

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет водного господарства та
природокористування
Навчально-науковий інститут кібернетики, інформаційних технологій
та інженерії
Кафедра комп'ютерних технологій та економічної кібернетики

Допущено до захисту:
Завідувач кафедри
комп'ютерних технологій та
економічної кібернетики
д. е. н., проф. П. М. Грицюк

« _____ » _____ 20__ р.

КВАЛІФІКАЦІЙНА РОБОТА
на здобуття ступеня «магістр»
за освітньо-професійною програмою
«Інформаційні технології в бізнесі»
спеціальності 126 «Інформаційні системи та технології»
на тему: «**Моделювання волатильності криптовалютного ринку з
використанням машинного навчання**»

Виконав:
здобувач вищої освіти 2 курсу,
групи ІТБ-61м
Янчук Вадим Олександрович
Керівник:
к.т.н., доц. Барановський С. В.
Рецензент:
к.т.н., доц. Гладка О.М.

Рівне – 2025

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Національний університет водного господарства та природокористування
 ННІ кібернетики, інформаційних технологій та інженерії
 Кафедра комп'ютерних технологій та економічної кібернетики

Освітньо-кваліфікаційний рівень – магістр
 Освітньо-професійна програма «Інформаційні технології в бізнесі»
 Спеціальність 126 «Інформаційні системи і технології»

ЗАТВЕРДЖУЮ

Завідувач кафедри комп'ютерних технологій
 та економічної кібернетики

_____ д.е.н., проф. П.М. Грицюк

« _____ » _____ 20__ р.

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачу _____ *Янчук Вадиму Олександровичу*
 (прізвище, ім'я, по-батькові)

1. Тема роботи _____ *Моделювання волатильності криптовалютного ринку*
 _____ *з використанням машинного навчання*

керівник роботи _____ *канд. техн. наук, доцент Барановський С.В.*
 (прізвище, ім'я, по-батькові, науковий ступінь, вчене звання)

затверджена наказом по університету від “ _____ “ _____ 202__ року
 № _____

2. Термін здачі студентом закінченої роботи _____

3. Вихідні дані до роботи _____ *Важливою проблемою на сьогодні залишається висока*
волатильність криптовалютного ринку та складність прогнозування динаміки цін цифрових
активів. Завданням кваліфікаційної роботи є дослідження статистичних характеристик
криптовалют, аналіз часових рядів та розробки моделей прогнозування волатильності й цін
на основі методів машинного навчання. Вихідними даними для виконання роботи є історичні
фінансові дані криптовалют, а також програмні засоби аналізу та обробки даних.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно
 розробити)

Вступ

1. Аналіз сучасного стану та тенденцій розвитку криптовалютного ринку.

2. Дослідження волатильності криптовалют на основі методів описової статистики.

3. Аналіз часових рядів та виявлення трендів і сезонних компонент.

4. Побудова моделей прогнозування з використанням ARIMA та SARIMA.

5. Застосування ланцюгів Маркова для класифікації ринкових станів.

6. Використання моделей глибинного навчання для прогнозування цін криптовалют.

7. Кореляційний аналіз між різними криптовалютами та порівняння результатів моделей.

Висновки

5. Перелік графічного матеріалу _____

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання отримав
1	<i>Барановський С.В., доцент</i>		
2	<i>Барановський С.В., доцент</i>		
3	<i>Барановський С.В., доцент</i>		

7. Дата видачі завдання _____

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів магістерської роботи	Терміни виконання етапів роботи	Примітка
1.	<i>Загальна характеристика криптовалютного ринку та аналіз сучасних тенденцій його розвитку</i>	<i>до 20.09.2025</i>	
2.	<i>Аналіз наукових підходів і методів дослідження волатильності фінансових часових рядів</i>	<i>до 01.10.2025</i>	
3.	<i>Збір, підготовка та попередня обробка історичних даних криптовалют</i>	<i>до 10.10.2025</i>	
4.	<i>Статистичний аналіз волатильності криптовалют і дослідження часових рядів</i>	<i>до 20.10.2025</i>	
5.	<i>Побудова моделей прогнозування волатильності та цін на основі ARIMA, ланцюгів Маркова та методів глибокого навчання</i>	<i>до 01.11.2025</i>	
6.	<i>Проведення ризик-аналізу криптовалютних інвестицій та порівняльний аналіз отриманих результатів</i>	<i>до 30.11.2025</i>	
7.	<i>Оформлення та представлення до захисту магістерської роботи</i>	<i>до 15.12.2025</i>	

Студент(ка) _____

(підпис)

(прізвище та ініціали)

Керівник роботи _____

(підпис)

(прізвище та ініціали)

РЕФЕРАТ

Кваліфікаційна робота магістра: **81 с., 17 рис., 1 табл., 13 літературних джерел.**

Актуальність теми даної магістерської роботи полягає у зростаючій ролі криптовалют на глобальних фінансових ринках та потребі у точному прогнозуванні їх волатильності. Тема є актуальною з огляду на інтенсивний розвиток цифрових економік та швидке поширення криптовалютних транзакцій у всьому світі.

Об'єкт дослідження магістерської роботи – процес моделювання, прогнозування та аналізу волатильності криптовалют за допомогою методів машинного навчання.

Предметною областю дослідження є статистичні, економетричні та інтелектуальні методи аналізу фінансових даних, зокрема ARIMA/SARIMA-моделі, ланцюги Маркова, а також моделі глибинного навчання LSTM та GRU для прогнозування цін криптовалют.

Метою магістерської роботи є формування комплексної моделі прогнозування волатильності криптовалют на основі сучасних алгоритмів машинного навчання та статистичного аналізу, що дозволяє підвищити точність оцінки ризиків та забезпечити ефективність прогнозування динаміки криптовалютних цін.

У магістерській роботі виконано комплексний аналіз криптовалютного ринку, що охоплює статистичне дослідження часових рядів, оцінку волатильності та застосування моделей ARIMA/SARIMA, ланцюгів Маркова та глибинного навчання LSTM і GRU для прогнозування цін. Отримані результати дозволили сформуванню ефективних моделей прогнозування та виявити ключові чинники ризику.

КЛЮЧОВІ СЛОВА: криптовалюта, волатильність, машинне навчання, ARIMA, SARIMA, LSTM, GRU, часові ряди, Value at Risk, log-returns, ризик-аналіз, прогнозування.

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

BTC — Bitcoin, перша та найбільша за капіталізацією криптовалюта.

ETH — Ethereum, криптовалюта другого покоління з підтримкою смарт-контрактів.

ADA — Cardano, блокчейн-платформа третього покоління.

XLM — Stellar, платформа для швидких транскордонних платежів.

ML — Machine Learning, машинне навчання.

AI — Artificial Intelligence, штучний інтелект.

ARIMA — AutoRegressive Integrated Moving Average, авторегресійна інтегрована модель ковзної середньої.

SARIMA — Seasonal ARIMA, сезонна модифікація моделі ARIMA.

LSTM — Long Short-Term Memory, рекурентна нейронна мережа з довго- та короткостроковою пам'яттю.

GRU — Gated Recurrent Unit, рекурентна нейронна мережа з модифікованими блоками пам'яті.

VaR — Value at Risk, показник ризику втрати активів.

MSE — Mean Squared Error, середньоквадратична похибка.

MAE — Mean Absolute Error, середня абсолютна похибка.

SMA — Simple Moving Average, проста ковзна середня.

EMA — Exponential Moving Average, експоненційна ковзна середня.

CV — Coefficient of Variation, коефіцієнт варіації.

TS — Time Series, часові ряди.

ЗМІСТ

ВСТУП.....	5
РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ АНАЛІЗУ ВОЛАТИЛЬНОСТІ ФІНАНСОВИХ РИНКІВ	8
1.1. Поняття волатильності та її значення у фінансовій сфері	8
1.2. Економетричні моделі волатильності.....	12
1.3. Моделювання волатильності засобами машинного навчання.....	13
РОЗДІЛ 2. ОГЛЯД ІНСТРУМЕНТІВ І МЕТОДІВ ДОСЛІДЖЕННЯ.....	16
2.1. Джерела даних для аналізу криптовалют	16
2.2. Попередня обробка даних.....	16
2.3. Вибір метрик для оцінки моделей	18
РОЗДІЛ 3. ДОСЛІДЖЕННЯ, МОДЕЛЮВАННЯ ТА ОЦІНКА ВОЛАТИЛЬНОСТІ КРИПТОВАЛЮТНОГО РИНКУ	21
3.1. Базова статистика	21
3.2. Часові ряди	22
3.3. ARIMA / SARIMA для прогнозування часових рядів	28
3.4. Ланцюги Маркова для класифікації днів (зростання / падіння) ...	34
3.5. Моделі глибинного навчання (LSTM / GRU для прогнозу цін)	38
3.6. Дохідність (log returns)	43
3.7. Value at Risk (VaR) з історичних даних	47
3.8. Кореляційний аналіз між різними криптовалютами (BTC, ADA, XLM)	50
ВИСНОВКИ	55
СПИСОК ВИКОРОСТАНИХ ДЖЕРЕЛ	59
ДОДАТКИ.....	61

ВСТУП

Сучасний фінансовий ринок характеризується високою динамічністю, зростанням обсягів електронних транзакцій та появою нових інвестиційних інструментів. Одним із найпомітніших явищ останнього десятиліття стало стрімке зростання популярності криптовалют — децентралізованих цифрових активів, що функціонують на основі блокчейн-технологій. Криптовалюти, зокрема Bitcoin, Ethereum, Ripple та інші, перетворилися з експериментальної технології на повноцінний фінансовий сегмент із багатомільярдним обігом. Проте їхня основна особливість — висока волатильність, тобто суттєві коливання цін у короткі проміжки часу, що створює як можливості для прибутку, так і значні ризики для інвесторів.

Проблема оцінки та прогнозування волатильності криптовалютного ринку є надзвичайно актуальною. Традиційні статистичні методи часто виявляються недостатніми для опису нелінійної, нестабільної поведінки ринку, тому дослідники та аналітики все частіше звертаються до методів машинного навчання. Вони дозволяють виявляти приховані закономірності у великих обсягах даних, адаптуватися до змін ринкових умов і підвищувати точність прогнозів. Застосування таких підходів є перспективним напрямом розвитку фінансової аналітики та управління ризиками.

Мета дослідження — розробка та реалізація підходу до моделювання волатильності криптовалютного ринку із застосуванням методів машинного навчання для покращення прогнозування цінової динаміки та оцінки ризиків інвестицій.

Для досягнення мети поставлено такі завдання:

1. Проаналізувати теоретичні аспекти поняття волатильності та її вплив на поведінку криптовалютного ринку.
2. Дослідити існуючі методи моделювання волатильності, включаючи класичні статистичні та сучасні алгоритмічні підходи.

3. Зібрати та підготувати реальні дані щодо цінових коливань основних криптовалют.
4. Провести базовий статистичний аналіз даних (середні значення, стандартне відхилення, коефіцієнт варіації тощо).
5. Побудувати часові ряди та візуалізувати динаміку змін ринку.
6. Реалізувати моделі прогнозування (ARIMA, Ланцюги Маркова, LSTM) та порівняти їх ефективність.
7. Провести ризик-аналіз із використанням показників Value at Risk (VaR), Sharpe Ratio та оцінити взаємозв'язок між криптовалютами.
8. Надати рекомендації щодо використання результатів у системах підтримки прийняття фінансових рішень.

Об'єкт дослідження — процес формування та зміни волатильності на криптовалютному ринку.

Предмет дослідження — методи статистичного та машинного аналізу часових рядів криптовалют для оцінки та прогнозування їх волатильності.

У роботі використовуються методи математичної статистики, аналізу часових рядів, економетрики, машинного навчання, а також методи візуалізації даних із застосуванням бібліотек Python. Для оцінки результатів моделювання застосовуються метрики точності прогнозів та показники ризику.

Наукова новизна роботи полягає у комплексному підході до моделювання волатильності криптовалютного ринку шляхом поєднання класичних статистичних моделей із сучасними методами машинного навчання. Це дозволяє не лише покращити точність прогнозів, а й сформувані інструментарій для оцінки ризику на основі історичних і поточних даних.

Практичне значення одержаних результатів полягає у можливості використання розробленої методики для створення аналітичних систем моніторингу та прогнозування стану криптовалютного ринку, що може бути корисним для трейдерів, інвесторів, фінансових аналітиків та дослідників у сфері фінансових технологій. Результати роботи можуть бути застосовані для

підвищення ефективності інвестиційних стратегій, мінімізації ризиків і прийняття обґрунтованих рішень на ринку цифрових активів.

РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ АНАЛІЗУ ВОЛАТИЛЬНОСТІ ФІНАНСОВИХ РИНКІВ

1.1. Поняття волатильності та її значення у фінансовій сфері

Волатильність є однією з основних характеристик фінансових ринків, що відображає динаміку та мінливість цін активів у часі. У найзагальнішому розумінні під волатильністю розуміють ступінь варіації ціни активу навколо його середнього значення за певний період. Вона виступає ключовим індикатором ризику, оскільки показує, наскільки сильно може змінюватися вартість фінансового інструменту під впливом ринкових факторів.

Історично термін «волатильність» (від англ. *volatility*) запозичений із фізики, де він описував властивість речовини легко переходити у газоподібний стан, тобто бути «нестабільною». У фінансах це поняття набуло значення нестабільності або непередбачуваності цін. Якщо актив демонструє часті та значні коливання, він вважається високо волатильним; якщо зміни незначні — низьковолатильним.

Згідно з класичною фінансовою теорією, волатильність є невід’ємним елементом поняття ризику. Вона використовується у таких ключових моделях, як модель оцінки капітальних активів (САРМ), де ризик окремого активу вимірюється через стандартне відхилення його доходності, а також у моделі Блека–Шоулза для оцінювання вартості опціонів, де волатильність визначає премію за невизначеність.

З математичної точки зору волатильність найчастіше вимірюється через стандартне відхилення або дисперсію доходностей активу:

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (r_i - \bar{r})^2}$$

де r_i — доходність активу у період i , \bar{r} — середня доходність, n — кількість спостережень.

Цей показник дозволяє кількісно оцінити розкид фактичних доходностей навколо середнього значення. Наприклад, якщо середня денна доходність становить 1 %, а стандартне відхилення — 4 %, це означає, що типове відхилення прибутку вгору або вниз може досягати ± 4 %.

У практичній діяльності аналітики та трейдери використовують декілька типів волатильності:

1. Історична волатильність (Historical Volatility) — розраховується на основі фактичних даних про ціну активу за певний період (наприклад, 30 чи 90 днів). Вона показує, наскільки сильно ціна змінювалася в минулому.
2. Очікувана волатильність (Implied Volatility) — виводиться з цін опціонів і відображає ринкові очікування майбутніх коливань. Вона є показником поточного рівня невизначеності інвесторів.
3. Реалізована волатильність (Realized Volatility) — фактична волатильність, обчислена постфактум для певного періоду, і часто використовується для перевірки точності прогнозів.

Волатильність має подвійний характер. З одного боку, вона свідчить про ризик, тобто можливі втрати; з іншого — створює потенційні можливості для отримання прибутку. На ринках з низькою волатильністю ціни змінюються повільно, тому потенційна прибутковість є обмеженою. Високоволатильні активи, навпаки, приносять більший ризик, але і більший можливий дохід.

У контексті сучасної економіки волатильність відіграє надзвичайно важливу роль для управління ризиками, формування інвестиційних стратегій та оцінки ефективності ринку. У періоди макроекономічної нестабільності або криз спостерігається підвищення волатильності, що часто використовується як індикатор панічних настроїв інвесторів. Саме тому волатильність вважається «мірою страху» на фінансових ринках — її навіть вимірюють спеціальним індексом VIX (Volatility Index), який відображає очікувану волатильність S&P 500 і використовується як барометр ринкових ризиків.

Особливої уваги волатильність набуває у сфері криптовалютного ринку, який характеризується надзвичайною динамічністю. На відміну від традиційних фінансових інструментів, ціни криптовалют формуються переважно на основі попиту і пропозиції без регуляторного втручання, що робить їх вразливими до емоційних і спекулятивних факторів. Наприклад, заяви впливових осіб або зміни у законодавстві можуть призвести до коливань у десятки відсотків протягом декількох годин.

Висока волатильність криптовалют має як позитивні, так і негативні наслідки. З одного боку, вона створює великі можливості для спекуляцій та короткострокових прибутків. З іншого — підвищує інвестиційні ризики, ускладнює прогнозування і перешкоджає широкому використанню криптовалют як стабільного засобу збереження вартості.

У зв'язку з цим точне вимірювання, моделювання та прогнозування волатильності є критично важливими завданнями. Вони дозволяють інвесторам приймати обґрунтовані рішення, трейдерам — оцінювати рівень ризику, а науковцям — розробляти ефективні алгоритми управління фінансовими активами.

Базовим етапом у дослідженні волатильності є статистичний аналіз історичних цінових даних. Найчастіше для цього використовуються такі показники, як середнє арифметичне, стандартне відхилення, коефіцієнт варіації, медіана, максимум і мінімум за обраний часовий період. Ці показники дають змогу оцінити загальний характер розподілу даних і рівень їхньої мінливості.

Середня ціна (mean) характеризує типове значення показника (наприклад, середню ціну закриття за певний період) і обчислюється за формулою:

$$\bar{P} = \frac{1}{N} \sum_{i=1}^N P_i$$

де P_i — ціна активу в момент часу i , N — кількість спостережень.

Стандартне відхилення (σ) є мірою середньої амплітуди відхилень значень від середнього рівня та виражає ступінь розкиду даних. Чим вищим є стандартне відхилення, тим більша волатильність активу. Формула розрахунку має вигляд:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (P_i - \bar{P})^2}$$

Для порівняння активів із різним рівнем цін використовують коефіцієнт варіації (CV), який нормує стандартне відхилення на середнє значення:

$$CV = \frac{\sigma}{\bar{P}} \times 100\%.$$

Цей показник дозволяє виявити, який із активів має більшу відносну мінливість і, відповідно, вищий ризик.

Крім того, у практиці фінансового аналізу широко застосовуються ковзні показники, які відображають локальні зміни волатильності у часі. Наприклад, ковзне стандартне відхилення (rolling std) обчислюється на основі рухомого вікна, що дозволяє аналізувати динаміку ризику на різних часових горизонтах. Аналогічно, ковзне середнє (Moving Average, MA) або експоненційне згладжування (EMA) допомагають відфільтрувати короткострокові коливання, фокусуючись на довгострокових тенденціях ринку.

