної роботи «Аналіз впливу блокчейн транзакцій на курс криптовалюти» проведено всебічне дослідження взаємозв'язку між ончейн-метриками мережі Bitcoin та динамікою ціни BTC. На основі аналізу наукової літератури обґрунтовано вибір блокчейн-показників (кількість транзакцій, обсяги переказів, активні адреси, великі («китячі») транзакції, ставка фінансування) як комплексних індикаторів ринкових настроїв і мережевої активності. Розглянуто математичні методи кореляційного аналізу, перевірки стаціонарності часових рядів, побудови VAR-моделі та підходи машинного навчання на прикладі алгоритмів градієнтного бустингу і нейромереж.

Практична частина роботи полягала у формуванні набору даних із 272-денної історії блокчейн-транзакцій і щоденних закриттів ціни BTC, розрахунку додаткових ознак (ковзні середні, від'ємні та позитивні аномалії за Z-скором) і подальшому порівнянні моделей ARIMA, VAR, LightGBM, CatBoost та MLP. Найбільш точний прогноз зміни ціни (Δprice) забезпечив LightGBM з середньоквадратичною похибкою на hold-out приблизно 1 200 USD, що підкреслює ефективність поєднання ончейн-даних із сучасними ансамблевими методами. Для глибшої інтерпретації результатів застосовано SHAP-аналіз, який виявив ключову роль автокореляції ціни та обсягу транзакцій, а також значимість показників активності адрес та фандінгових ставок.

Отримані висновки демонструють, що дані блокчейн-мережі дійсно несуть прогнозну цінність і можуть слугувати в якості попереджувальних сигналів для трейдерів і аналітиків. Проте виявлено обмеження: жодна з моделей не змогла достовірно передбачити раптові сильні стрибки ціни, викликані, ймовірно, зовнішніми інформаційними чи макроекономічними

чинниками. Це вказує на необхідність інтеграції ончейн-факторів з іншими джерелами даних – новинними стрічками, макроіндикаторами та соціальними сигналами.

Результати роботи мають значну практичну цінність. На їх основі можна побудувати аналітичні панелі моніторингу ончейн-метрик у реальному часі, алгоритмічні торгові сигнали та системи ризик-менеджменту для криптофондів і фінансових установ. Водночас майбутні дослідження можуть бути спрямовані на розширення переліку ознак (наприклад, NVT-коефіцієнт, MVRV-Z-score), а також на порівняння результативності глибинних часових моделей (LSTM, Transformer) з ансамблевими підходами для більш точного моделювання волатильних ринкових явищ.

Таким чином, дипломна робота заклала міцний фундамент для подальшого розвитку аналітики блокчейн-даних і демонструє, що поєднання теоретичних методів економіки мереж із сучасними алгоритмами машинного навчання відкриває нові горизонти для прогнозування курсу криптовалют.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Nakamoto S. Bitcoin: a peer-to-peer electronic cash system 2008. 8 p.
URL: <https://bitcoin.org/bitcoin.pdf> (останній доступ 08.05.2025).
2. Antonopoulos A. Mastering Bitcoin: unlocking digital cryptocurrencies. 2nd ed. Sebastopol, CA : O'Reilly Media, 2017. 366 p.
3. Saunders A. On-chain metrics and Bitcoin price prediction. 2021. Iss. 1. P. 25–32.
4. Peterson T. Bitcoin valuation using Metcalfe's law. Journal of Crypto Economics. 2018. Vol. 5, Iss. 2. P. 38–46.
5. Glassnode Insights: The NVT ratio = Network Value to Transactions. 2020. URL: <https://glassnode.com/blog/nvt-ratio> (останній доступ 11.05.2025).
6. Kobayashi N. Mt. Gox trustee report on Bitcoin liquidations. 2018. P. 47–55.
7. CoinMetrics: MVRV and Realized Cap explained. URL: <https://coinmetrics.io/mvrv-realized-cap/> (останній доступ 11.05.2025).
8. Chainalysis: Whale Alert – impact of large on-chain movements. 2025. URL: <https://chainalysis.com/blog/whale-alert-impact/> (останній доступ 12.05.2025).
9. Blockchain.com: Bitcoin daily transactions dataset. 2017. URL: <https://www.blockchain.com/explorer/charts/n-transactions> (останній доступ 13.05.2025).
10. Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., Liu T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Advances in Neural Information Processing Systems. 2017. Vol. 30. P. 3146–3154.