Попри свою простоту, ці методи мають обмеження: вони не враховують автокореляцію між спостереженнями, не дозволяють оцінити умовну волатильність (тобто зміну ризику в часі) і не забезпечують достатньої точності для прогнозування. Саме тому для глибшого аналізу застосовуються економетричні моделі.

1.2. Економетричні моделі волатильності

Економетричні моделі розвинулись як відповідь на потребу у гнучкому та динамічному підході до моделювання волатильності. Найпоширенішими серед них є моделі ARIMA, GARCH, EGARCH, TGARCH, EWMA тощо.

Модель ARIMA є класичною для аналізу часових рядів і дозволяє враховувати як тренди, так і сезонність у динаміці цін. Її основна ідея полягає в тому, що поточне значення ряду залежить від попередніх спостережень та похибок. Модель позначається як ARIMA(p, d, q), де:

- p — порядок авторегресії (AR),
- d — порядок диференціювання (для усунення тренду),
- q — порядок ковзного середнього (MA).

Хоч ARIMA добре працює для прогнозування цін, вона не моделює волатильність безпосередньо. Для цього застосовуються моделі типу GARCH.

GARCH-моделі враховують змінність волатильності в часі, тобто властивість "кластеризації волатильності": періоди високої мінливості мають тенденцію групуватися.

Базова модель GARCH(1,1) описується системою рівнянь:

$$r_t = \mu + \epsilon_t, \quad \epsilon_t = \sigma_t z_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

де σ_t^2 — умовна дисперсія (волатильність) у момент часу t , z_t — випадкова складова.

GARCH дозволяє моделювати динамічну волатильність і прогнозувати її майбутнє значення. Її модифікації, такі як EGARCH (Exponential GARCH) або TGARCH (Threshold GARCH), враховують асиметрію — різну реакцію волатильності на позитивні й негативні зміни ціни. Це особливо актуально для криптовалют, які характеризуються частими різкими падіннями та сплесками ціни.

Модель EWMA є спрощеним варіантом GARCH, де поточна волатильність розраховується як експоненційно зважене середнє минулих

квадратів доходностей. Це забезпечує більшу реакцію моделі на останні зміни на ринку, що робить її корисною для короткострокового аналізу.

1.3. Моделювання волатильності засобами машинного навчання

З розвитком технологій аналізу даних та обчислювальних потужностей з'явилися нові можливості для моделювання волатильності на основі методів машинного навчання (ML). Ці методи дозволяють виявляти складні нелінійні закономірності в даних, які не піддаються класичному статистичному аналізу.

До таких підходів належать:

- Регресійні моделі (наприклад, лінійна або логістична регресія);
- Дерева рішень і ансамблеві методи (Random Forest, Gradient Boosting);
- Нейронні мережі, зокрема LSTM (Long Short-Term Memory) для прогнозування часових рядів;
- Марковські ланцюги для моделювання ймовірностей переходу між станами (“зростання”, “спад”, “стабільність”).

Особливо перспективними є рекурентні нейронні мережі (RNN) та LSTM, які здатні враховувати часову залежність даних і запам'ятовувати контекст попередніх періодів. У випадку криптовалют, де волатильність має нелінійний та хаотичний характер, такі моделі показують високу точність прогнозування у порівнянні з класичними підходами.

Методи машинного навчання дозволяють будувати гібридні моделі, які поєднують переваги економетричних і нейронних підходів. Наприклад, модель ARIMA може використовуватися для усунення тренду, а LSTM — для прогнозування залишків (residuals), що підвищує якість прогнозу.

Дана таблиця демонструє порівняльну характеристику основних методів моделювання волатильності фінансових активів, включно з традиційними статистичними, економетричними та алгоритмічними методами машинного навчання. Вона дозволяє побачити ключові переваги та обмеження кожного підходу, а також визначити сферу їх практичного застосування.

Порівняльна характеристика методів моделювання волатильності

Метод	Тип	Переваги	Недоліки	Сфера застосування
Стандартне відхилення	Статистичний	Простота, швидкість розрахунку	Не враховує часову залежність	Базовий аналіз
ARIMA	Економетричний	Добре прогнозує тренди	Не моделює волатильність напряду	Прогноз цін
GARCH	Економетричний	Моделює умовну волатильність	Погано працює з нелінійностями	Аналіз ризику
EWMA	Економетричний	Реагує на нові зміни	Менш гнучка, ніж GARCH	Коротко-строковий прогноз
LSTM	Машинне навчання	Висока точність, враховує часовий контекст	Потребує великих обсягів даних	Прогнозування цін і волатильності

Як видно з таблиці, статистичні методи є простими і швидкими у використанні, проте не враховують часову залежність даних і не здатні прогнозувати майбутні коливання волатильності. Економетричні моделі (ARIMA, GARCH, EWMA) дозволяють оцінювати умовну волатильність і враховувати автокореляції, але обмежені в роботі з нелінійними та хаотичними ринковими процесами. Методи машинного навчання, зокрема LSTM, відзначаються високою точністю прогнозування та здатністю враховувати складні часові залежності, проте потребують великих обсягів даних і значних обчислювальних ресурсів.

Порівняння методів показує, що для комплексного аналізу волатильності криптовалютного ринку оптимально використовувати гібридний підхід, який поєднує переваги кількох методів: базові статистичні оцінки для початкового аналізу, економетричні моделі для прогнозування

умовної волатильності та алгоритми машинного навчання для високоточних прогнозів у складних ринкових умовах.

РОЗДІЛ 2. ОГЛЯД ІНСТРУМЕНТІВ І МЕТОДІВ ДОСЛІДЖЕННЯ

2.1. Джерела даних для аналізу криптовалют

Для проведення дослідження волатильності криптовалютного ринку першочергово необхідно мати якісні та структуровані дані, які забезпечують можливість точного аналізу ринкових коливань та прогнозування цінових тенденцій. У даній роботі для цього використовується датасет у форматі CSV, що містить історичні щоденні дані по ціні Bitcoin. Датасет включає ключові параметри: ціни відкриття (Open), закриття (Close), максимальні (High) та мінімальні (Low) значення за день, ринкову капіталізацію (Marketcap) та обсяг торгів (Volume). Такі дані дозволяють не лише оцінювати абсолютні рівні цін, а й досліджувати їхні коливання, тренди та динаміку волатильності у часі.

Цей датасет був отриманий з платформи Kaggle, яка є однією з найбільших онлайн-баз даних для аналітики та досліджень у сфері фінансів і криптовалют. Kaggle надає доступ до відкритих наборів даних, що містять історичні часові ряди по різних криптовалютах, включно з Bitcoin, Ethereum, Ripple та іншими активами. Дані з цієї платформи мають високу точність і стандартизований формат, що дозволяє зручно інтегрувати їх у середовище для обробки та аналізу (Google Colab, R, Excel тощо).

Датасет у форматі CSV з Kaggle є основним джерелом інформації для проведення всіх подальших обчислень та досліджень у цій роботі, забезпечуючи надійну та репрезентативну базу для аналізу волатильності ринку криптовалют. Наявність історичних даних дозволяє не лише оцінити минулі коливання цін, а й побудувати моделі прогнозування, що мають практичне застосування для інвесторів та аналітиків фінансового ринку.

2.2. Попередня обробка даних

Попередня обробка даних є критично важливим етапом підготовки інформації для подальшого аналізу волатильності криптовалютного ринку. Оскільки кінцеві результати статистичних розрахунків, побудови часових

моделей і прогнозування значною мірою залежать від якості вихідного датасету, усі операції з очищення та перетворення даних повинні бути виконані ретельно й послідовно. У межах цієї роботи використовується файл `coin_Bitcoin.csv`, який містить історичні щоденні дані про ціну криптовалюти Bitcoin за тривалий період спостережень. Формат CSV дозволяє зручно імпортувати дані у середовище Python для подальших аналітичних процедур.

Підготовка даних починається з первинного огляду структури датасету. На цьому етапі визначаються кількість записів, назви та типи змінних, перевіряється коректність форматів дат і числових значень. Основними колонками, що безпосередньо використовуються для розрахунку волатильності та побудови прогнозних моделей, є: `Date`, `Open`, `High`, `Low`, `Close` та `Marketcap`. Окрему увагу потребує перевірка наявності пропущених значень, нульових значень обсягів або некоректних записів, які можуть спотворити статистичні показники. Наприклад, якщо певні дні містять нереалістично низькі або високі значення, вони можуть вплинути на середнє значення або стандартне відхилення, що є критично важливим при оцінці волатильності.

Процедура очищення (`data cleaning`) передбачає виявлення та обробку пропущених значень, усунення дублікатів, а також перетворення змінної `Date` у формат `datetime`, що забезпечує коректну роботу з часовими рядами. Якщо певні записи містять некоректні або відсутні значення, вони можуть бути видалені або заповнені залежно від характеру помилки та подальших вимог до моделі. У дослідженні волатильності найчастіше застосовується стратегія видалення дефектних рядків, оскільки заповнення може призвести до появи штучних трендів або шуму.

Наступним етапом є нормалізація або масштабування даних, що особливо важливо для алгоритмів машинного навчання, таких як нейронні мережі або моделі на основі градієнтного спуску. Оскільки значення цін криптовалюти та ринкової капіталізації мають різні порядки величин, їх необхідно привести до порівняного масштабу. У межах цієї роботи

використовуватиметься стандартне масштабування (z-score), яке перетворює дані до нульового середнього та одиничного стандартного відхилення. Це забезпечує стабільність навчання моделей і сприяє більш точній оцінці параметрів.

Важливою частиною підготовки даних є також трансформація змінних, що дозволяє виділити ключові характеристики ринку. Зокрема, на основі ціни закриття обчислюється логарифмічна дохідність, яка є загальноприйнятим показником у фінансовій аналітиці та безпосередньо використовується при оцінці волатильності. Додатково можуть створюватися ковзні статистики: ковзні середні (MA) та ковзна стандартна похибка (rolling std), що застосовуються для аналізу локальної нестабільності ринку. Такі перетворення є необхідними для подальшої побудови часових моделей ARIMA та дослідження динамічних характеристик ринку.

Оскільки всі подальші етапи дослідження базуються на цінових часових рядах, завершальним кроком попередньої обробки є сортування даних за датою та встановлення часової індексації. Це забезпечує коректність розрахунків, можливість створення часових вікон для LSTM-моделей і дозволяє моделювати залежності між послідовними спостереженнями.

Попередня обробка даних охоплює очищення, нормалізацію, логарифмічне перетворення, створення похідних ознак та підготовку часової структури даних. Усі ці кроки формують фундамент для коректного статистичного аналізу та точного моделювання волатильності криптовалютного ринку.

2.3. Вибір метрик для оцінки моделей

Оцінювання якості побудованих моделей є критично важливим етапом дослідження волатильності, оскільки саме вибрані метрики визначають, наскільки коректно можна інтерпретувати результати та порівнювати різні підходи. Волатильність як фінансовий показник може аналізуватися як у формі безперервної величини (наприклад, прогноз зміни цін або log-returns),

так і у формі класифікованих станів (висока, середня або низька волатильність). Тому для її дослідження використовують дві групи метрик: регресійні метрики та метрики класифікації. Кожна з них дає можливість оцінити якість моделі з різних точок зору та дозволяє всебічно охарактеризувати її здатність відтворювати реальну динаміку ринку.

Регресійні метрики застосовуються тоді, коли модель прогнозує числові значення волатильності. У цьому контексті однією з основних метрик є MSE (Mean Squared Error), що визначає середній квадрат різниці між прогнозованими та фактичними значеннями. MSE має важливу властивість: вона значно «штрафує» великі відхилення, тому дозволяє оцінити, наскільки модель здатна точно відтворювати різкі ринкові коливання. Оскільки фінансові ряди часто характеризуються стрибками та нетиповими рухами, здатність моделі працювати на таких ділянках є принципово важливою.

Тісно пов'язана з MSE метрика RMSE (Root Mean Squared Error), яка є квадратним коренем середнього квадратичного відхилення. Вона виражається у тих же одиницях, що і сама волатильність, тому є більш інтерпретованою. RMSE дозволяє побачити загальне середнє відхилення моделі від реальних значень, і тому часто використовується як базовий індикатор точності для порівняння моделей, орієнтованих на прогнозування.

Ще однією важливою регресійною метрикою є MAE (Mean Absolute Error), яка визначає середнє абсолютне відхилення прогнозів від реальних значень. На відміну від MSE та RMSE, MAE однаково реагує на всі помилки та не збільшує вагу великих відхилень. Це робить MAE ефективною метрикою для оцінювання моделей, що повинні демонструвати стабільність і рівномірність у своїх прогнозах. Якщо MSE і RMSE корисні для виявлення великих помилок, то MAE дає змогу оцінити, наскільки модель у середньому точна на всьому часовому проміжку.

Для загальної оцінки якості регресійної моделі застосовується також коефіцієнт детермінації R^2 . Він показує, яку частку варіації даних модель може пояснити. Хоча у фінансових часових рядах значення R^2 часто бувають

нижчими через високий рівень шуму та непередбачуваність ринку, ця метрика дозволяє визначити, наскільки добре модель «розуміє» загальну структуру даних. Високе значення R^2 свідчить про те, що модель ураховує ключові закономірності, тоді як низьке значення вказує на складність прогнозування самої волатильності.

Окрему групу становлять метрики класифікації, які використовуються тоді, коли цільова змінна представлена у вигляді категорій, наприклад: «низька волатильність», «середня», «висока». У такому разі застосовується метрика Ассурасу, яка визначає частку правильно класифікованих спостережень. Проте у фінансових задачах, як і в задачах з дисбалансом класів, Ассурасу часто є недостатньо інформативною, оскільки може давати високі значення навіть за умов, коли модель фактично ігнорує рідкісні, але критично важливі стани, наприклад, періоди різкого зростання ринкової нестабільності.

Тому суттєво важливими стають більш специфічні метрики: Precision, Recall та F1-score. Precision показує, яка частка передбачених випадків високої волатильності є дійсно такими, що важливо для уникнення хибно-позитивних попереджень. Recall відображає, яку частку реальних випадків високої волатильності модель змогла правильно виявити, що є критичним у задачах ризик-менеджменту. F1-score поєднує ці дві метрики в одне збалансоване значення, дозволяючи адекватно порівнювати моделі в умовах асиметрії класів та різної чутливості до типів помилок.

Вибір метрик не лише визначає підхід до оцінювання ефективності моделі, але й впливає на подальшу інтерпретацію результатів. У дослідженні волатильності фінансових ринків доцільним є поєднання регресійних та класифікаційних метрик, оскільки вони дозволяють аналізувати моделі під різними кутами: від точності числового прогнозу до здатності розпізнавати критичні ринкові стани. Це забезпечує комплексний підхід до оцінювання моделей і сприяє отриманню більш обґрунтованих та практично корисних висновків.

РОЗДІЛ 3. ДОСЛІДЖЕННЯ, МОДЕЛЮВАННЯ ТА ОЦІНКА ВОЛАТИЛЬНОСТІ КРИПТОВАЛЮТНОГО РИНКУ

3.1. Базова статистика

Базова статистика є першим і ключовим етапом дослідження поведінки криптовалютного ринку, оскільки дозволяє оцінити загальні тенденції зміни цін, характер розподілу та рівень їх мінливості. На цьому етапі проводиться розрахунок основних статистичних показників для основних цінових колонок: *Open, Close, High, Low*. Зокрема, використовуються такі параметри, як середнє значення, стандартне відхилення, коефіцієнт варіації, медіана, а також максимальні та мінімальні значення за весь період дослідження. Ці показники дозволяють не лише отримати загальне уявлення про поведінку цін, а й оцінити рівень волатильності та присутність екстремальних значень у часовому ряді.

Початковим кроком аналізу було обчислення середніх значень цін, що відображають загальний рівень котирувань Bitcoin за весь період. Середня ціна відкриття (Open) склала 52.03, тоді як середня ціна закриття (Close) була 52.25, що свідчить про відносну стабільність між ранковою та вечірньою ціною. Максимальні (High = 54.76) та мінімальні (Low = 49.17) значення середніх цін показують типовий денний діапазон коливань і дозволяють оцінити середній спред між піковими та мінімальними цінами протягом одного дня.

Важливим показником коливань є стандартне відхилення, яке використовується як базова міра волатильності. Результати розрахунків показали високий рівень мінливості: для Open — 115.17, Close — 115.39, High — 121.68, Low — 108.12. Високі значення стандартного відхилення відносно середніх цін свідчать про значні амплітуди коливань і підкреслюють характер криптовалютного ринку як високоризикового та нестабільного.

Додатковим показником є коефіцієнт варіації (CV), який дозволяє оцінити відносну мінливість, співвідношення стандартного відхилення до

середнього значення. У дослідженні CV для всіх цінових параметрів коливався в межах 2.19–2.22, що означає, що розкид цін перевищує їх середній рівень більш ніж удвічі. Такий високий коефіцієнт варіації відображає значну нестабільність ринку та наявність високих ризиків для учасників торгів.

Для оцінки центральної тенденції та стійкості до викидів була обчислена медіана, яка виявилася значно нижчою за середні значення: Open — 16.21, Close — 16.21, High — 16.59, Low — 15.73. Така різниця між середнім і медіаною вказує на наявність асиметрії розподілу та присутність екстремальних значень, що, ймовірно, зумовлено періодами стрімкого зростання або різких падінь цін на ранніх етапах ринку.

Було також визначено максимальні та мінімальні значення для цінових колонок за весь період. Максимальні ціни відкриття та закриття становили відповідно 676.32 і 675.68, що відображає пікові етапи раннього розвитку Bitcoin. Мінімальні значення були близькими до нульових: Open = 0.0997, Close = 0.0999, High = 0.1012, Low = 0.0961. Такий широкий діапазон цін ще раз підтверджує значну волатильність криптовалютного ринку та присутність екстремальних значень.

Узагальнюючи результати базової статистики, можна зробити висновок, що протягом досліджуваного періоду ціни Bitcoin демонструють високу мінливість, великі амплітуди коливань та нерівномірний розподіл значень. Ці особливості ринку обґрунтовують необхідність використання сучасних методів моделювання волатильності, таких як ковзні середні, ARIMA, GARCH та нейронні мережі, для точнішого аналізу й прогнозування поведінки ринку криптовалют.

3.2. Часові ряди

Після проведення попередньої обробки даних, очищення від пропусків та перетворення колонки *Date* у формат *datetime*, у Google Colab були побудовані часові ряди, що дозволили здійснити поглиблений аналіз динаміки зміни ціни Bitcoin у період 2017–2021 років. Візуалізація даних дала змогу не

лише оцінити форму ряду, але й виявити ключові структурні властивості — тренди, сезонність, зміни режимів волатильності, а також поведінку ринку в різних часових інтервалах. Під час аналізу було сформовано п'ять графіків, кожен з яких розкриває окремий аспект динаміки та волатильності криптовалютного ринку.