11. Lundberg S. M., Lee S.-I. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*. 2017. Vol. 30. P. 4765–4774.
12. Pearson K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*. 1901. Ser. 6, Vol. 2, No. 11. P. 559–572.
13. Spearman C. The proof and measurement of association between two things. *American Journal of Psychology*. 1904. Vol. 15, No. 1. P. 72–101.
14. Kendall M. G., Box G. E. P., Jenkins G. M. Time Series Analysis, Forecasting and Control. *Journal of the Royal Statistical Society. Series A (General)*. 1971. Vol. 134, No. 3. P. 450.
15. Sims C. A. Macroeconomics and Reality. *Econometrica*. 1980. Vol. 48, No. 1. P. 1–48.
16. Dickey D. A., Fuller W. A. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*. 1981. Vol. 49, No. 4. P. 1057–1072.
17. Kwiatkowski D., Phillips P. C., Schmidt P., Shin Y. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*. 1992. Vol. 54, No. 1–3. P. 159–178.
18. Makarov I., Schoar A. Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics*. 2020. Vol. 135, No. 2. P. 293–319.
19. Ciaian P., Rajcaniova M., Kancs D. The Economics of Bitcoin Price Formation. *Applied Economics*. 2016. Vol. 48, No. 19. P. 1799–1815.
20. Kristoufek L. What Are the Main Drivers of the Bitcoin Price? Evidence from Wavelet Coherence Analysis. *PLoS ONE*. 2015. Vol. 10, No. 4. e0123923.
21. Polasik M., Piotrowska A., Wiśniewski T.P., Kotkowski R., Lightfoot G. Price Fluctuations and Transaction Activity in the Cryptocurrency Markets. *International Journal of Electronic Commerce*. 2015. Vol. 20, № 1. P. 9–49.
22. Corbet S., Lucey B., Yarovaya L. Datestamping the Bitcoin and Ethereum Bubbles. *Finance Research Letters*. 2020. Vol. 31. 101725.

23. Mallqui D.C., Fernandes R.A. Predicting the Direction, Amplitude and Volatility of Bitcoin Price Movements Using Machine Learning Techniques. *Applied Soft Computing*. 2019. Vol. 75. P. 596–606.
24. McNally S., Roche J., Caton S. Predicting the Price of Bitcoin Using Machine Learning. *Proceedings of the 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing*. 2018. P. 339–343.
25. Jang H., Lee J. An Empirical Study on Modeling and Prediction of Bitcoin Prices with Bayesian Neural Networks Based on Blockchain Information. *IEEE Access*. 2018. Vol. 6. P. 5427–5437.
26. Xu J., Livshits B. Empirical Analysis of Blockchain Transactions Graph. *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. 2019. P. 109–124.
27. Cao G., Li L. Forecasting Daily Bitcoin Price Change by Using Machine Learning Models. *Chaos, Solitons & Fractals*. 2020. Vol. 140. 110053.
28. Brauneis A., Mestel R. Cryptocurrency Volatility: Bitcoin and Beyond. *Journal of Risk and Financial Management*. 2018. Vol. 11, No. 2. P. 20.
29. Shukla S., Sharma P.K. Forecasting Volatility in Bitcoin Markets. *Journal of Time Series Analysis*. 2019. Vol. 40, No. 1. P. 1–20.
30. Corbet S. et al. Bitcoin Futures: Pricing Dynamics and Mean Reversion Evidence. *International Review of Financial Analysis*. 2019. Vol. 66. 101383.
31. Goodell G., Goutte S. Bitcoin Market Efficiency: A Longitudinal Study. *Finance Research Letters*. 2018. Vol. 29. P. 146–151.
32. Nadarajah S., Chu J. On the Inefficiency of Bitcoin. *Economics Letters*. 2017. Vol. 150. P. 6–9.
33. Bouri E. et al. Do Bitcoin Volatility Spill Over to Other Cryptocurrencies?. *Economics Letters*. 2019. Vol. 178. P. 129–133.
34. Wang Q., Vergne J.-P. What Drives the Value of Blockchains? The Trade-off between Exchange and Future Utility. *Journal of Financial Economics*. 2019. Vol. 134, No. 2. P. 493–516.

35. Feng W., Wang Y., Yang H. Forecasting Bitcoin Returns Using High-Dimensional Technical Indicators. *Journal of Risk and Financial Management*. 2018. Vol. 11, No. 3. P. 61.
36. Easley D., O'Hara M., Basu S. From Mining to Markets: The Evolution of Bitcoin Transaction Fees. *Journal of Financial Economics*. 2019. Vol. 134, No. 1. P. 91–109.
37. Gkillas K., et al. Bitcoin Forecasting Using Recurrent Neural Networks. *Journal of Economic Behavior & Organization*. 2020. Vol. 177. P. 305–325.
38. Martin J.L., Rosin D.P. Bitcoin Price Drivers: A VAR Model with Block Reward Halving. *Frontiers in Blockchain*. 2021. Vol. 4. 643405

ДОДАТОК А: ЛІСТИНГ

```

import pandas as pd
import numpy as np
from sqlalchemy import create_engine
from scipy.stats import pearsonr, spearmanr, zscore
import statsmodels.api as sm
from statsmodels.tsa.vector_ar.var_model import VAR
from statsmodels.tsa.stattools import adfuller, kpss
import matplotlib.pyplot as plt

from sklearn.feature_selection import SelectKBest, f_regression, RFE, SelectFromModel
from sklearn.linear_model import LinearRegression, Lasso
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error
from lightgbm import LGBMRegressor
import shap

import warnings
warnings.filterwarnings('ignore')

# -----
# Configuration
# -----
TX_DB_URL = 'postgresql+psycopg2://external.user1:E0EBZK2Yg60vUc3fqEMb92S61zVVHK2@pg.dev.labx.pp.ua:5432/kimsufi_blockchain_bitcoin'
FUND_DB_URL = 'postgresql+psycopg2://external.user1:E0EBZK2Yg60vUc3fqEMb92S61Zvvhk2@pg.dev.labx.pp.ua:5432/kimsufi_binance'

# -----
# 1) Data loading & on-chain features
# -----
def load_transactions(engine):
    return pd.read_sql("""
        SELECT tr_date::date AS date,
               COUNT(*) AS tx_count,
               SUM(tr_sum_btc) AS tx_volume
        FROM blockchain_bitcoin.transactions_data
        GROUP BY date
        ORDER BY date
    """, engine, parse_dates=['date'])

def load_funding(engine):
    return pd.read_sql("""
        SELECT (to_timestamp(timestamp) AT TIME ZONE 'UTC')::date AS date,
               AVG(mark_price) AS price,
               AVG(funding) AS funding_rate
        FROM binance.funding_data
        WHERE symbol = 'BTCUSDT'
        GROUP BY date
        ORDER BY date
    """, engine, parse_dates=['date'])

def add_onchain_features(df, engine, z_thresh=2.0):
    # викачуємо статистики по адресах