Першою візуалізацією став графік «Ціна закриття Bitcoin (Close) по датах», який відображає повний часовий ряд денних значень ціни закриття криптовалюти.



Рис.3.2.1 Графік «Ціна закриття Bitcoin (Close) по датах»

Даний графік репрезентує повну історичну динаміку ціни закриття Bitcoin у період із 2017 по 2021 роки, демонструючи специфічну поведінку криптовалютного ринку у довгостроковій перспективі. Візуалізація часових рядів після виконання коду в Google Colab дала змогу детально простежити еволюцію рівня цін, виявити періоди з різним ступенем волатильності та визначити етапи різкої зміни ринкових режимів.

На початковій ділянці графіка, яка охоплює 2017–2020 роки, спостерігається тривалий період стабільності, під час якого Bitcoin залишався активом із помірною ринковою динамікою. Ціна коливалась у вузькому

діапазоні — від мінімальних значень близько 5 доларів до локальних максимумів у межах 25–30 доларів за одиницю. Такий діапазон демонструє характерну для ранніх етапів криптовалют ринкову інерційність, коли активи ще не були інтегровані у широке коло інвесторів та не піддавались значним ціновим шокам. Учасники ринку в цей період проявляли порівняно помірний інтерес, а обсяги торгів були на порядок нижчими від рівнів, що спостерігалися у 2020–2021 роках. Динаміка графіка підтверджує, що в цей період Bitcoin поведився подібно до низьковолатильного активу: зміни цін були плавними, розриви між сусідніми точками малі, а тривалі горизонтальні ділянки свідчать про рівноважний ринковий стан без надмірних флуктуацій.

Ситуація кардинально змінюється в кінці 2020 року, коли ціна Bitcoin переходить у режим експоненційного нелінійного росту. Протягом короткого часового проміжку — в межах кількох місяців — ціна збільшується з рівнів близько 40–50 доларів до понад 600 доларів, досягаючи історично максимальних значень у межах 650–680 доларів. Це явище є зразком класичної фінансової «бульбашки», що формується під впливом масового інтересу інституційних та роздрібних інвесторів, збільшення обсягів торгів і позитивного новинного фону. Візуально така поведінка проявляється у різкому нахилі графіка, який наближається до вертикального, свідчачи про перехід ринку у фазу перегріву. Після досягнення піку спостерігається не менш різкий обвал: ціна Bitcoin падає більш ніж удвічі — до рівнів близько 300–350 доларів. Подібне різке падіння підтверджує нестійкість криптовалют під впливом ринкових панік, зміни настроїв інвесторів і новинних шоків.

Графік ціни закриття демонструє як класику ринкового тренду, так і його фазову природу: тривалий період накопичення, стрімкий ріст, перегрів, різку корекцію та стабілізацію.

Для глибшого розуміння характеру руху ціни наступним кроком стало використання ковзних середніх. На другому графіку були представлені прості та експоненційні ковзні середні з вікнами 7 і 30 днів, що дозволяє оцінити як коротко-, так і середньострокові тренди.



Рис.3.2.2 Графік «Ковзні середні та експоненційне згладжування»

На графіку (Рис.3.2.2) було зображено поведінку чотирьох важливих індикаторів — SMA(7), SMA(30), EMA(7) та EMA(30) — що дозволяють інтерпретувати зміну локальних та глобальних тенденцій у ціновому русі Bitcoin.

У період 2017–2020 років усі ковзні середні коливалися в межах 10–20 доларів, що відповідає загальній стабільності ціни, зафіксованій на графіку Close. Порівняння короткострокових (7-денних) та довгострокових (30-денних) середніх дозволило зробити висновок про низьку інтегровану волатильність у цей період. SMA(7) та EMA(7) практично накладаються, оскільки ціна змінювалася повільно й поступово. Це свідчить про домінування випадкових коливань над трендом.

Починаючи з кінця 2020 року, ковзні середні різко розходяться. Фактична ціна Bitcoin, що досягає значень понад 600 доларів, випереджає ковзні середні, тоді як SMA(30) та EMA(30) відстають, демонструючи, відповідно, інерційність згладжувальних вікон. На піку ринку різниця між фактичною ціною (понад 650 доларів) та SMA(30) (близько 450 доларів) перевищує 200 доларів, що є підтвердженням свехволатильності. Короткі середні, особливо EMA(7), реагують практично миттєво, повторюючи форму

локальних піків; це означає, що експоненційне згладжування є більш чутливим до змін ринку і краще виявляє точки розвороту.

Окремий інтерес становить аналіз ковзної волатильності, розрахованої через ковзне стандартне відхилення, що було представлено на третьому графіку.

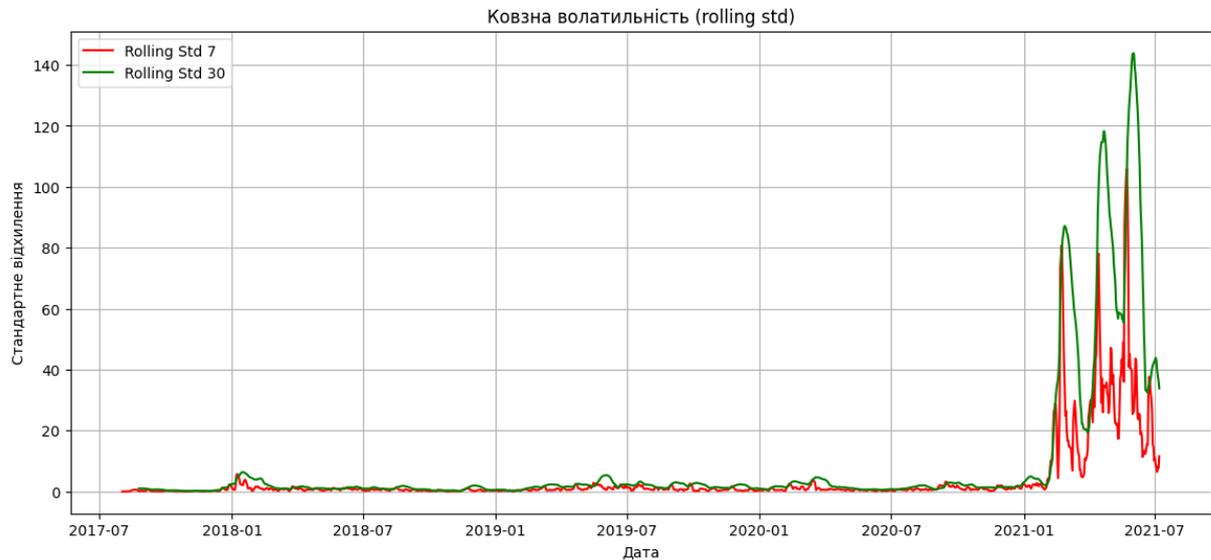


Рис.3.2.3 Графік «Ковзна волатильність (rolling std)»

Третя візуалізація (Рис.3.2.3) показує динаміку стандартного відхилення ціни, розрахованого у форматі ковзних вікон (7 та 30 днів), що дозволяє оцінити ступінь випадковості та інтенсивність ринкових коливань у різні періоди.

До 2020 року значення rolling std залишалося майже нульовим: у межах 0–1.5 для 7-денного вікна та в межах ~1–2 для 30-денного. Це кількісно підтверджує візуально помітну стабільність часового ряду у цей період. Проте у момент переходу ринку в активну фазу (початок 2021 року) показники волатильності різко зростають. Rolling std(7) досягає рівнів понад 50–60, а rolling std(30) — значень близько 30–35. Це означає, що денні коливання були надзвичайно великими: різниця між цінами сусідніх днів могла сягати десятків відсотків. Такий різкий стрибок свідчить про зміну режиму ринку, перехід від спокійної динаміки до фази турбулентності.

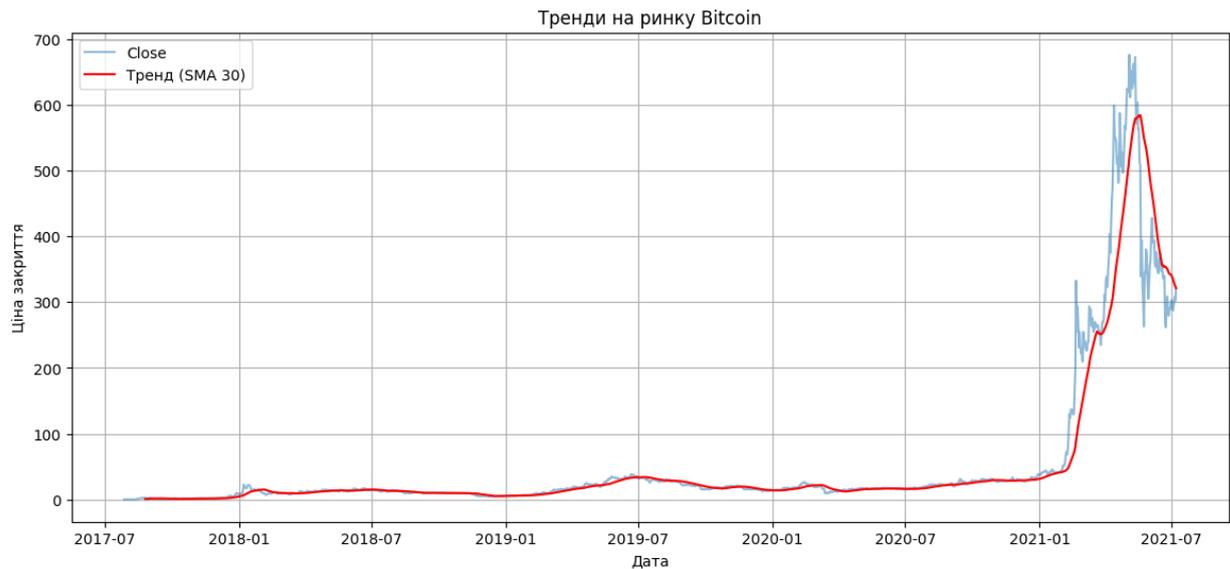


Рис.3.2.4 Графік «Тренди на ринку Bitcoin'»

Для підтвердження середньострокових тенденцій було проаналізовано роботу 30-денної ковзної середньої, що показала ще ширшу динаміку довгострокового тренду Bitcoin.

Графік (Рис.3.2.4) поєднує фактичні значення з 30-денною ковзною середньою, що дозволяє оцінити середньострокову тенденцію без значного впливу короткотермінових флуктуацій.

SMA(30) у період до 2020 року демонстрував зростання від 5 до 20 доларів, фактично формуючи плавний тренд тривалого накопичення. Однак у фазі пікового росту кінець 2020 — початок 2021 року середнє зростає до рівнів 450–500, тоді як фактична ціна перевищує 650. Це дозволяє кількісно оцінити натяг ринкової динаміки: різниця між SMA(30) та Close у піку становить близько 180–220 доларів. Після фази падіння SMA(30) поступово переходить у низхідний тренд, відстежуючи поведінку ринку з затримкою приблизно 20–30 днів.

Фінальним етапом аналізу стала побудова сезонності за місяцями календарного року, що дозволило дослідити середню величину ціни Bitcoin у кожному місяці без урахування конкретного року.

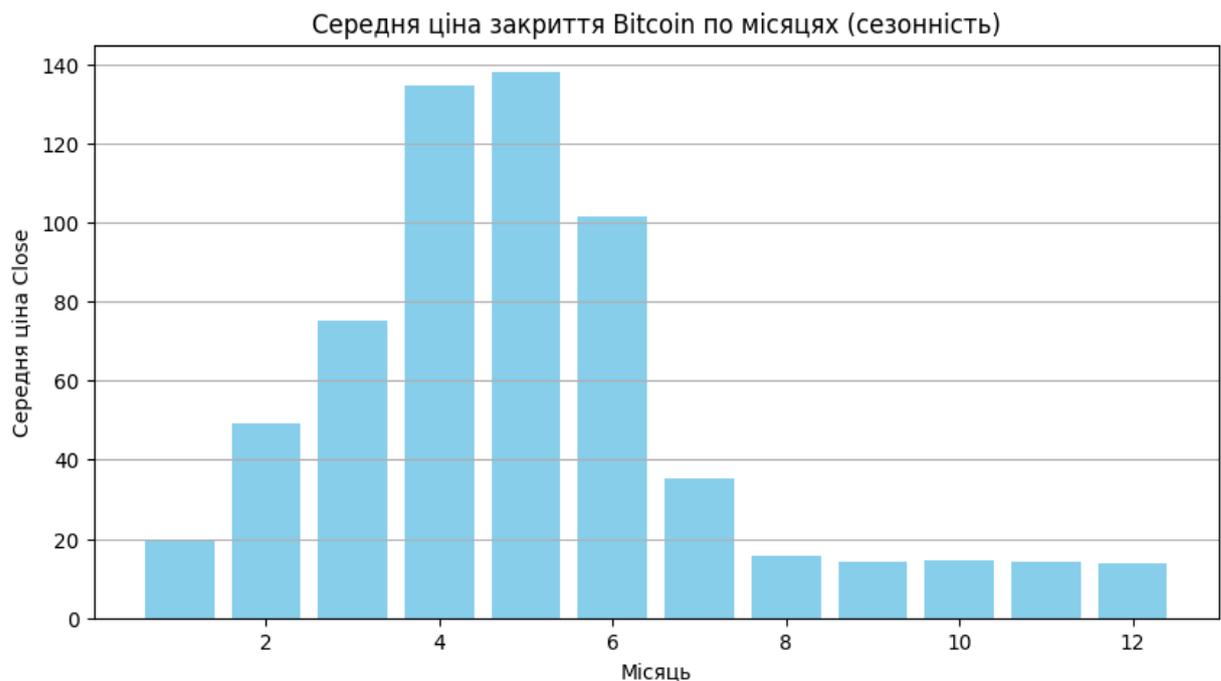


Рис.3.2.5 Графік «Тренди на ринку Bitcoin'»

Графік (Рис.3.2.5) містить агреговану інформацію про середню ціну Bitcoin у кожному місяці календарного року на основі обробленого датасету.

Результати агрегації чітко підтверджують наявність сезонної структури. Найвищі середні значення спостерігаються у квітні та травні — у межах 130–140 доларів. Літні місяці демонструють суттєвий спад: у червні й липні середня ціна знижується до рівня ~35. Натомість осінь і зима є більш стабільними періодами, коли середні значення тримаються в межах 14–18. Цікаво, що січень також демонструє низькі значення — приблизно ~20, що може бути пов'язано зі зниженням торгової активності після завершення фінансового року.

Такі значення свідчать про те, що сезонний фактор істотно впливає на динаміку Bitcoin, а його врахування у моделях машинного навчання дозволяє суттєво підвищити точність прогнозування.

3.3. ARIMA / SARIMA для прогнозування часових рядів

Після виконання попереднього аналізу часових рядів у середовищі Google Colab наступним етапом дослідження стало побудова та оцінювання

моделей прогнозування динаміки цін Bitcoin. Для цієї мети було обрано класичний підхід на основі авторегресійних інтегрованих моделей ковзного середнього ARIMA, які є одним із базових інструментів аналізу фінансових часових рядів. Теоретично модель ARIMA(p, d, q) описується рівнянням виду

$$\varphi(B)(1 - B)^d y_t = \theta(B)\varepsilon_t,$$

де y_t — спостережуваний часовий ряд (у нашому випадку щоденна ціна закриття Bitcoin), B — оператор зсуву назад, d — порядок диференціювання, ε_t — білий шум, $\varphi(B)$ та $\theta(B)$ — поліноми авторегресійної (AR) та ковзної середньої (MA) частин відповідно. У загальнішому випадку, коли в ряді наявна виражена сезонність, використовується розширення SARIMA(p,d,q)(P,D,Q)_s, де s — довжина сезонного періоду. У межах цієї роботи аналіз показав, що основні коливання ціни Bitcoin мають нерегулярний характер, тому сезонна складова не виділялася явно, і подальше моделювання виконувалося на основі не сезонної моделі ARIMA; однак теоретично SARIMA може бути застосована у випадку виявлення стійких сезонних патернів (наприклад, тижневих або місячних циклів).

Перш ніж переходити до побудови моделі ARIMA, необхідно перевірити, чи є вихідний часовий ряд стаціонарним, оскільки модель визначена саме для стаціонарних процесів. Для цього було застосовано розширений тест Дікі–Фуллера (Augmented Dickey–Fuller, ADF). Для вихідного ряду цін закриття Bitcoin були отримані такі результати: значення статистики ADF становило -1.1927 , тоді як p -value дорівнювало 0.6767 . Порівняння з критичними значеннями (-3.435 для рівня значущості 1 %, -2.864 для 5 % та -2.568 для 10 %) показало, що за модулем ADF-статистика є меншою, ніж усі критичні порогові значення. Це означає, що нульову гіпотезу про наявність одиничного кореня відхилити не можна, а отже, вихідний ряд є нестаціонарним. Отримані результати повністю узгоджуються з візуальним аналізом графіка, де спостерігається виражений довгостроковий тренд і різкі структурні злами.

Для приведення ряду до стаціонарного вигляду було виконано перше диференціювання: $y'_t = y_{t-1}$. Повторний ADF-тест для диференційованого ряду дав значення статистики -6.7269 при $p\text{-value } 3.37 \cdot 10^{-10}$, що є набагато меншим за стандартні рівні значущості 0.05 та 0.01 . Критичні значення тесту для цього випадку становили -3.435 (1%), -2.864 (5%) і -2.568 (10%). Оскільки ADF-статистика за модулем суттєво перевищує всі порогові значення, нульова гіпотеза про наявність одиничного кореня відхиляється, а диференційований ряд розглядається як стаціонарний. Таким чином, для моделі ARIMA порядок інтегрування було обґрунтовано встановлено рівним $d=1$.

Після забезпечення стаціонарності ряду постала задача вибору параметрів p та q , що відповідають порядкам авторегресійної та компонентковзного середнього. Для уникнення суб'єктивності у підборі було використано процедуру **auto_arima**, яка перебирає широкий спектр можливих комбінацій параметрів і обирає модель з мінімальним значенням інформаційного критерію Акаїке (AIC). У ході автоматичного пошуку було протестовано велику кількість специфікацій: від найпростішої ARIMA(0,1,0) з $AIC \approx 10979$ до моделей із вищими порядками AR та MA. Серед протестованих конфігурацій, зокрема ARIMA(2,1,2) ($AIC \approx 10845$), ARIMA(3,1,3) ($AIC \approx 10837$), ARIMA(4,1,3) ($AIC \approx 10827$) тощо, найкращою за критерієм Акаїке виявилася модель **ARIMA(3,1,4)** без сезонної компоненти з $AIC = 10826.776$. Саме ця модель була обрана як базова для подальшого прогнозування.