```

```

chunks = []
for chunk in pd.read_sql("""
    SELECT tr_date::date AS date,
           SUM(count_addresses) AS total_addr_refs,
           AVG(count_addresses) AS avg_addr_per_tx,
           COUNT(DISTINCT address_out) AS unique_receivers,
           SUM(CASE WHEN tr_sum_btc > 10 THEN 1 END) AS large_tx_count
    FROM blockchain_bitcoin.transactions_data
    GROUP BY date
    ORDER BY date
""", engine, parse_dates=['date'], chunksize=50_000):
    chunks.append(chunk)
stats = pd.concat(chunks).set_index('date')
df = df.join(stats, how='inner')

# інші ознаки
df['avg_tx_value'] = df['tx_volume'] / df['tx_count']
df['tx_count_ma7'] = df['tx_count'].rolling(7).mean()
df['addr_refs_ma7'] = df['total_addr_refs'].rolling(7).mean()
df['large_tx_ratio'] = df['large_tx_count'] / df['tx_count']
df['z_large_tx'] = zscore(df['large_tx_count'].fillna(0))
df['anomaly'] = (df['z_large_tx'].abs() > z_thresh).astype(int)
return df

# -----
# 2) Feature selection
# -----
def feature_selection(df):
    tmp = df.copy()
    tmp['price_change'] = tmp['price'].pct_change().shift(-1)
    fs = tmp.dropna(subset=['price_change'])
    candidates = [
        'tx_volume', 'tx_count', 'funding_rate', 'price',
        'total_addr_refs', 'avg_addr_per_tx', 'large_tx_count',
        'avg_tx_value', 'tx_count_ma7', 'addr_refs_ma7', 'large_tx_ratio'
    ]
    X = fs[candidates].dropna()
    y = fs.loc[X.index, 'price_change']

    print("\n=== Feature Selection ===")
    sel = SelectKBest(f_regression, k=5).fit(X, y)
    print(" Univariate: ", list(X.columns[sel.get_support()]))
    rfe = RFE(LinearRegression(), n_features_to_select=5).fit(X, y)
    print(" RFE:      ", list(X.columns[rfe.get_support()]))
    lsm = SelectFromModel(Lasso(alpha=0.001), threshold='mean').fit(X, y)
    print(" Lasso:     ", list(X.columns[lsm.get_support()]))
    rf = RandomForestRegressor(n_estimators=100, random_state=42).fit(X, y)
    imp = pd.Series(rf.feature_importances_, index=X.columns)
    print(" RandomForest: ", list(imp.nlargest(5).index))

# -----
# 3) Correlation & OLS
# -----
def run_correlations(df):
    print("\n=== Correlations ===")
    print(" Pearson price~tx_volume: ", pearsonr(df['price'],
df['tx_volume']))

```

```

        print("    Spearman    price~tx_volume:",    spearmanr(df['price'],
df['tx_volume']))
    X = sm.add_constant(df[['tx_volume','tx_count']].dropna())
    y = df.loc[X.index, 'price']
    print(sm.OLS(y, X).fit().summary())

# -----
# 4) VAR + Impulse Response
# -----
def run_var_irf(df, maxlags=7, periods=10):
    data = df[['price','tx_volume','tx_count']].dropna()
    m = VAR(data)
    res = m.fit(maxlags=maxlags, ic='aic')
    print("\n=== VAR Summary ===")
    print(res.summary())

    irf = res.irf(periods)
    irf.plot(orth=False)
    plt.suptitle('Impulse Response (shocks ~ price)')
    plt.tight_layout()
    plt.show()

# -----
# 5) Linear interpretation  $\Delta$ Price ~ features
# -----
def interpret_linear(df, feats):
    d = df.copy()
    d[' $\Delta$ price'] = d['price'].shift(-1) - d['price']
    d = d.dropna(subset=[' $\Delta$ price'] + feats)
    Xc = d[feats]
    yc = d[' $\Delta$ price']
    lin = LinearRegression().fit(Xc, yc)
    coefs = pd.Series(lin.coef_, index=feats).sort_values()
    plt.figure(figsize=(6,6))
    coefs.plot(kind='barh')
    plt.title('Linear model: coefficients for  $\Delta$ Price')
    plt.xlabel('Price (USD) per unit feature')
    plt.tight_layout()
    plt.show()

    print(f" Intercept: {lin.intercept_:.2f}")

# -----
# 6) Anomaly analysis & plot
# -----
def analyze_anomalies(df):
    d = df.copy()
    d[' $\Delta$ price'] = d['price'].diff()
    d[' $\Delta$ pct'] = d['price'].pct_change()*100
    anoms = d[d['anomaly']==1]
    impact_usd = anoms[' $\Delta$ price'].mean()
    impact_pct = anoms[' $\Delta$ pct'].mean()

    print("\n=== Anomaly Impact ===")
    print(f" Mean  $\Delta$ price on anomaly days: {impact_usd:.2f} USD")
    print(f" Mean % $\Delta$ price on anomaly days: {impact_pct:.2f}%")

    plt.figure(figsize=(10,4))
    plt.plot(d.index, d[' $\Delta$ price'], label='Price')