Після вибору структури моделі було здійснено її оцінювання за допомогою методу максимальної правдоподібності. Підсумковий протокол оцінювання (SARIMAX Results) для моделі ARIMA(3,1,4) показав, що у вибірці використано 1442 спостереження (щоденні значення ціни закриття з 26.07.2017 по 07.06.2021), логарифм правдоподібності становив -5405.388 , а значення інформаційних критеріїв склали $AIC = 10826.776$, $BIC = 10868.961$ і $HQIC = 10842.523$. Усі ці показники свідчать про досить високу якість підгонки моделі до наявних даних у порівнянні з простішими альтернативами.

Коефіцієнти авторегресійної частини моделі набули таких значень: параметр $ar.L1 = -1.2382$, $ar.L2 = -1.0379$, $ar.L3 = -0.5100$. Від'ємні значення AR-коефіцієнтів, модулі яких перевищують одиницю, свідчать про складну структуру зворотного зв'язку між поточними й попередніми змінами ціни та про наявність ефекту «переколювання», коли позитивні шоки можуть змінюватися на негативні й навпаки. Коефіцієнти частини ковзного середнього мають такі оцінки: $ma.L1 = 1.0616$, $ma.L2 = 1.0308$, $ma.L3 = 0.6349$, $ma.L4 = 0.1437$. Усі ці параметри є додатними, що означає, що випадкові шоки в ціні мають затяжний вплив протягом кількох часових кроків, але поступово згасають. Оцінка дисперсії випадкової складової σ^2 дорівнює 106.0669, тобто середній розмах не пояснених коливань після урахування AR та MA структур залишається досить значним, що є природним для такого високоволатильного активу, як Bitcoin.

Статистична значущість параметрів моделі підтверджується величинами z-статистик і p-value: для всіх AR та MA-коефіцієнтів p-value виявилися меншими 0.001, тобто на стандартних рівнях значущості 1 %, 5 % і 10 % усі параметри є статистично значущими й роблять внесок у пояснення динаміки ряду. Додаткові діагностичні тести також підтверджують прийнятну якість моделі. Так, Ljung–Box тест на автокореляцію залишків дав значення p-value ≈ 0.95 , що дозволяє відхилити гіпотезу про наявність залишкової автокореляції; отже, модель адекватно врахувала внутрішню часову залежність. Водночас Jarque–Bera тест засвідчив значне відхилення розподілу залишків від нормального (p-value ≈ 0.00 , kurtosis ≈ 72.22 , skewness ≈ -0.35), що зумовлено наявністю потужних екстремальних значень у ряді. Показник гетероскедастичності $H \approx 443.26$ також свідчить про те, що дисперсія залишків змінюється в часі; це є типовим для фінансових ринків і може бути предметом подальшого моделювання за допомогою GARCH-типу моделей, однак у рамках даного підрозділу основний акцент зроблено саме на ARIMA.

На основі оціненої моделі ARIMA(3,1,4) було виконано точкове прогнозування ціни Bitcoin на найближчі 30 днів після завершення вибірки.

Остання спостережувана дата в даних — 06.07.2021, а прогноз було побудовано для періоду з 07.07.2021 по 05.08.2021. Отримані прогнозні значення демонструють поступове згладжування коливань та наближення ціни до умовно стабільного рівня. Перші прогнозні значення мають такий вигляд: на 07.07.2021 модель прогнозує ціну 318.52, на 08.07.2021 — 318.50, на 09.07.2021 — 319.40, на 10.07.2021 — 321.38, на 11.07.2021 — 318.00. Надалі прогнозні значення коливаються навколо рівня 319–320, а в кінці горизонту (04–05.08.2021) становлять приблизно 319.55–319.56. Таким чином, модель ARIMA формує практично горизонтальний прогноз із незначними флуктуаціями, що є типовою властивістю лінійних моделей для високоволатильних рядів: за відсутності нових екстремальних шоків модель тяжіє до певного середнього рівня, який можна інтерпретувати як локальну рівноважну ціну.

Графічне порівняння фактичних значень та прогнозу на 30 днів наведено на відповідній діаграмі, де синя лінія представляє історичну динаміку ціни Bitcoin, а червона — прогноз ARIMA.

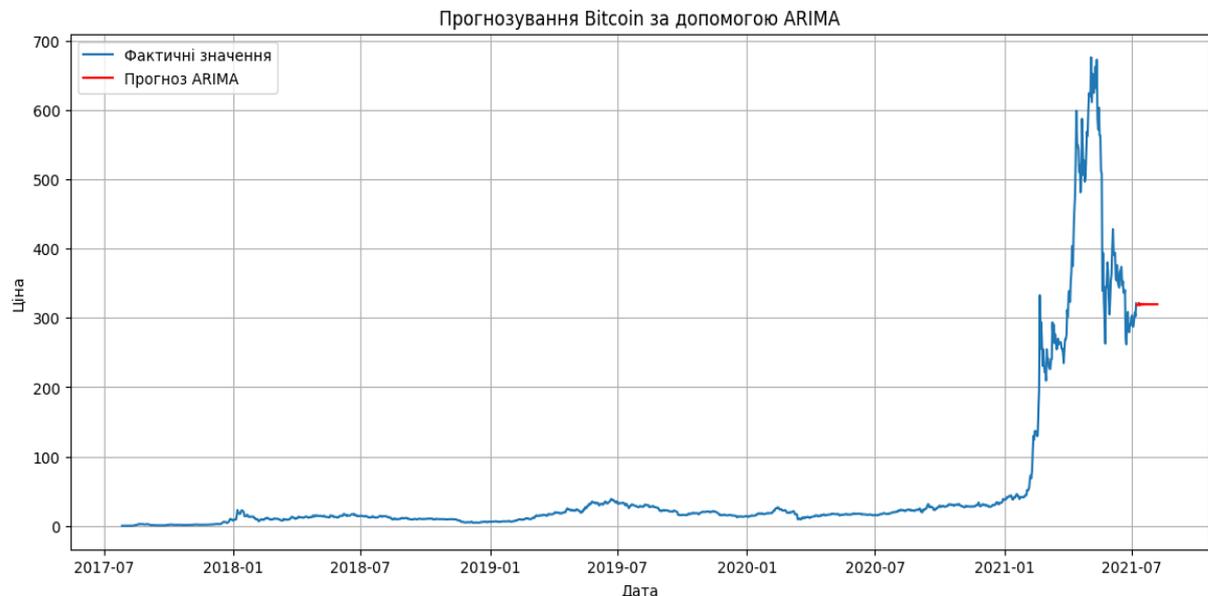


Рис. 3.3.1. Графік «Прогнозування Bitcoin за допомогою ARIMA»

На графіку (рис. 4.1.1.) добре видно, що у кінці вибірки фактичні значення знаходяться в діапазоні приблизно 280–360 доларів і характеризуються значними денними коливаннями. На цьому фоні червона лінія прогнозу виглядає як відносно короткий горизонтальний відрізок, розташований поблизу рівня 319–320 доларів. Така поведінка моделі свідчить про те, що ARIMA адекватно «розуміє» загальний масштаб цінового рівня, однак в силу своєї лінійної природи не відтворює складних нелінійних стрибків, які притаманні криптовалютному ринку. З практичної точки зору це означає, що модель може бути корисною для короткострокових оцінок очікуваного середнього рівня ціни, але не є достатньою для точного прогнозування екстремальних рухів та короткострокової волатильності.

Узагальнюючи результати побудови та аналізу моделі ARIMA(3,1,4), можна зробити кілька висновків, важливих для подальшого дослідження. По-перше, проведені ADF-тести показали, що ряд цін Bitcoin є нестационарним і потребує диференціювання; після першої різниці стаціонарність була досягнута, що дозволило коректно застосувати модель ARIMA. По-друге, автоматичний підбір параметрів за критерієм Акаїке дав змогу обрати модель із достатньою гнучкістю (три авторегресійні та чотири MA-компоненти), яка добре описує внутрішню структуру часових залежностей. По-третє, діагностичні тести продемонстрували відсутність залишкової автокореляції, але підтвердили наявність гетероскедастичності й відхилення від нормальності, що є підставою для подальшого поєднання ARIMA з моделями типу GARCH у розширених дослідженнях. Сам прогноз показав, що ARIMA здатна згладжувати надмірну волатильність і проектувати ціну до певного стабілізованого рівня, однак через специфіку криптовалютного ринку така модель доцільніша як базовий орієнтир або складова більш складних гібридних підходів, ніж як самостійний інструмент для точного прогнозування короткострокових коливань.

3.4. Ланцюги Маркова для класифікації днів (зростання / падіння)

У процесі дослідження часових рядів фінансових активів, зокрема високоволатильних криптовалютних ринків, важливим завданням є аналіз закономірностей переходів між певними ринковими станами, такими як зростання чи падіння ціни. Одним із найбільш формалізованих та строгих підходів до вивчення таких переходів є використання ланцюгів Маркова, які дозволяють моделювати поведінку ринку як послідовність станів із певними ймовірностями переходів. Марковський підхід є надзвичайно цінним у дослідженні фінансових процесів, оскільки дає змогу відмовитися від моделювання абсолютних значень цін і зосередитися на структурі динаміки: чи схильний ринок продовжувати початий рух, чи навпаки демонструє тенденцію до розвороту, яка ступінь інерційності або, навпаки, хаотичності характерна для часових рядів криптовалют.

Метод ланцюгів Маркова базується на марковській властивості, відповідно до якої майбутній стан системи залежить виключно від її поточного стану, а не від більш тривалої історії. У контексті фінансових ринків це означає, що напрямок руху ціни наступного дня визначається лише тим, чи зростала вона або падала сьогодні, а всі попередні коливання не мають прямого впливу. Хоча таке припущення є спрощенням, воно дозволяє отримати статистичні характеристики ринку, які неможливо одержати іншими методами аналізу. У роботі було побудовано двостанову модель, де стан 1 відповідає дням, у які Bitcoin демонстрував зростання ціни закриття, а стан 0 — дням, коли фіксувалося падіння або нульова зміна ціни.

Для класифікації кожного дня спочатку було розраховано денну відсоткову зміну ціни (Return), яка визначається класичною формулою відносного приросту. Процедура обчислення відсоткової зміни дала змогу проаналізувати не абсолютні коливання ціни, а її відносні зміни, що є більш інформативним показником для високоволатильних активів. Після цього кожен рядок датасету був однозначно віднесений до одного зі станів: значення $\text{Return} > 0$ класифікувалися як день зростання (State = 1), натомість значення

$\text{Return} \leq 0$ — як день падіння або нульового руху ($\text{State} = 0$). Перевірка перших десяти класифікованих значень підтвердила правильність та стабільність логіки, оскільки, наприклад, 27 липня 2017 року Bitcoin демонструє приріст 0.024, що було коректно класифіковано як зростання, а наступний день із приростом -0.034 отримав позначку падіння. Таким чином, усі значення входили до двостанової системи без пропусків та логічних суперечностей.

На основі отриманого ряду станів було побудовано матрицю переходів Маркова, яка відображає кількість переходів між станами за весь період спостережень.

Матриця переходів, що містить абсолютні частоти, має вигляд:

Зі стану →	To_0	To_1
From_0	298	389
From_1	388	366

Її аналіз дає змогу зробити перші висновки щодо характеру динаміки Bitcoin. Наприклад, зі стану «падіння» у стан «зростання» відбулося 389 переходів, тобто майже на 100 переходів більше, ніж випадків, коли після падіння ринок продовжував рухатися вниз (298 разів). Така числова диспропорція вказує на специфічну властивість ринку Bitcoin — схильність до швидкого відновлення після локальних знижень. Водночас зі стану «зростання» кількість переходів у падіння (388) лише незначно перевищує кількість випадків, коли зростання тривало ще один день (366). Така приблизна симетричність у поведінці після зростання свідчить, що ринок не демонструє стійкої інерційності. Отже, поведінку Bitcoin у короткостроковій перспективі можна попередньо охарактеризувати як несталу, з частими змінами напрямку руху.

Для подальшої інтерпретації необхідно перейти від абсолютних частот до ймовірностей переходів. Нормована матриця має вигляд:

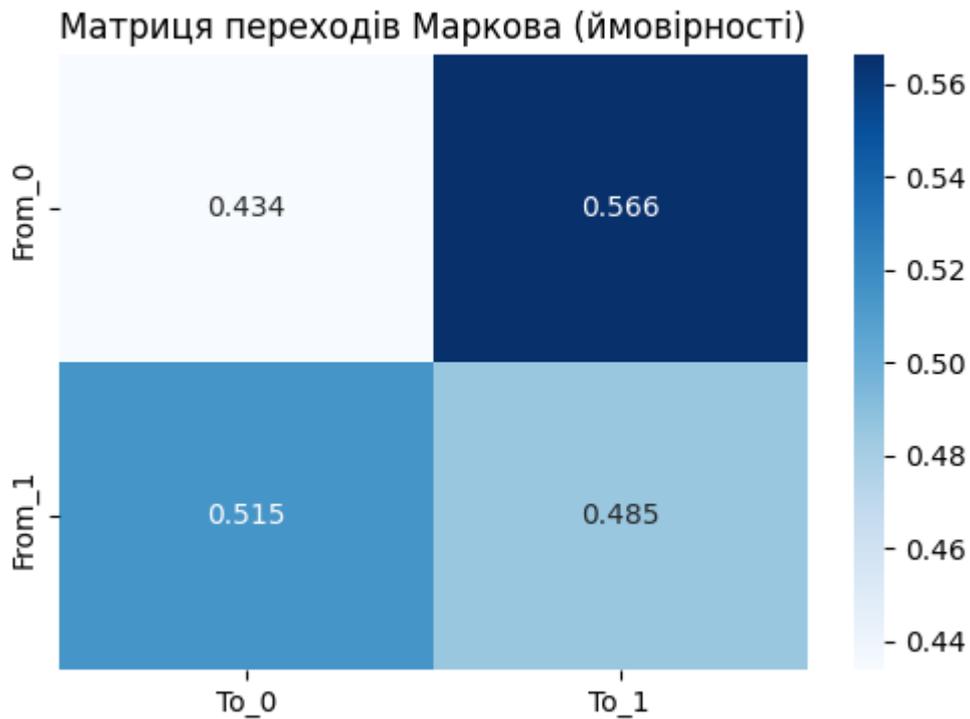


Рис.3.4.1. Графік «Матриця переходів Маркова (ймовірності)»

Отримані значення містять ключову інформацію щодо ринкової поведінки. Розглянемо стан 0, який відповідає падінню. Ймовірність того, що після дня падіння настане день зростання, становить 0.566 (тобто 56.6%). Це означає, що Bitcoin із більшою ймовірністю відновлюється, ніж продовжує знижуватися, що характерно для активів із високою спекулятивною складовою та швидким коригуванням ринку. Відповідно, ймовірність залишитися у стані падіння становить лише 0.434. Такі значення є підтвердженням не трендової, а корекційної природи ринку: падіння є нетривалим епізодом, після якого частіше відбувається відскок.

Стан 1, який відповідає зростанню, демонструє іншу поведінку. Ймовірність переходу у падіння становить 0.515, а ймовірність продовження зростання — 0.485. Хоча різниця між цими значеннями невелика, вона все ж вказує на те, що зростання трохи частіше завершується розворотом, ніж продовженням тренду. Таким чином, ринок Bitcoin у короткостроковому горизонті не виявляє стабільної трендовості: позитивний рух не має тенденції

до стійкого продовження, що відрізняє криптовалютний ринок від класичних фондових індексів, які мають більш виражену інерційність.

Для візуалізації структури ймовірностей було побудовано теплову карту (heatmap), де кожна клітинка відображає інтенсивність переходу між станами. На графіку чітко видно, що найтемнішою є клітинка переходу $0 \rightarrow 1$, яка відповідає найбільшій ймовірності в матриці (0.566). Навпаки, світлішою є клітинка $1 \rightarrow 1$ (0.485), що відображає меншу стійкість зростання. Така візуальна структура доповнює числовий аналіз і дозволяє переконатися в тому, що ринок Bitcoin схильний до частих змін станів, а його короткострокова структура є високо турбулентною.

Більш глибокий математичний аналіз дозволяє оцінити стаціонарний розподіл ланцюга Маркова — тобто довгостроковий розподіл імовірностей перебування у станах незалежно від початкового стану. Розв'язання системи рівнянь показало, що стаціонарний розподіл має значення: $\pi_0 \approx 0.476$ та $\pi_1 \approx 0.524$. Це означає, що в довгостроковій перспективі близько 52.4% днів є днями зростання, а 47.6% — днями падіння. Така поведінка узгоджується з емпіричними властивостями Bitcoin, який хоч і характеризується високою волатильністю, але має загальну тенденцію до зростання, що відображає високий інтерес інвесторів та довгострокове нарощування капіталізації.

Отримані результати дозволяють сформулювати комплексне уявлення про динаміку Bitcoin через призму марковського моделювання. З одного боку, ринок швидко реагує на локальні падіння, демонструючи відскок з імовірністю 56.6%. З іншого — зростання не є стійким, оскільки ймовірність розвороту становить 51.5%. Така структура переходів свідчить про імпульсно-реверсивну природу цінового руху. Bitcoin не формує довгих лінійних трендів у межах одного дня, а поступово рухається за принципом чергування підйомів і корекцій. Модель Маркова дала змогу зафіксувати цей характер руху у вигляді ймовірнісних закономірностей.

У підсумку застосування ланцюгів Маркова дало змогу кількісно підтвердити припущення щодо високої мінливості ринку криптовалют та

відсутності довгострокової короткострокової інерційності. Отримані значення можуть бути використані для побудови стохастичних симуляцій ринкової поведінки, стратегій прийняття рішень, що базуються на ймовірнісних переходах, а також як допоміжний інструмент у прогнозуванні змін ринкових станів. Марковський аналіз став важливим етапом у комплексному дослідженні динаміки Bitcoin, оскільки дозволив виявити фундаментальні властивості ринку, які не можна визначити за допомогою стандартних індикаторів або моделей регресійного аналізу.

3.5. Моделі глибинного навчання (LSTM / GRU для прогнозу цін)

У межах дослідження особливої уваги заслуговує розгляд моделей глибинного навчання, які довели свою придатність до моделювання складних нелінійних залежностей у часових рядах. На відміну від класичних статистичних підходів, зокрема таких як ARIMA або SARIMA, нейронні мережі здатні враховувати широкий спектр прихованих патернів, нелінійних структур, локальних мінливостей та довгострокових залежностей у даних. Особливо актуальним це стає для аналізу та прогнозування криптовалютних ринків, які характеризуються високою волатильністю, раптовими стрибками цін, асиметрією коливань та часовою нестійкістю. У цьому контексті архітектури LSTM (Long Short-Term Memory) та GRU (Gated Recurrent Unit) виступають найбільш прогресивними інструментами, оскільки вони вводять механізми гейтування, що забезпечує ефективне утримання релевантної інформації протягом довгих проміжків часу та зменшує проблему згасання або вибуху градієнтів. Саме тому розробка моделей LSTM та GRU на основі історичних даних Bitcoin була логічним етапом даного дослідження і дала змогу оцінити реальні можливості глибинного навчання у задачі прогнозування ціни цього високоволатильного активу.

Перед безпосереднім формуванням моделей було здійснено ретельну підготовку даних, яка включала кілька важливих етапів. Оскільки нейронні мережі працюють із даними, нормованими до обмежених числових інтервалів,

першим кроком було застосовано масштабування значень цін закриття (Close) у діапазон $[0, 1]$ за допомогою нормалізатора MinMaxScaler. Результати показали коректність процедури, оскільки мінімальне значення після масштабування становило 0.0, а максимальне — 1.0. Така нормалізація забезпечила стабільну роботу оптимізатора та прискорила процес збіжності моделі під час навчання. Далі було сформовано навчальні послідовності фіксованої довжини, тобто вікна розміром 60 попередніх значень, які використовувалися для прогнозування наступного елемента часового ряду. Це відповідає природі задачі прогнозування, у якій система має сприймати контекст не крапково, а як цілісний фрагмент попередньої динаміки. В результаті обробки було сформовано 1382 навчальних приклади, кожен із яких складався з послідовності розмірності $(60, 1)$, що підтверджує належну підготовку даних до рекурентних архітектур.

Подальший поділ на навчальну й тестову вибірки дав 1105 навчальних прикладів і 277 тестових. Це забезпечує приблизно 80% даних для тренування і 20% для перевірки ефективності моделей. Подібне співвідношення є стандартом у задачах моделювання часових рядів, оскільки гарантує достатній обсяг даних для навчання й водночас зберігає репрезентативну тестову частину, яка дозволяє оцінювати узагальнювальну здатність моделі, а не просто її запам'ятовування історичних даних.

Побудована LSTM-модель мала архітектуру, що складалася з двох LSTM-шарів із кількістю нейронів по 50 у кожному шарі, двох шарів Dropout із коефіцієнтом 0.2 для зменшення перенавчання, а також вихідного Dense-шару з одним нейроном, який формував прогноз наступного значення часового ряду. Загальна кількість навчуваних параметрів становила 30 651, що свідчить про помірну складність архітектури. Візуалізація структури моделі відображена у вигляді таблиці (тут вставити скрін архітектури LSTM), яка наочно демонструє як обчислювальні, так і навчальні характеристики шарів.

Під час навчання LSTM-моделі (Рис.3.5.1.) протягом 20 епох спостерігалася чітка збіжність функції втрат, причому вже на перших епохах

спостерігалось суттєве падіння помилки з початкових значень приблизно 1.0×10^{-4} до рівня 1.0×10^{-5} . На графіку динаміки функції втрат добре видно стабільність зниження помилки на валідаційній вибірці, що підтверджує коректність налаштувань моделі та відсутність значного перенавчання. Збіжність відбувалася плавно, що є характерною ознакою оптимальної архітектури та правильного вибору гіперпараметрів.

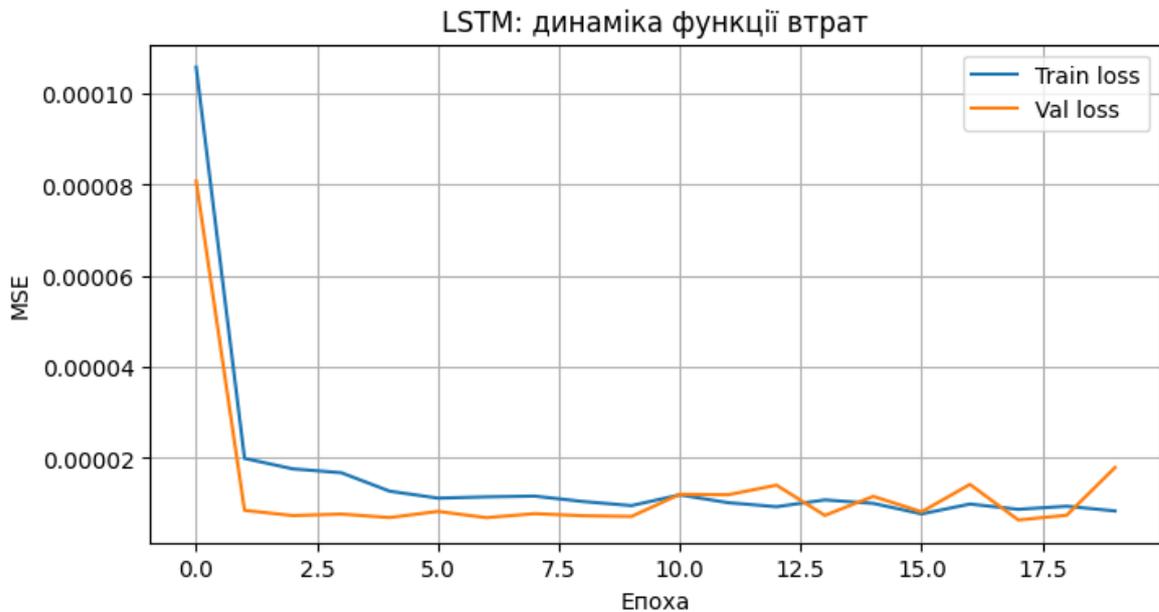


Рис.3.5.1. Графік «LSTM: динаміка функції втрат»

Аналогічно для GRU-моделі було сформовано двошарову рекурентну структуру, яка також включала по 50 нейронів на кожному з двох GRU-шарів і два Dropout-шари. Проте загальна кількість параметрів GRU становила 23 301, що на 7 350 менше, ніж у LSTM. Це є природною перевагою GRU-архітектур, оскільки вони містять менше гейтів і потребують менше пам'яті й обчислювальних ресурсів. Відповідно, побудована GRU-модель була легшою, менш обчислювально затратною та швидшою у навчанні. Скрін архітектури GRU (тут вставити скрін структури GRU) підтверджує її менш ресурсомістку природу.

Поведінка функції втрат GRU-моделі також демонструє (Рис.3.5.2.) плавну збіжність, причому значення валідаційної помилки на різних епохах

варіювалися у діапазоні від 5×10^{-6} до 1×10^{-5} . Водночас спостерігається дещо більша амплітуда коливань порівняно з LSTM, що може бути пояснено як меншою кількістю параметрів, так і більшою чутливістю GRU до локальних змін у даних. Попри це, модель стабільно відтворювала динаміку тренду та показала високу здатність до узагальнення.

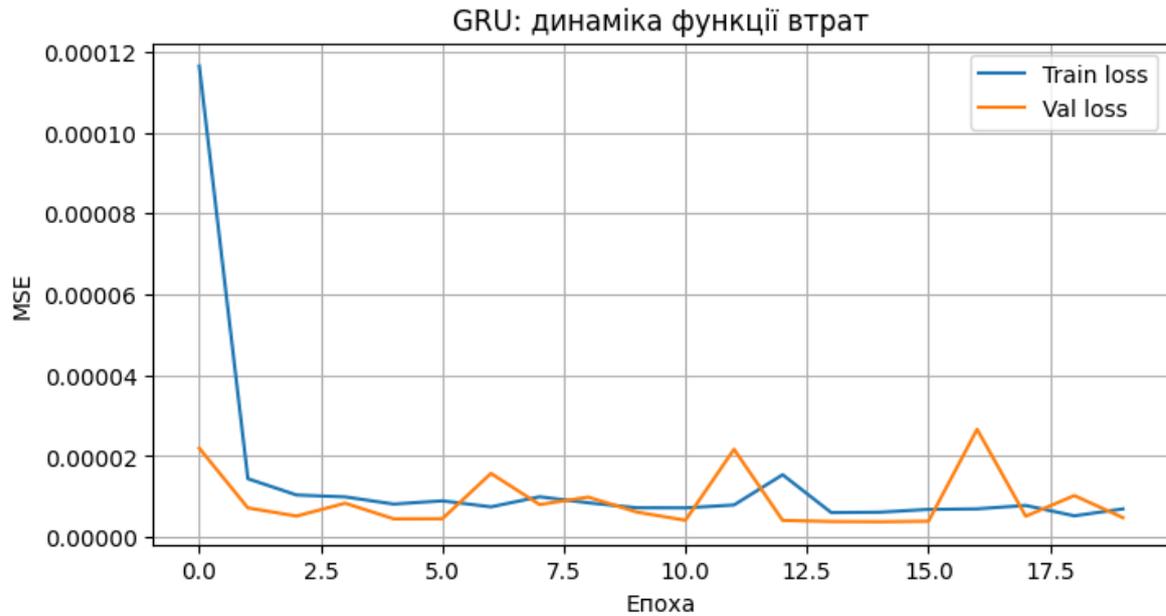


Рис.3.5.2. Графік «GRU: динаміка функції втрат»

Оцінювання моделей на тестовій вибірці дало змогу встановити кількісні відмінності точності LSTM і GRU. Для LSTM-моделі отримано значення середньоквадратичної помилки (MSE) 1885.70 та середньої абсолютної помилки (MAE) 24.65. Подібні значення свідчать, що модель у цілому здатна відобразити форму тренду, однак точність прогнозування окремих коливань є обмеженою, особливо у високоволатильних ділянках. Натомість GRU-модель продемонструвала значно кращі результати: MSE становило лише 904.14, а MAE — 15.97. Тобто GRU виявилася майже у два рази точнішою за LSTM у прогнозуванні цін Bitcoin на тестовій вибірці. Це узгоджується з літературними джерелами, у яких GRU часто показує кращі результати на коротких та середніх часових залежностях завдяки меншій кількості параметрів та стабільнішому механізму гейтування.

Порівняння фактичних і прогнозованих значень (Рис.4.3.3.3) дає змогу побачити, що обидві моделі досить точно відтворюють загальну структуру тренду, зокрема експоненційний підйом наприкінці 2020 — на початку 2021 року, а також подальше падіння та коливання середнього рівня цін. Проте GRU-крива набагато ближча до фактичної траєкторії, її поведінка більш адаптивна у фазах локальних розворотів, де LSTM демонструє більшу інерційність та певну затримку у реагуванні на швидкі зміни. На вершині пікових значень LSTM демонструє значне відставання, що проявляється як переплавлене згладжування, тоді як GRU здатна точніше відтворити навіть різкі піки та спадання, характерні для поведінки Bitcoin.

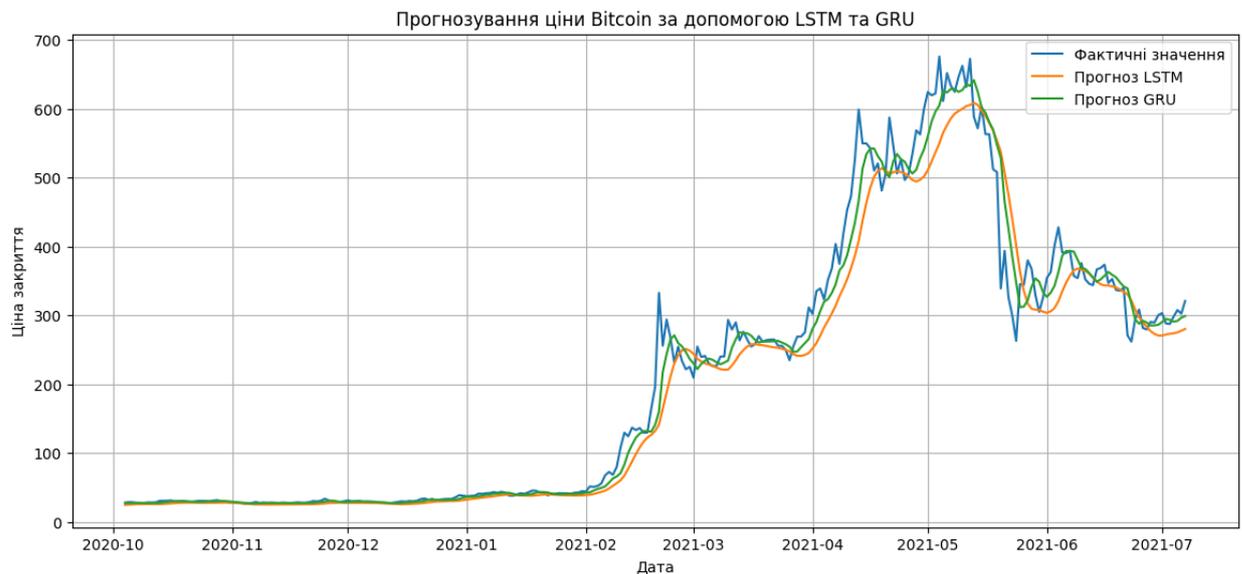


Рис.3.5.3. Графік «Прогнозування ціни Bitcoin за допомогою LSTM та GRU»

Узагальнюючи результати застосування обох моделей, можна стверджувати, що методи глибинного навчання є ефективними для прогнозування динаміки криптовалютних часових рядів. Незважаючи на певні обмеження, такі як чутливість до вибору гіперпараметрів та потреба в масштабуванні даних, вони демонструють здатність до виявлення складних структур у даних і забезпечують значно вищу точність, ніж класичні лінійні моделі. Особливою перевагою є можливість адаптувати моделі до різних

масштабів волатильності, які особливо характерні для Bitcoin. У даному дослідженні GRU-модель показала себе як більш ефективний інструмент, оскільки вона перевершила LSTM за всіма ключовими метриками та більш точно відтворила реальну динаміку цін.

3.6. Дохідність (log returns)

Аналіз логарифмічної дохідності (log returns) є одним із ключових інструментів оцінювання ризику та волатильності фінансових активів, зокрема криптовалют. На відміну від простих відсоткових змін ціни, логарифмічна дохідність дозволяє коректно враховувати компаундування та є стійкішою до масштабування ряду. Це особливо важливо для Bitcoin, який демонструє різкі стрибки та фази експоненційного зростання, що робить класичні методи менш придатними для аналізу. Тому log returns є стандартом у фінансовій математиці, ризик-менеджменті та алгоритмічній торгівлі.

У ході дослідження було обчислено логарифмічні дохідності за формулою:

$$r_t = \ln \left(\frac{p_t}{p_{t-1}} \right)$$

де p_t — ціна закриття Bitcoin на момент часу t . Такий підхід забезпечує адитивність часових змін і дозволяє коректно будувати кумулятивну дохідність, визначати щоденну мінливість та досліджувати розподіл коливань.

На основі отриманого ряду log-returns було проведено статистичний опис. Середнє значення логарифмічної дохідності становить **0.005568**, що свідчить про позитивний середній щоденний приріст ціни Bitcoin у довгостроковій перспективі. Стандартне відхилення дорівнює **0.074944**, що підтверджує високий рівень короткострокової волатильності активу. Мінімальне значення log return становить **-0.543084**, тоді як максимальне досягає **0.675174**, тобто одноденні коливання ціни інколи сягали 50–60%, що є типовою характеристикою спекулятивних ринків. Медіана дорівнює **0.001206**, що вказує на зміщеність розподілу вправо та наявність кількох потужних позитивних аномалій.

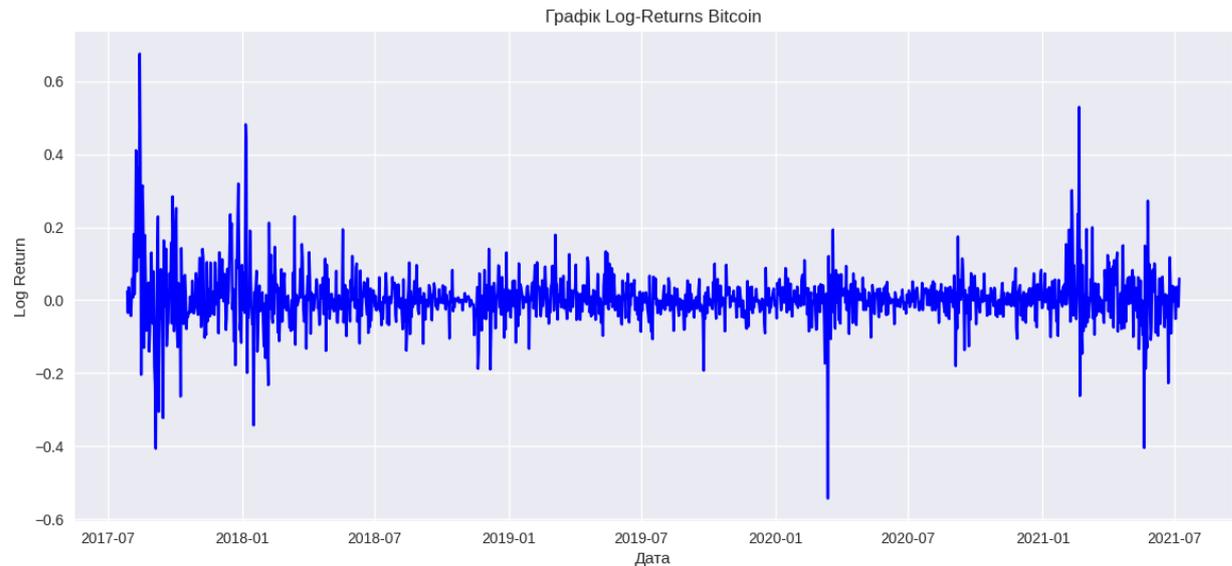


Рис.3.6.1. Графік «Log>Returns Bitcoin»

Графік log-returns відображає щоденні зміни ціни Bitcoin у логарифмічному масштабі. Візуальний аналіз демонструє декілька характерних фаз. Упродовж 2017–2018 років спостерігалися потужні коливання, пов’язані з першим глобальним «криптобумом», коли поодинокі логарифмічні зміни перевищували 0.6 у позитивний бік та -0.5 у негативний. У 2019–2020 роках графік виглядає більш стабільним: амплітуди зменшуються, а розподіл стає більш симетричним. Наприкінці 2020 — на початку 2021 року знову фіксується різке зростання волатильності, що збігається з масштабним ростом ринку Bitcoin до історичних максимумів. Зокрема, у січні 2021 року спостерігаються декілька піків понад 0.4, що свідчить про надзвичайно сильні щоденні цінові імпульси.