```

```

plt.scatter(anoms.index, anoms['Δprice'],
            color='red', s=50, label='Anomaly')
plt.title('ΔPrice with detected anomalies')
plt.ylabel('ΔPrice (USD)')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()

# -----
# 7) Train LightGBM & hold-out + Actual vs Predicted price
# -----
def train_lightgbm(df, feats, test_size=0.2):
    d = df.copy()
    d['Δprice'] = d['price'].shift(-1) - d['price']
    # видаляємо NaN у цільовій і в ознаках
    d = d.dropna(subset=['Δprice'] + feats)

    X = d[feats]
    y = d['Δprice']
    split = int(len(d)*(1-test_size))
    X_train, X_test = X.iloc[:split], X.iloc[split:]
    y_train, y_test = y.iloc[:split], y.iloc[split:]

    model = LGBMRegressor(
        random_state=42,
        n_jobs=-1,
        max_depth=6,
        n_estimators=500,
        learning_rate=0.1,
        num_leaves=31,
        lambda_l2=0.5,
        feature_fraction=0.8,
        bagging_fraction=0.8,
        bagging_freq=1
    )
    # просто вчимо на всіх ітераціях, без early stopping
    model.fit(X_train, y_train)

    y_pred = model.predict(X_test)
    rmse = np.sqrt(mean_squared_error(y_test, y_pred))
    print(f"\n=== LightGBM hold-out ===")
    print(f" RMSE on Δprice: {rmse:.2f} USD")

    # Plot: Actual vs Predicted ΔPrice
    plt.figure(figsize=(10,4))
    plt.plot(y_test.index, y_test, label='Actual ΔPrice', linewidth=2)
    plt.plot(y_test.index, y_pred, '-', label='Predicted ΔPrice', linewidth=2)
    plt.title('LightGBM: Actual vs Predicted ΔPrice')
    plt.ylabel('ΔPrice (USD)')
    plt.legend()
    plt.grid(True)
    plt.tight_layout()
    plt.show()

    return model, X_test

# -----
# 8) SHAP analysis