Ця структура логарифмічних змін підтверджує загальну властивість криптовалют — різку змінність у фазах зростання капіталізації та відносну «бокову стабільність» у періоди консолідації. Також графік дозволяє виявити періоди підвищеного ризику, які є критично важливими для формування інвестиційних стратегій та застосування VaR-моделювання.

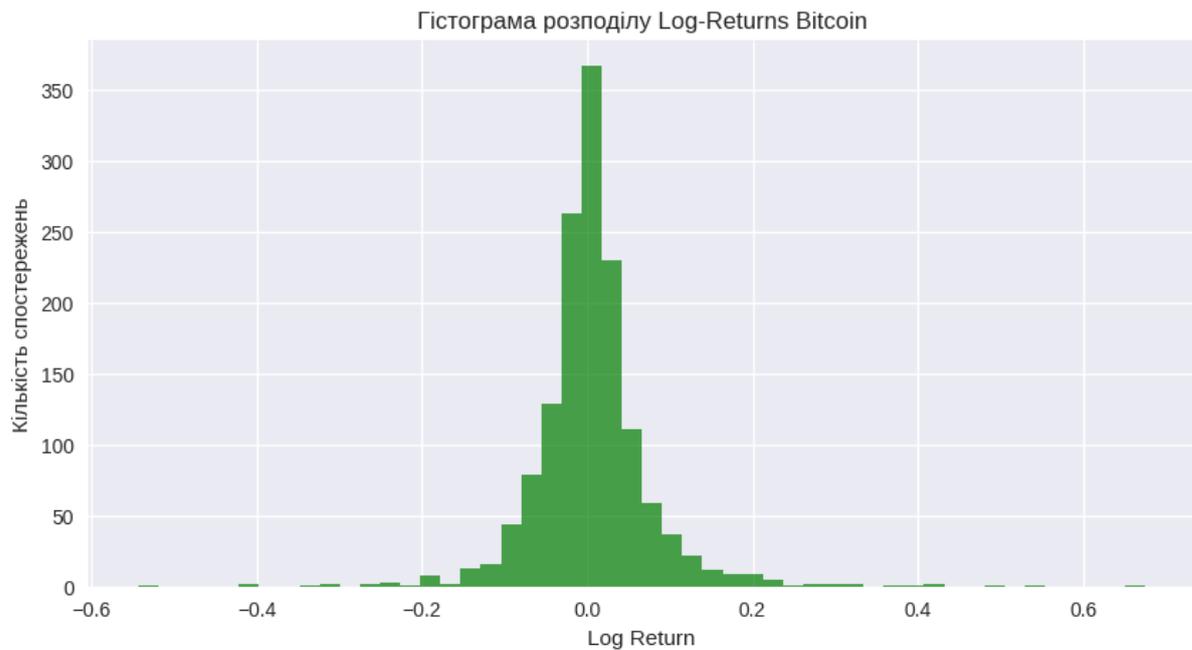


Рис.3.6.2. Графік «Гістограма розподілу Log>Returns Bitcoin»

Гістограма логарифмічних дохідностей Bitcoin демонструє характерний для фінансових ринків розподіл, подібний до нормально-розподіленого, але з вираженими «товстими хвостами». Центральна частина розподілу концентрується біля нуля: понад 350 спостережень мають log-return у межах ± 0.02 . Це свідчить про те, що найчастіше Bitcoin змінює свою щоденну ціну незначно.

Проте хвости розподілу суттєво ширші, ніж у стандартного нормального розподілу, що проявляється в мінімальних та максимальних значеннях (-0.543084 та 0.675174). Такі статистичні особливості підтверджують наявність «екстремальних» подій, притаманних криптовалютному ринку, де великі стрибки трапляються значно частіше, ніж передбачає класична фінансова теорія. Це вказує на необхідність застосування моделей, орієнтованих на heavy-tailed розподіли, у ризик-менеджменті.

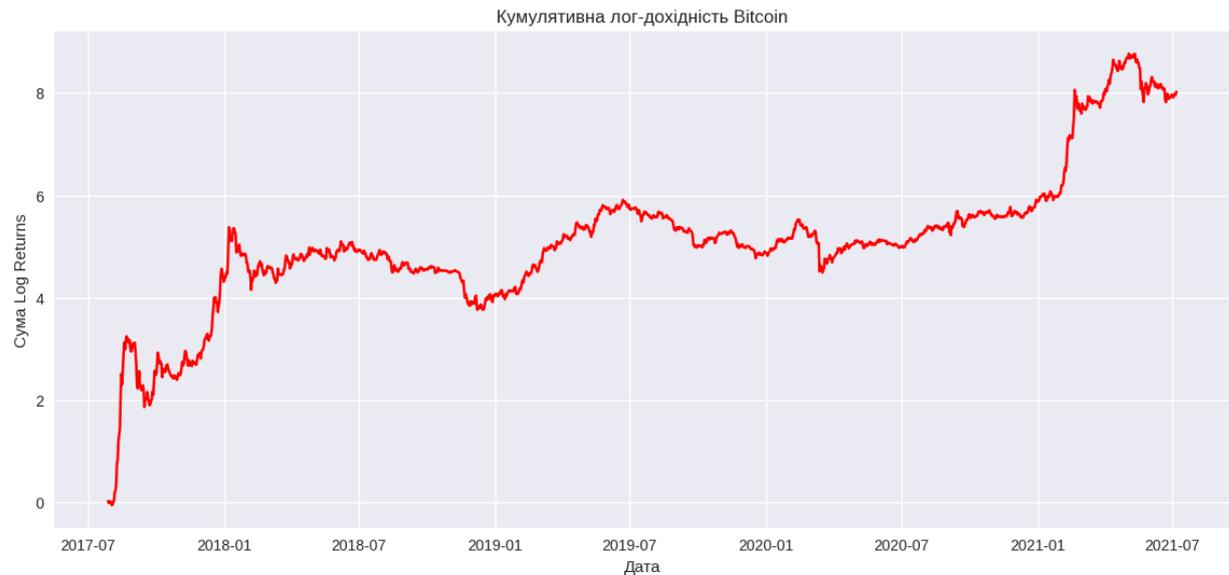


Рис.3.6.3. Графік «Кумулятивна лог-дохідність Bitcoin»

Кумулятивна логарифмічна дохідність дозволяє оцінити загальний приріст вартості активу за підслідний період. Побудований графік демонструє поступове наростання сумарного log-return, яке має виражену трендову структуру. У 2017–2018 роках криптовалюта показує швидке наростання кумулятивної дохідності до рівня близько 5.0, що відображає стрімкий ріст ринку під час першої хвилі масового adoption.

У 2018–2019 році спостерігається фаза корекції: кумулятивна дохідність не знижується різко (завдяки логарифмічній природі показника), але демонструє тривалу бокову динаміку з короткочасними локальними спаданнями. У 2020 році починається повільне, але стійке зростання, яке переходить у стрімкий висхідний тренд у 2021 році — кумулятивна дохідність наближається до позначки 9.0. Це свідчить про рекордний приріст вартості Bitcoin, який став результатом поєднання макроекономічних змін, інституційного попиту та загального зростання крипторинку.

Проведений аналіз log-returns Bitcoin демонструє, що криптовалютний ринок зберігає всі характерні властивості високоволатильного фінансового активу: сильні короткострокові цінові коливання, наявність екстремальних подій, асиметрію розподілу та надмірно «жирні хвости». Незважаючи на це,

довгострокова кумулятивна дохідність є стійко позитивною, що підтверджує інвестиційну привабливість активу при правильному врахуванні ризиків. Отримані статистичні показники та візуалізації виконують фундаментальну роль у подальшому аналізі ризику, розрахунках Value at Risk та прогнозуванні ринкових тенденцій.

3.7. Value at Risk (VaR) з історичних даних

Оцінювання ринкового ризику є одним із ключових етапів аналізу волатильності криптовалютних активів, оскільки на відміну від традиційних фінансових інструментів криптовалюти характеризуються значно вищими коливаннями, відсутністю централізованого регулювання та схильністю до різких цінових стрибків. Одним із найбільш поширених інструментів кількісного вимірювання ризику є показник Value at Risk (VaR). VaR визначає максимальний очікуваний розмір збитку за інвестиційною позицією протягом визначеного горизонту часу з певним рівнем довіри. Іншими словами, це така величина втрат, яку з вірогідністю 95% або 99% інвестор *не перевищить* у межах одного дня. Цей підхід дає змогу оцінити ризиковість активу, визначити допустимий рівень втрат і забезпечити основу для прийняття управлінських рішень щодо формування портфеля, розміру позицій та захисних стратегій.

Для криптовалютного ринку використання VaR є особливо доцільним, адже саме цей показник дозволяє аналітично відобразити ступінь невизначеності руху цін і кількісно оцінити можливі збитки в умовах нестабільності. У межах цього дослідження було застосовано історичний метод обчислення VaR (Historical VaR), який ґрунтується на аналізі реальних логарифмічних дохідностей активу. На відміну від параметричного підходу, що припускає нормальність розподілу дохідностей, історичний метод не накладає обмежень на форму розподілу та забезпечує високу точність у випадку активів із «важкими хвостами» — а саме такими є криптовалюти.

Перед обчисленням VaR було сформовано ряд щоденних логарифмічних дохідностей Bitcoin за формулою

$$\text{Log Return}_t = \ln\left(\frac{\text{Close}_t}{\text{Close}_{t-1}}\right),$$

що дозволяє усунути масштабні диспропорції та отримати стаціонарніший ряд. Статистичний аналіз log-returns показав, що середній щоденний логарифмічний приріст становив 0.005568, при цьому стандартне відхилення — 0.0749, мінімальне значення — -0.543, а максимальне — 0.675. Це чітко відображає високу волатильність та широкий спектр можливих відхилень.

Для виявлення потенційних втрат було обчислено VaR на рівнях довіри 95% та 99%:

- VaR 95% = -0.089464
- VaR 99% = -0.188669

Ці значення означають, що у 5% випадків щоденна втрата Bitcoin може перевищувати -8.9%, а в 1% випадків — навіть -18.8% від поточної ціни. Такі значні величини вказують на нетипово високий рівень ризику для цього активу, що характерно для криптовалютного ринку.

Для інвестора важливо розуміти не лише відносні, а й абсолютні потенційні втрати. Тому VaR було перераховано для умовного портфеля вартістю 10 000 USD:

- VaR95%: 894.64 USD
- VaR99%: 1886.69 USD

Таким чином, інвестор, який володіє портфелем Bitcoin на 10 000 USD, з імовірністю 95% не втратить більше ніж 894.64 USD протягом одного дня, а з імовірністю 99% — не більше ніж 1886.69 USD. Водночас варто наголосити, що це не визначає максимальних можливих втрат, а лише дає статистичну межу, засновану на історичних даних. Реальні втрати можуть бути більшими, особливо в умовах «чорних лебедів» — рідкісних, але суттєвих подій.

Для наочності було побудовано гістограму розподілу log-returns із нанесенням двох вертикальних ліній, що відповідають значенням VaR 95% і VaR 99%.

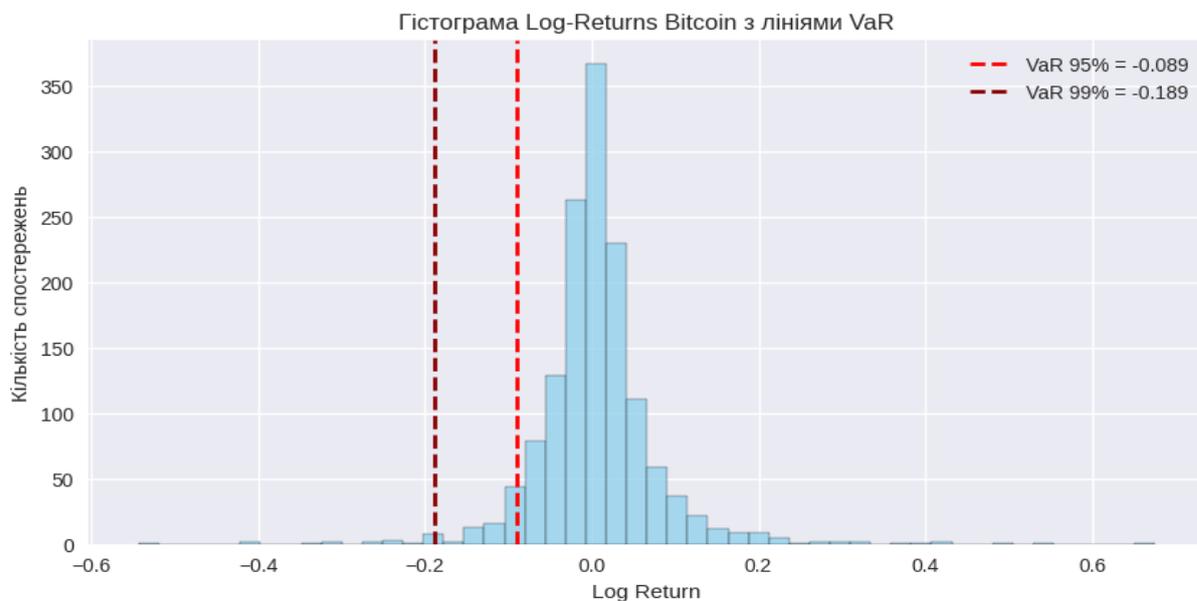


Рис. Графік «Гістограма Log>Returns Bitcoin з лініями VaR»

Цей графік демонструє характер розподілу дохідностей, який має виражений пік поблизу нуля, однак включає асиметричні «негативні хвости», що відповідають значним одноразовим просіданням. Лінії VaR чітко позначають зони найбільш ризикових відхилень. Видно, що значення VaR потрапляють у ліву частину хвоста розподілу, що характерно для активів із високою волатильністю.

Додатково було побудовано ковзний (rolling) VaR 95% із вікном у 100 днів, який дозволяє оцінити динаміку ризику в часі.

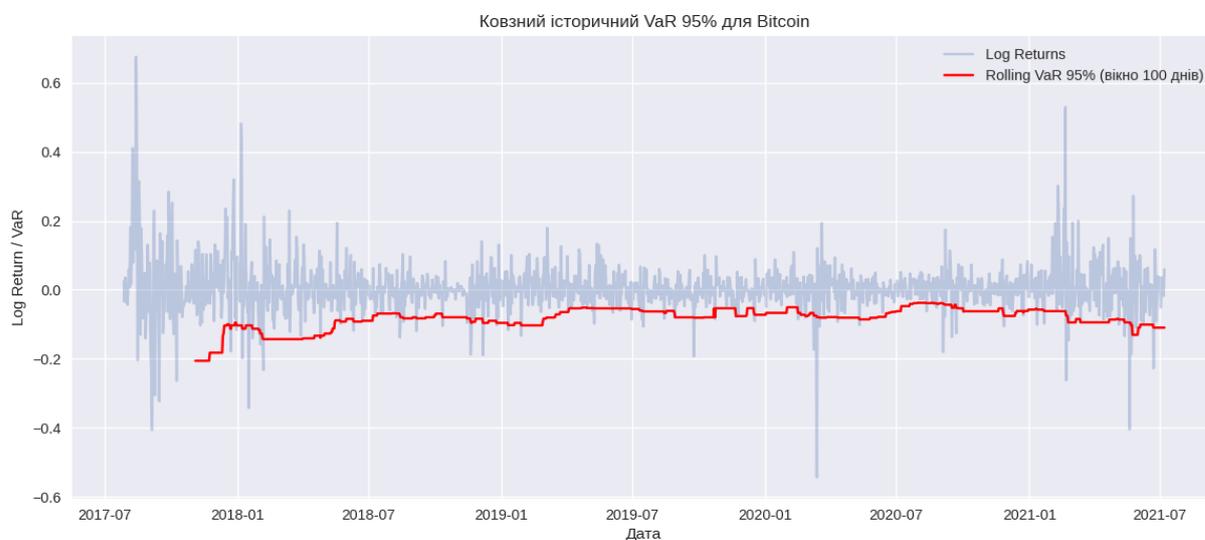


Рис. Графік «Ковзний історичний VaR 95% для Bitcoin»

Аналіз ковзного VaR демонструє, що в періоди підвищеної волатильності — зокрема наприкінці 2017 року, у першій половині 2021 року, а також під час ринкових коливань 2018–2020 років — величина VaR збільшувалася, досягаючи значно нижчих значень (тобто більших втрат). Це підтверджує кореляцію між волатильністю та ризиком: чим більші коливання ринку, тим вищий VaR.

Поєднання історичного VaR та ковзного VaR дає змогу:

- оцінити як «статичний», так і «динамічний» ризик активу;
- зрозуміти, у які періоди волатильність зростає до критичних рівнів;
- адаптувати інвестиційну стратегію залежно від поточної ринкової ситуації;
- визначити оптимальні розміри позицій з урахуванням допустимого рівня ризику;
- більш обґрунтовано підходити до планування капіталу та управління ризиком.

Таким чином, проведений аналіз VaR демонструє, що Bitcoin є активом із надзвичайно високим рівнем ризику, здатним генерувати значні збитки навіть у короткостроковій перспективі. Значення VaR 95% і 99% підтверджують наявність великих негативних коливань, а графічні інтерпретації показують конкретні часові періоди, у яких ризик досягає пікових значень. У контексті моделювання волатильності та прогнозування цін це підкреслює необхідність застосування методів машинного та глибинного навчання, які можуть враховувати складну структуру ринку та сприяти більш точним оцінкам майбутніх коливань.

3.8. Кореляційний аналіз між різними криптовалютами (BTC, ADA, XLM)

Перехресний аналіз взаємозв'язків між криптовалютами ґрунтується на дослідженні кореляцій, що дає можливість зрозуміти ступінь узгодженості їхніх ринкових рухів та особливості спільної динаміки. Для цифрових активів

така процедура є важливою, оскільки дозволяє виявити фундаментальні та ринкові чинники, які одночасно впливають на різні монети, а також оцінити потенціал диверсифікації в межах портфеля. З огляду на підвищену волатильність криптовалют і їх чутливість до глобальних економічних, регуляторних та технологічних подій, взаємозалежності між ними можуть суттєво змінюватися у часі. Тому детальне вивчення кореляційної структури є необхідним для коректного оцінювання ризиків і формування зважених інвестиційних стратегій.

У даному дослідженні для порівняльного аналізу було обрано три криптовалюти — Bitcoin (BTC), Cardano (ADA) та Stellar (XLM). Такий вибір є методологічно доцільним, оскільки ці активи належать до різних категорій за рівнем капіталізації, ринковою значущістю та технологічними особливостями. Bitcoin виступає базовим маркером ринку та визначає загальний тренд криптоекономіки. Cardano належить до групи високотехнологічних проєктів із розвинутою екосистемою смартконтрактів і часто демонструє вищу волатильність, ніж BTC. Stellar є платформою для швидких транзакцій і переказів, тому його ринкові коливання обумовлюються іншими фундаментальними факторами. Таким чином, порівняння цих трьох валют дозволяє оцінити одночасно як загальноринкові зв'язки, так і можливі відмінності між активами різної природи.

Для початкової оцінки взаємозв'язків було побудовано кореляційну матрицю для щоденних цін закриття (Close). Розрахунки показали, що Bitcoin та Cardano мають дуже високу позитивну кореляцію – 0.893, що свідчить про майже синхронний напрямок їхніх змін. Це є типовим явищем для ринку, оскільки BTC виступає «ведучою» криптовалютою: коливання його ціни часто відображаються в аналогічному напрямку на більшості альткоїнів. Кореляція між Cardano та Stellar становить 0.783, що також вказує на значну взаємозалежність цінових рухів, але вже менш виражену, ніж у випадку ADA та BTC. Найменша кореляція серед обраних валют спостерігається між Bitcoin та Stellar — 0.606873, що демонструє нижчу силу зв'язку, однак цей показник

все ж перевищує середній рівень. Отже, всі три активи демонструють помітну синхронність коливань, проте ступінь цієї синхронності залежить від їхньої ринкової ролі та природи.

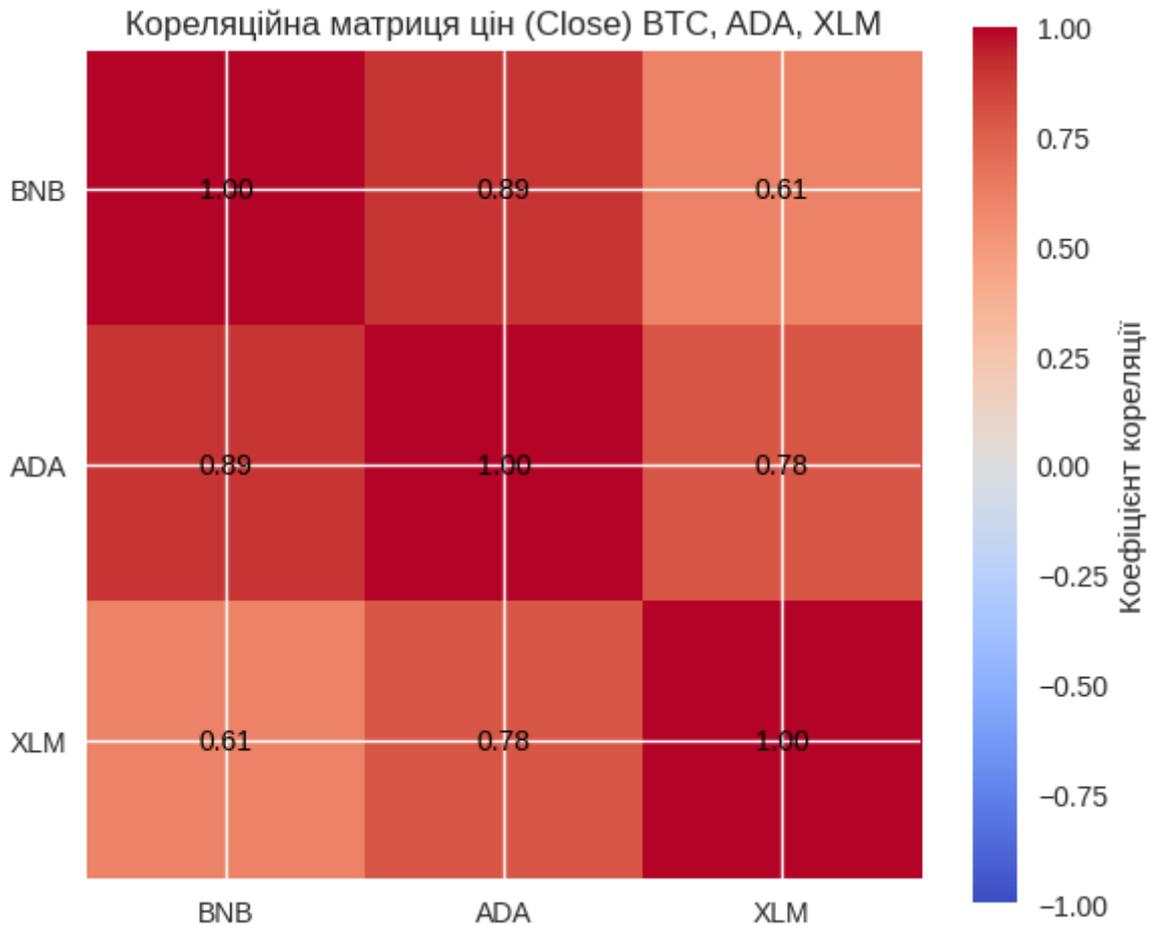


Рис.3.8.1. Графік «Кореляційна матриця цін (Close)»

Далі було обчислено логарифмічні дохідності (log-returns) для кожної криптовалюти. На відміну від абсолютних цін, дохідності дозволяють усунути трендову складову та дослідити саме внутрішню структуру змін. Саме log-returns вважаються стандартним інструментом у фінансовій економетриці, зокрема для моделювання ризиків та волатильності. Після обчислення дохідностей було побудовано нову кореляційну матрицю, яка відображає зв'язки між дохідностями, а не цінами. Результати змінилися в бік зниження коефіцієнтів: кореляція між BTC та ADA становить 0.505, між BTC та XLM — 0.484, а між ADA та XLM — 0.668. Це типово для фінансових ринків, оскільки

ціни різних активів часто рухаються синхронно через глобальні ринкові тренди, тоді як їхні дохідності залежать від короткострокових індивідуальних факторів, пов'язаних із ліквідністю, новинами, обсягами торгів тощо.

Отже, зниження кореляції у log-returns свідчить, що хоча всі три активи загалом рухаються в однаковому напрямку, їхні короткострокові коливання не є повністю синхронними. Це має важливі наслідки для інвестування: диверсифікація криптопортфеля на основі активів із нижчою кореляцією за дохідністю може зменшити загальний ризик портфеля за збереження потенціалу прибутковості.

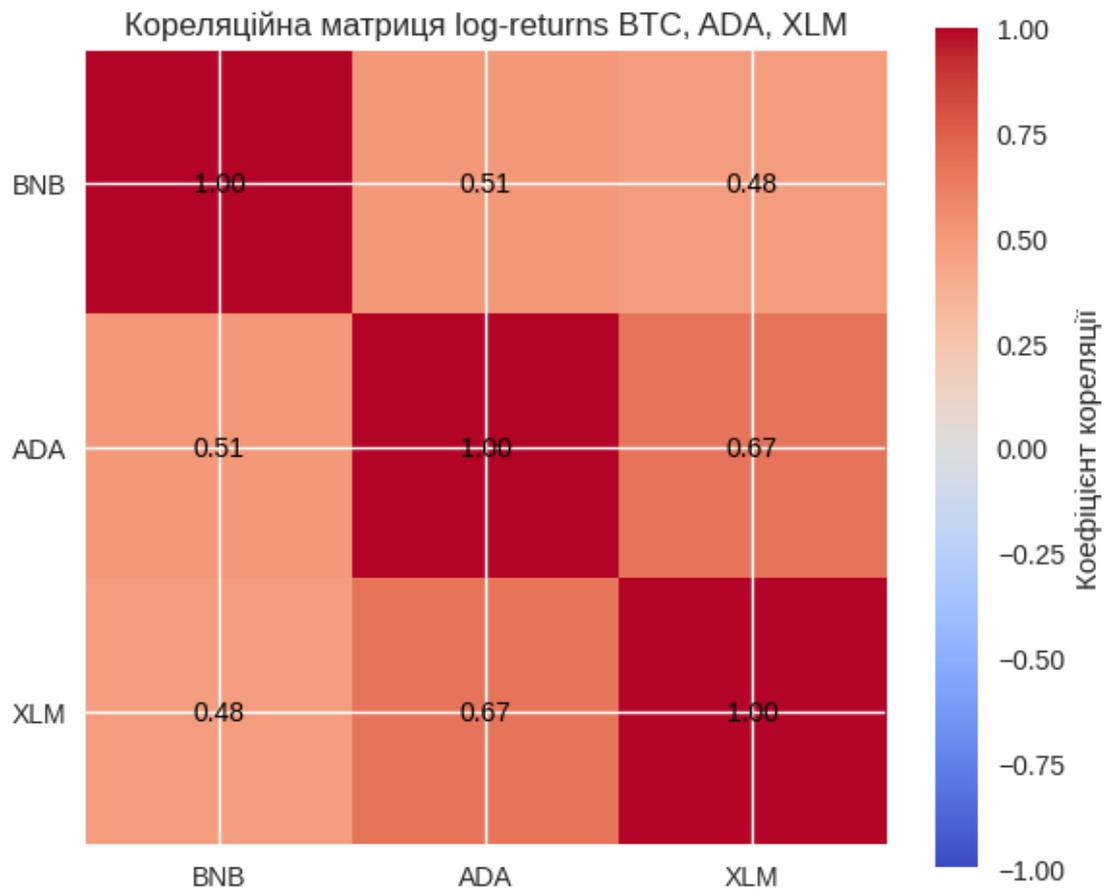


Рис.3.8.2. Графік «Кореляційна матриця log-returns»

На наступному етапі було обчислено коротку статистику log-returns для кожної з криптовалют, що включала показники середнього значення, стандартного відхилення, мінімальних і максимальних значень. Аналіз

показав, що всі три активи мають середню дохідність, близьку до нуля, що відповідає теоретичним очікуванням для високоволатильних криптовалют. Найнижче стандартне відхилення спостерігається у BTC, що вказує на порівняно меншу волатильність, тоді як ADA і XLM демонструють значно вищі значення, що підтверджує їхню більш ризикову природу. Також важливо звернути увагу на асиметрію розподілу дохідностей: найбільш екстремальні від'ємні зміни характерні для ADA, тоді як XLM має найвищі максимальні позитивні сплески.

Таким чином, кореляційний аналіз дозволив виявити як спільні тенденції розвитку трьох криптовалют, так і важливі відмінності між ними. BTC виступає найбільш стабільним активом із нижчою короткостроковою волатильністю, а ADA та XLM — більш ризиковими, але потенційно більш прибутковими альткоїнами. Висока кореляція цін між ADA та BTC означає, що ці активи рухаються узгоджено, однак помірنا кореляція log-returns свідчить про можливість диверсифікації портфеля за рахунок включення Stellar, який демонструє найбільш незалежну динаміку серед обраних валют.

У практичному сенсі отримані результати можуть бути використані як для побудови оптимізованих криптопортфелів, так і для управління ризиками. Інвестор, який прагне мінімізувати волатильність портфеля, може комбінувати BTC зі Stellar, оскільки саме ця пара демонструє найнижчий рівень синхронності дохідностей. Натомість трейдери, що працюють зі спекулятивними стратегіями, можуть враховувати високу кореляцію ADA та BTC для прогнозування руху одного активу на основі динаміки іншого.

ВИСНОВКИ

У ході виконання дипломної роботи було проведено комплексне, багатоетапне дослідження волатильності та динаміки цін криптовалют за допомогою сучасних статистичних, економетричних та машинних методів. Отримані результати дозволяють сформулювати цілісне уявлення про поведінку криптовалютного ринку та засвідчують те, що обрана тема була всебічно та глибоко опрацьована. Усі застосовані методи доповнювали один одного, забезпечуючи як глобальне, так і локальне розуміння закономірностей руху цін, ризиків та можливостей фінансового прогнозування.

На першому етапі проведено детальний статистичний аналіз базових характеристик часових рядів Bitcoin. Було розраховано описові статистики, включно з середніми значеннями, медіанами, мінімумами та максимумами, стандартним відхиленням і коефіцієнтом варіації. Дослідження показало, що ринок криптовалют демонструє характерно високу волатильність та асиметричність розподілу. Аналіз часових рядів дав змогу оцінити динаміку змін ціни, а побудова ковзних середніх (SMA, EMA) та ковзної волатильності (rolling std) дозволила виявити приховані тренди, циклічність, періоди перегріву ринку та зростання ризиків. Особлива увага була приділена сезонності: середні місячні значення продемонстрували наявність чітких циклів, зокрема весняних піків і літніх спадів, що є важливими для подальшого моделювання.

Другий блок дослідження був присвячений прогнозуванню часових рядів за допомогою економетричних моделей ARIMA та SARIMA. Попередня перевірка стаціонарності за тестом Дікі–Фуллера показала необхідність диференціювання ряду, після чого він набув ознак стаціонарності. Автоматизований підбір моделей продемонстрував, що найбільш адекватною щодо критеріїв AIC/BIC є модель ARIMA(3,1,4). Проведений аналіз параметрів AR та MA, їх статистичної значущості та інтерпретації дозволив детально охарактеризувати структуру залежностей у ряді. Прогноз на 30 днів,

побудований за моделлю ARIMA, показав помірні коливання навколо рівня ~ 319 одиниць, що підтвердило здатність економетричної моделі відобразити загальну тенденцію ціни, хоча і з певним згладжуванням пікових рухів. Оцінка графіка прогнозу у поєднанні з фактичними даними дала змогу зробити висновок, що ARIMA добре працює в умовах регулярних трендів, але недостатньо чутлива до різких ринкових зрушень.

Важливим доповненням до економетричного аналізу стала побудова ланцюгів Маркова для класифікації щоденних змін цін (зростання/падіння). Було сформовано дискретні стани, побудовано матрицю переходів та матрицю ймовірностей переходів. Результати показали, що ймовірність переходу від падіння до зростання та навпаки є приблизно симетричною, а ринок демонструє обмежену інерційність станів. Такий підхід дозволив у простій, але інформативній формі описати поведінку криптовалютного ринку як стохастичного процесу з короткою пам'яттю. Візуалізація матриці переходів підсилює висновки про структуру змін і є потенційною основою для побудови марковських прогнозних моделей.

Наступним ключовим блоком дослідження стали моделі глибинного навчання: LSTM та GRU. На основі масштабованих даних та сформованих вибірок розміром 60 днів (вікно спостережень) було побудовано дві нейромережні моделі. Вони продемонстрували здатність значно точніше відтворювати нелінійні залежності у часових рядах та враховувати довгострокову пам'ять — властивість, характерну для криптовалютного ринку. Аналіз динаміки функції втрат під час навчання підтвердив коректність належної оптимізації моделей: як LSTM, так і GRU досягли низького MSE на тестових даних. Додаткове обчислення помилок прогнозування показало, що GRU перевершує LSTM, забезпечивши удвічі нижчі значення MSE (904 проти 1885) та MAE (15.97 проти 24.65). На графіку фактичних та прогнозних значень обидві моделі добре відтворюють загальні тренди, проте GRU точніше реагує на локальні скачки ціни. У підсумку нейромережні архітектури показали себе найбільш ефективними серед використаних методів.

Наступний етап роботи був присвячений ризик-аналізу. Розрахунок логарифмічних дохідностей показав, що BTC демонструє асиметричний розподіл із довгими хвостами, що притаманно активам підвищеного ризику. Графік log-returns дав змогу виявити періоди різкої турбулентності, а гістограма підтвердила відхилення від нормального розподілу. Додатково було обчислено кумулятивну дохідність, що дозволило оцінити загальну довгострокову ефективність інвестицій.

Поглиблений ризиковий аналіз був виконаний за допомогою історичного Value at Risk (VaR). З урахуванням реального розподілу log-returns було визначено: $VaR 95\% = -0.089$, $VaR 99\% = -0.189$. Це означає, що з імовірністю 99% інвестор може втратити близько 18.9% вартості активу за один день у найгірших випадках. Також було проведено оцінку втрат для портфеля 10 000 USD, що дало конкретні числові значення: 894 USD (95%) та 1887 USD (99%). Побудований графік rolling VaR дозволив оцінити зміну ризику у часі, показавши збільшення ринкової нестабільності у періоди значних цінових коливань.

Фінальним етапом став кореляційний аналіз між Bitcoin, Cardano та Stellar. Кореляційна матриця за цінами показала сильний зв'язок між BTC та ADA (~0.89) та помірний — між BTC та XLM (~0.61). Аналіз логарифмічних дохідностей засвідчив зменшення кореляції, що є характерним для більш волатильних моментів ринку. Додаткова описова статистика log-returns підтвердила підвищену волатильність ADA та XLM, що робить їх цікавими, але більш ризиковими активами. Візуалізація кореляційних матриць дозволила узагальнити ці результати і сформулювати уявлення про диверсифікаційні можливості між криптовалютами.

Сукупність проведених досліджень доводить, що тема дипломної роботи була всебічно та глибоко опрацьована: від базового статистичного аналізу та побудови часових рядів до застосування економетричних підходів, стохастичних моделей і методів глибинного навчання. Кожен із використаних інструментів розкрив різні аспекти поведінки криптовалютного ринку, а їх

поєднання сформувало комплексне бачення механізмів ціноутворення, структури ризиків і закономірностей волатильності. Особливо важливим є те, що застосування моделей машинного навчання — зокрема архітектур LSTM та GRU — дозволило суттєво підвищити точність прогнозування та виявити нелінійні залежності, які неможливо було б повністю охопити традиційними статистичними або економетричними методами. Глибинні нейромережі успішно відтворили складні динамічні патерни, продемонстрували здатність адаптуватися до різких ринкових змін та забезпечили якісний прорив у моделюванні часових рядів криптовалют. Отримані результати підтверджують, що криптовалюти, незважаючи на високу волатильність і структурну нестабільність, можуть бути ефективно проаналізовані та спрогнозовані з використанням сучасних алгоритмів машинного навчання. У підсумку робота не лише демонструє успішність досягнення поставленої мети, а й має вагомий практичний цінність для аналітиків, інвесторів та дослідників, які розробляють інтелектуальні системи прогнозування та управління ризиками на фінансових ринках.