```

```

# -----
def run_shap(model, X_test):
    explainer = shap.Explainer(model)
    sv = explainer(X_test)

    # SHAP bar plot
    shap.summary_plot(sv, X_test, plot_type='bar', show=False)
    plt.title("SHAP feature importance")
    plt.tight_layout()
    plt.show()

    # SHAP beeswarm
    shap.summary_plot(sv, X_test, show=False)
    plt.title("SHAP beeswarm")
    plt.tight_layout()
    plt.show()

# -----
# 9) Main pipeline
# -----
def main():
    eng_tx = create_engine(TX_DB_URL).execution_options(stream_results=True)
    eng_fund = create_engine(FUND_DB_URL)

    tx = load_transactions(eng_tx)
    fund = load_funding(eng_fund)
    df = tx.set_index('date').join(fund.set_index('date'), how='inner').dropna()

    # on-chain features + anomaly flag (z_thresh=2.0)
    df = add_onchain_features(df, eng_tx, z_thresh=2.0)

    # 2) Feature selection
    feature_selection(df)

    # 3) Correlations & OLS
    run_correlations(df)

    # 4) VAR + IRF
    run_var_irf(df)

    # 5) Linear coeffs for  $\Delta$ Price
    feats = [
        'tx_volume', 'tx_count', 'funding_rate', 'price',
        'total_addr_refs', 'avg_addr_per_tx', 'large_tx_count',
        'avg_tx_value', 'tx_count_ma7', 'addr_refs_ma7',
        'large_tx_ratio', 'anomaly'
    ]
    interpret_linear(df, feats)

    # 6) Anomaly impact & plot
    analyze_anomalies(df)

    # 7) Train LightGBM & plot Actual vs Predicted  $\Delta$ Price
    lgb_model, X_test = train_lightgbm(df, feats)

    # 8) SHAP analysis (importance + beeswarm)
    run_shap(lgb_model, X_test)

if __name__ == '__main__':

```

main()

ДОДАТОК Б: ПРЕЗЕНТАЦІЙНІ МАТЕРІАЛИ

Підготував: Кривоус Максим
Олександрович
Науковий керівник: Канцедал
Георгій Олегович

Аналіз впливу блокчейн транзакцій на курс криптовалюти

Актуальність проблеми

Економічна значущість. Bitcoin перетворився з експериментального проекту у масштабний фінансовий інструмент із капіталізацією понад сотні мільярдів доларів США та високою волатильністю. Розуміння факторів ціноутворення має як наукове, так і практичне значення для учасників ринку.

Унікальність даних. На відміну від традиційних фінансових активів, Bitcoin забезпечує повну прозорість блокчейну - всі транзакції публічно доступні для аналізу. Це створює унікальну можливість досліджувати взаємозв'язок між мережевою активністю та ринковою ціною в режимі реального часу.

Практична потреба. Інвестори та аналітики постійно шукають фундаментальні показники для оцінки "справедливої" вартості криптовалют та прогнозування цінових коливань. Ончейн-метрики можуть слугувати такими індикаторами, доповнюючи традиційний технічний аналіз.

Вплив великих гравців. Можливість відстежувати великі транзакції ("whale"-активність) дозволяє виявляти потенційні ринкові рухи та зміни в поведінці крупних власників, що безпосередньо впливає на ринкові очікування та короткострокову волатильність.

Слайд 2

Предмет, об'єкт та мета роботи

Предмет дослідження: методи аналізу часових рядів та прогностичні моделі машинного навчання для оцінки впливу новинного настрою на ринок криптовалют.

Об'єкт дослідження: процеси формування ціни Біткоїна під впливом блокчейн транзакцій.

Мета роботи: створити та порівняти моделі машинного навчання для визначення залежності між настроєм криптовалютних новин та змінами курсу Біткоїна.

Слайд 1

Постановка задачі

наявні ончейн-дані блокчейну Bitcoin та історичні ціни за період 272 дні (червень 2023 - березень 2024 років);

дослідження впливу блокчейн-транзакцій на курс Bitcoin через аналіз ключових ончейн-метрик;

використовуються статистичні методи (кореляційний аналіз, VAR) та алгоритми машинного навчання (LightGBM, CatBoost, MLP);

результати моделей порівнюються за метриками RMSE, MAE та R^2 для визначення найточнішого підходу;

за допомогою функціонально-вартісного аналізу підтверджується економічна доцільність розробленого рішення.

Слайд 2

Про розроблений програмний продукт

Програмний продукт реалізований на Python з використанням бібліотек pandas, numpy, lightgbm, sklearn, shap у середовищі Jupyter Notebook.

Складається з двох частин: збору та обробки ончейн-даних та машинного навчання.

Перша частина збирає дані з PostgreSQL баз блокчейну Bitcoin та біржових даних, формує ончейн-метрики (кількість транзакцій, активні адреси, обсяги переказів, великі транзакції) та набори ознак для навчання. Друга частина тренує статистичні моделі (VAR, кореляційний аналіз) та алгоритми машинного навчання (LightGBM, CatBoost, MLP), оцінює їх ефективність за метриками RMSE, MAE та R^2 , проводить SHAP-аналіз важливості ознак і візуалізує результати порівняння.

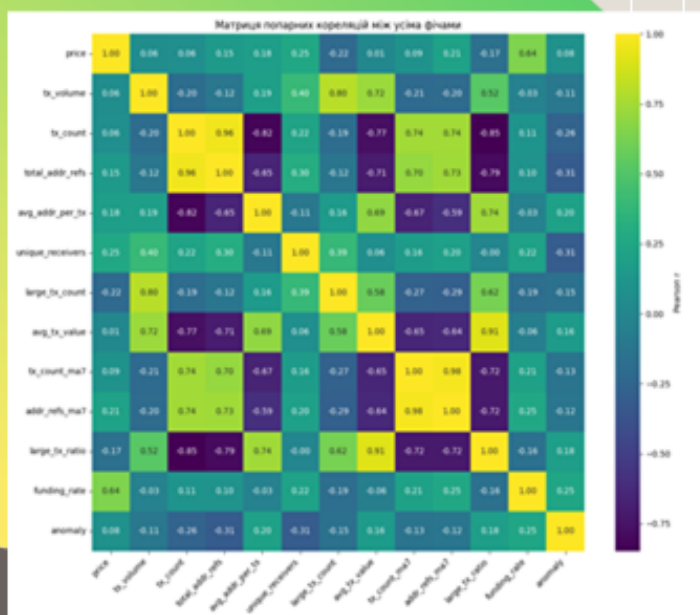
Слайд 4



Матриця парних коефіцієнтів кореляції Пірсона

- total_addr_refs vs Price: позитивна кореляція (≈ 0.15)
- tx_volume vs Price: дуже слабка (≈ 0.06)
- large_tx_ratio vs Price: мінусова (≈ -0.17)
- funding_rate vs Price: середня позитивна (≈ 0.64)
- anomaly, avg_tx_value, avg_addr_per_tx: практично нульова

Слайд 5

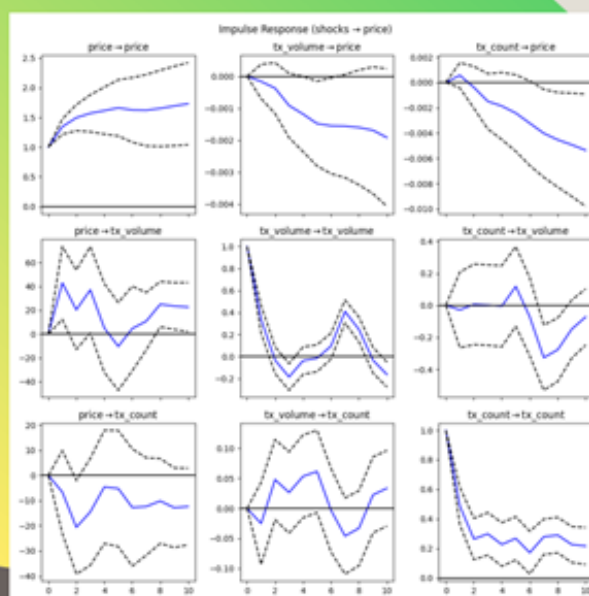


Результат роботи VAR-моделі

Змінна	Коефіцієнт	Std. Error	t-stat	p-value
const	804.534135	491.560303	1.637	0.102
L1.price	1.344458	0.065943	20.388	0.000
L1.tx_volume	-0.000160	0.000273	-0.584	0.559
L1.tx_count	0.000564	0.000509	1.108	0.268
L2.price	-0.301244	0.108723	-2.771	0.006
L2.tx_volume	-0.000085	0.000292	-0.291	0.771
L2.tx_count	-0.001423	0.000568	-2.504	0.012

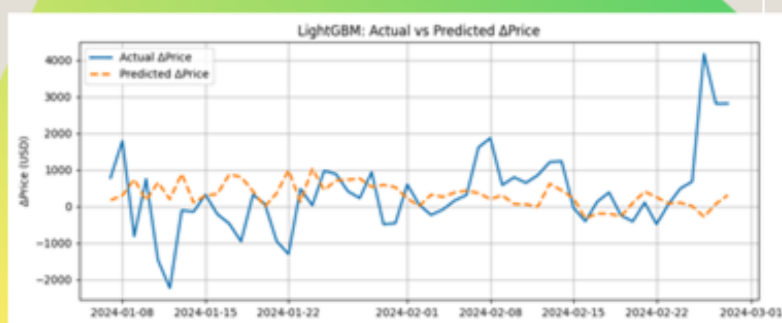
Слайд 6

Аналіз імпульсних відгуків



Слайд 7

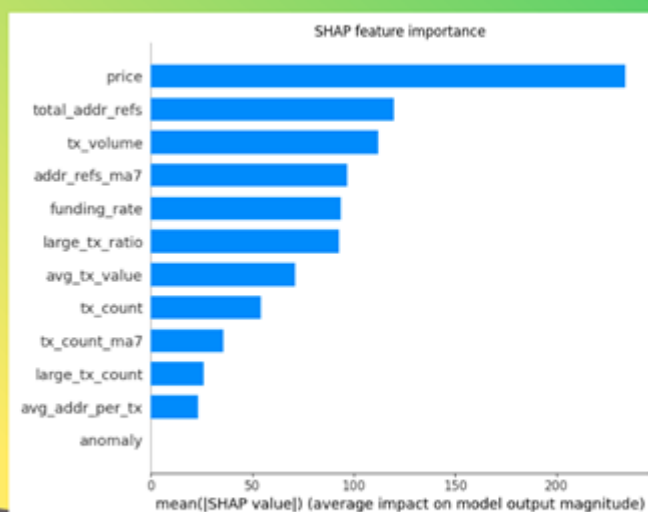
Порівняння фактичної та прогнозованої добових змін ціни BTC



Синя лінія – справжнє
коливання ціни
Помаранчева – прогнозоване

Слайд 8

Аналіз SHAP-графіку



Слайд 9

Висновки

Було виконано дослідження та експеримент з побудови моделей для аналізу впливу блокчейн-транзакцій на курс Bitcoin. Для криптовалютного ринку подібна проблема прогнозування є актуальною та складною. Виконано аналіз статистичними методами та алгоритмами машинного навчання за допомогою розробленого програмного продукту. LightGBM модель показала найкращі результати серед усіх архітектур з $RMSE \approx 4700$ USD. Порівняльний аналіз виявив обмежені можливості прогнозування екстремальних цінових стрибків виключно на основі ончейн-метрик.

Перспективи розвитку:

- створення системи підтримки інвестиційних рішень для криптовалютного ринку на базі ончейн-аналітики як конкурентоздатного продукту;
- додавання інших джерел даних (макроекономічні індикатори, соціальні мережі, технічні індикатори) та розширення на інші криптовалюти для побудови мультиактивних прогнозних моделей;
- подальші дослідження для збільшення точності прогнозування екстремальних рухів ціни та кращого розуміння впливу ончейн-факторів на ціноутворення криптовалют.

Слайд 11

Дякую за увагу!

Слайд 12