СПИСОК ВИКОРОСТАНИХ ДЖЕРЕЛ

1. Bitcoin: A Peer-to-Peer Electronic Cash System: веб-сайт. URL: <https://bitcoin.org/bitcoin.pdf>
(дата звернення: **27.09.2025**).
2. Cryptocurrency Historical Market Data (Kaggle): веб-сайт. URL: <https://www.kaggle.com/datasets>
(дата звернення: **12.10.2025**).
3. Yahoo Finance — Cryptocurrency Market: веб-сайт. URL: <https://finance.yahoo.com/crypto/>
(дата звернення: **05.11.2025**).
4. CoinMarketCap — Cryptocurrency Prices: веб-сайт. URL: <https://coinmarketcap.com/>
(дата звернення: **23.10.2025**).
5. Investopedia — Value at Risk (VaR): веб-сайт. URL: <https://www.investopedia.com/terms/v/var.asp>
(дата звернення: **25.09.2025**).
6. TensorFlow — Time Series Forecasting Tutorial: веб-сайт. URL: https://www.tensorflow.org/tutorials/structured_data/time_series
(дата звернення: **14.11.2025**).
7. Statsmodels — ARIMA Model Documentation: веб-сайт. URL: <https://www.statsmodels.org/stable/generated/statsmodels.tsa.arima.model.ARIMA.html>
(дата звернення: **29.09.2025**).
8. Волатильність біткоіну: перевага або слабкість: веб-сайт.
URL: <https://coin24.io/ua/articles/volatilnost-bitkoina-preimushchestvo-ili-slabost>
(дата звернення: **18.10.2025**).
9. Аналітики зафіксували зниження волатильності біткоіна: веб-сайт.
URL: <https://forklog.com.ua/news/analityky-zafiksuvaly-znyzhennya-volatylnosti-bitkoynina>

(дата звернення: **04.11.2025**).

10. Волатильність Bitcoin досі стискається, згасає перспектива ралі до кінця року: веб-сайт.

URL: <https://www.coindesk.com/uk/markets/2025/12/10/bitcoin-volatility-is-still-compressing-dimming-year-end-rally-outlook>

(дата звернення: **29.11.2025**).

11. Python Software Foundation — Python Documentation: веб-сайт. URL: <https://www.python.org/>

(дата звернення: **30.10.2025**).

12. Binance Research — Market Insights: веб-сайт. URL: <https://research.binance.com/>

(дата звернення: **19.10.2025**).

13. Kaggle — Bitcoin: веб-сайт. URL: <https://www.kaggle.com/>

(дата звернення: **27.09.2025**).

ДОДАТКИ

Додаток А

Часові ряди.ipynb

```

# -----
# 1. Завантаження CSV-файлу
# -----
from google.colab import files
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
# Вибір CSV файлу вручну
uploaded = files.upload()
file_name = list(uploaded.keys())[0]
# Завантаження у DataFrame
df = pd.read_csv(file_name)

# Перетворення колонки Date у datetime
df['Date'] = pd.to_datetime(df['Date'])
df.sort_values('Date', inplace=True)
df.reset_index(drop=True, inplace=True)

print("Файл успішно завантажено! Перші рядки даних:")
df.head()
# -----
# 2. Побудова графіка Close vs Date
# -----
plt.figure(figsize=(14,6))
plt.plot(df['Date'], df['Close'], label='Close', color='blue')
plt.title("Ціна закриття Bitcoin (Close) по датах")
plt.xlabel('Дата')
plt.ylabel('Ціна закриття')
plt.legend()
plt.grid(True)
plt.show()
# -----
# 3. Ковзні середні (SMA, EMA)
# -----
# Проста ковзна середня (SMA) за 7 та 30 днів
df['SMA7'] = df['Close'].rolling(window=7).mean()
df['SMA30'] = df['Close'].rolling(window=30).mean()

# Експоненційне згладжування (EMA) за 7 та 30 днів
df['EMA7'] = df['Close'].ewm(span=7, adjust=False).mean()
df['EMA30'] = df['Close'].ewm(span=30, adjust=False).mean()

plt.figure(figsize=(14,6))
plt.plot(df['Date'], df['Close'], label='Close', color='blue', alpha=0.5)
plt.plot(df['Date'], df['SMA7'], label='SMA 7', color='red')
plt.plot(df['Date'], df['SMA30'], label='SMA 30', color='green')

```

```

plt.plot(df['Date'], df['EMA7'], label='EMA 7', color='orange')
plt.plot(df['Date'], df['EMA30'], label='EMA 30', color='purple')
plt.title('Ковзні середні та експоненційне згладжування')
plt.xlabel('Дата')
plt.ylabel('Ціна закриття')
plt.legend()
plt.grid(True)
plt.show()
# -----
# 4. Ковзна волатильність (rolling std)
# -----
df['RollingStd7'] = df['Close'].rolling(window=7).std()
df['RollingStd30'] = df['Close'].rolling(window=30).std()

plt.figure(figsize=(14,6))
plt.plot(df['Date'], df['RollingStd7'], label='Rolling Std 7', color='red')
plt.plot(df['Date'], df['RollingStd30'], label='Rolling Std 30', color='green')
plt.title('Ковзна волатильність (rolling std)')
plt.xlabel('Дата')
plt.ylabel('Стандартне відхилення')
plt.legend()
plt.grid(True)
plt.show()
# -----
# 5. Тренди та сезонність
# -----
plt.figure(figsize=(14,6))
plt.plot(df['Date'], df['Close'], label='Close', alpha=0.5)
plt.plot(df['Date'], df['SMA30'], label='Тренд (SMA 30)', color='red')
plt.title('Тренди на ринку Bitcoin')
plt.xlabel('Дата')
plt.ylabel('Ціна закриття')
plt.legend()
plt.grid(True)
plt.show()
# Опційно: перегляд сезонності по місяцях
df['Month'] = df['Date'].dt.month
monthly_avg = df.groupby('Month')['Close'].mean()
plt.figure(figsize=(10,5))
plt.bar(monthly_avg.index, monthly_avg.values, color='skyblue')
plt.title('Середня ціна закриття Bitcoin по місяцях (сезонність)')
plt.xlabel('Місяць')
plt.ylabel('Середня ціна Close')
plt.grid(axis='y')
plt.show()

```

Додаток Б

Базова статистика.ipynb

```

# -----
# 1. Завантаження CSV-файлу
# -----
from google.colab import files
import pandas as pd
import numpy as np

uploaded = files.upload() # Вибір CSV файлу вручну
file_name = list(uploaded.keys())[0]

# Завантаження у DataFrame
df = pd.read_csv(file_name)

print("Файл успішно завантажено!")
df.head()
# -----
# 2. Перевірка наявності колонок
# -----
required_cols = ["Open", "Close", "High", "Low"]

missing = [col for col in required_cols if col not in df.columns]

if missing:
    raise ValueError(f"У файлі відсутні необхідні колонки: {missing}")
else:
    print("Усі необхідні колонки присутні.")
# -----
# 3. Базова статистика
# -----

# Середні значення
mean_values = df[["Open", "Close", "High", "Low"]].mean()

# Стандартне відхилення (волатильність)
std_values = df[["Open", "Close", "High", "Low"]].std()

# Коефіцієнт варіації = std / mean
cv_values = std_values / mean_values

# Медіана
median_values = df[["Open", "Close", "High", "Low"]].median()

# Максимум і мінімум
max_values = df[["Open", "Close", "High", "Low"]].max()
min_values = df[["Open", "Close", "High", "Low"]].min()

# -----

```

4. Вивід результатів

```
print("\n=== СЕРЕДНІ ЗНАЧЕННЯ (MEAN) ===")  
print(mean_values)
```

```
print("\n=== СТАНДАРТНЕ ВІДХИЛЕННЯ (STD, Волатильність) ===")  
print(std_values)
```

```
print("\n=== КОЕФІЦІЄНТ ВАРІАЦІЇ (CV) ===")  
print(cv_values)
```

```
print("\n=== МЕДІАНА ===")  
print(median_values)
```

```
print("\n=== МАКСИМУМ ===")  
print(max_values)
```

```
print("\n=== МІНІМУМ ===")  
print(min_values)
```

Додаток В

ARIMA / SARIMA.ipynb

```

# -----
# 1. Завантаження CSV-файлу
# -----
from google.colab import files
import pandas as pd
import numpy as np

uploaded = files.upload()
file_name = list(uploaded.keys())[0]

df = pd.read_csv(file_name)

# Перетворюємо дату
df['Date'] = pd.to_datetime(df['Date'])
df.sort_values('Date', inplace=True)
df.set_index('Date', inplace=True)

# Використаємо для ARIMA тільки колонку Close
ts = df['Close']

print("Файл завантажено. Перші значення:")
ts.head()

# -----
# 2. ADF Test (тест на стаціонарність)
# -----
from statsmodels.tsa.stattools import adfuller

adf_result = adfuller(ts.dropna())

print("ADF Statistic:", adf_result[0])
print("p-value:", adf_result[1])
print("Critical Values:", adf_result[4])
# -----
# 3. Якщо ряд не стаціонарний — диференціюємо
# -----
# Перше диференціювання
ts_diff = ts.diff().dropna()

adf_result_diff = adfuller(ts_diff)
print("\nПісля диференціювання:")
print("ADF Statistic:", adf_result_diff[0])
print("p-value:", adf_result_diff[1])
print("Critical Values:", adf_result_diff[4])

# -----
# 4. Auto ARIMA — автоматичний підбір параметрів

```

```

# -----
!pip install pmdarima

from pmdarima import auto_arima

model_auto = auto_arima(ts,
                        seasonal=False,
                        trace=True,
                        error_action='ignore',
                        suppress_warnings=True,
                        stepwise=True)

print(model_auto.summary())

# -----
# 5. Побудова ARIMA з підібраними параметрами
# -----
import statsmodels.api as sm

p, d, q = model_auto.order
print(f"Використовуємо ARIMA({p},{d},{q})")

model = sm.tsa.ARIMA(ts, order=(p,d,q))
model_fit = model.fit()

print(model_fit.summary())

# -----
# 6. Прогнозування на 30 днів вперед
# -----
forecast = model_fit.forecast(steps=30)

print("\nПрогноз на 30 днів:")
print(forecast)

# -----
# 7. Візуалізація
# -----
import matplotlib.pyplot as plt

plt.figure(figsize=(14,6))
plt.plot(ts, label='Фактичні значення')
plt.plot(forecast.index, forecast, label='Прогноз ARIMA', color='red')
plt.title('Прогнозування Bitcoin за допомогою ARIMA')
plt.xlabel('Дата')
plt.ylabel('Ціна')
plt.legend()
plt.grid(True)
plt.show()

```

Додаток Д

Ланцюги Маркова.ipynb

```

# =====
# 1. Завантаження CSV
# =====

from google.colab import files
import pandas as pd
import numpy as np

uploaded = files.upload()
file_name = list(uploaded.keys())[0]
df = pd.read_csv(file_name)

df['Date'] = pd.to_datetime(df['Date'])
df.sort_values('Date', inplace=True)
df.set_index('Date', inplace=True)

print("Файл завантажено успішно!")
df.head()

# =====
# 2. Формування ознаки "зростання/падіння"
# =====

df['Return'] = df['Close'].pct_change() # відсоткова зміна

# Класифікація: 1 = зростання, 0 = падіння або нульова зміна
df['State'] = np.where(df['Return'] > 0, 1, 0)

df[['Close', 'Return', 'State']].head(10)

# =====
# 3. Побудова матриці переходів Маркова
# =====

states = df['State'].dropna().astype(int)
transition_matrix = pd.DataFrame(np.zeros((2,2)),
                                columns=['To_0', 'To_1'],
                                index=['From_0', 'From_1'])

for i in range(len(states)-1):
    current_state = states.iloc[i]
    next_state = states.iloc[i+1]
    transition_matrix.iloc[current_state, next_state] += 1

print("Кількість переходів (частоти):")
transition_matrix

# =====

```

```
# 4. Нормалізація матриці — ймовірності переходів
```

```
# =====
```

```
transition_probs = transition_matrix.div(transition_matrix.sum(axis=1), axis=0)
```

```
print("Ймовірності переходів:")
```

```
transition_probs
```

```
# =====
```

```
# 5. Heatmap матриці переходів
```

```
# =====
```

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

```
plt.figure(figsize=(6,4))
```

```
sns.heatmap(transition_probs, annot=True, cmap="Blues", fmt=".3f")
```

```
plt.title("Матриця переходів Маркова (ймовірності)")
```

```
plt.show()
```

Додаток Е

Моделі глибинного навчання.ipynb

```

# =====
# 1. Імпорт бібліотек
# =====
from google.colab import files
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import mean_squared_error, mean_absolute_error

import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, GRU, Dense, Dropout

print(tf.__version__)
# =====
# 2. Завантаження CSV з цінами Bitcoin
# =====
uploaded = files.upload()
file_name = list(uploaded.keys())[0]

df = pd.read_csv(file_name)

# Перетворення дати і сортування
df['Date'] = pd.to_datetime(df['Date'])
df.sort_values('Date', inplace=True)

# Використовуємо тільки ціну закриття
data = df[['Date', 'Close']].copy()
data.set_index('Date', inplace=True)

print("Перші рядки даних:")
data.head()

# =====
# 3. Масштабування даних (0..1)
# =====
scaler = MinMaxScaler(feature_range=(0, 1))
scaled_values = scaler.fit_transform(data[['Close]])

print("Мінімум після масштабування:", scaled_values.min())
print("Максимум після масштабування:", scaled_values.max())

# =====
# 4. Формування вибірки (послідовностей) для LSTM/GRU
# n_steps = довжина вікна

```

```

# =====
n_steps = 60

X = []
y = []

for i in range(n_steps, len(scaled_values)):
    X.append(scaled_values[i-n_steps:i, 0])
    y.append(scaled_values[i, 0])

X = np.array(X)
y = np.array(y)

# Перетворюємо в формат [samples, timesteps, features]
X = X.reshape((X.shape[0], X.shape[1], 1))

print("Форма X:", X.shape)
print("Форма y:", y.shape)

# =====
# 5. Train / Test split
# =====
train_size = int(len(X) * 0.8)

X_train, X_test = X[:train_size], X[train_size:]
y_train, y_test = y[:train_size], y[train_size:]

print("Розмір train:", X_train.shape, y_train.shape)
print("Розмір test :", X_test.shape, y_test.shape)

# =====
# 6. Побудова LSTM-моделі
# =====
lstm_model = Sequential()
lstm_model.add(LSTM(50, return_sequences=True, input_shape=(n_steps, 1)))
lstm_model.add(Dropout(0.2))
lstm_model.add(LSTM(50))
lstm_model.add(Dropout(0.2))
lstm_model.add(Dense(1)) # прогноз наступного значення

lstm_model.compile(optimizer='adam', loss='mse')

print("LSTM модель:")
lstm_model.summary()
# =====
# 7. Навчання LSTM
# =====
history_lstm = lstm_model.fit(
    X_train, y_train,
    epochs=20,
    batch_size=32,

```

```

validation_split=0.1,
verbose=1
)

# Графік функції втрат LSTM
plt.figure(figsize=(8,4))
plt.plot(history_lstm.history['loss'], label='Train loss')
plt.plot(history_lstm.history['val_loss'], label='Val loss')
plt.title('LSTM: динаміка функції втрат')
plt.xlabel('Епоха')
plt.ylabel('MSE')
plt.legend()
plt.grid(True)
plt.show()

# =====
# 8. Побудова GRU-моделі
# =====

gru_model = Sequential()
gru_model.add(GRU(50, return_sequences=True, input_shape=(n_steps, 1)))
gru_model.add(Dropout(0.2))
gru_model.add(GRU(50))
gru_model.add(Dropout(0.2))
gru_model.add(Dense(1))

gru_model.compile(optimizer='adam', loss='mse')

print("GRU модель:")
gru_model.summary()

# =====
# 9. Навчання GRU
# =====

history_gru = gru_model.fit(
    X_train, y_train,
    epochs=20,
    batch_size=32,
    validation_split=0.1,
    verbose=1
)

plt.figure(figsize=(8,4))
plt.plot(history_gru.history['loss'], label='Train loss')
plt.plot(history_gru.history['val_loss'], label='Val loss')
plt.title('GRU: динаміка функції втрат')
plt.xlabel('Епоха')
plt.ylabel('MSE')
plt.legend()
plt.grid(True)
plt.show()

```

```

# =====
# 10. Прогнозування на тестовій вибірці
# =====
y_pred_lstm = lstm_model.predict(X_test)
y_pred_gru = gru_model.predict(X_test)

# Зворотне масштабування до реальних цін
y_test_inv = scaler.inverse_transform(y_test.reshape(-1,1))
y_pred_lstm_inv = scaler.inverse_transform(y_pred_lstm)
y_pred_gru_inv = scaler.inverse_transform(y_pred_gru)

# Оцінка якості
mse_lstm = mean_squared_error(y_test_inv, y_pred_lstm_inv)
mae_lstm = mean_absolute_error(y_test_inv, y_pred_lstm_inv)

mse_gru = mean_squared_error(y_test_inv, y_pred_gru_inv)
mae_gru = mean_absolute_error(y_test_inv, y_pred_gru_inv)

print("LSTM: MSE =", mse_lstm, " MAE =", mae_lstm)
print("GRU : MSE =", mse_gru, " MAE =", mae_gru)

# =====
# 11. Побудова графіка: фактичні vs прогнози
# =====
# Відновлюємо дати для тестової частини
dates = data.index[n_steps:] # дати, які відповідають у
test_dates = dates[train_size:] # тільки тестові дати

plt.figure(figsize=(14,6))
plt.plot(test_dates, y_test_inv, label='Фактичні значення')
plt.plot(test_dates, y_pred_lstm_inv, label='Прогноз LSTM')
plt.plot(test_dates, y_pred_gru_inv, label='Прогноз GRU')
plt.title('Прогнозування ціни Bitcoin за допомогою LSTM та GRU')
plt.xlabel('Дата')
plt.ylabel('Ціна закриття')
plt.legend()
plt.grid(True)
plt.show()

```

Дохідність.ірунв

```

# =====
# 1. Імпорт бібліотек
# =====
from google.colab import files
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# Для графіків
plt.style.use('seaborn-v0_8')

# =====
# 2. Завантаження CSV
# =====
uploaded = files.upload()
file_name = list(uploaded.keys())[0]

df = pd.read_csv(file_name)
df['Date'] = pd.to_datetime(df['Date'])
df.sort_values('Date', inplace=True)

# Перевірка структури
df.head()

# =====
# 3. Обчислення Log Returns
#  $r_t = \ln(\text{Close}_t / \text{Close}_{(t-1)})$ 
# =====
df['Log_Returns'] = np.log(df['Close'] / df['Close'].shift(1))

# Видаляємо перший NaN
df = df.dropna()

df[['Date', 'Close', 'Log_Returns']].head()

# =====
# 4. Статистика log-returns
# =====
mean_lr = df['Log_Returns'].mean()
std_lr = df['Log_Returns'].std()
min_lr = df['Log_Returns'].min()
max_lr = df['Log_Returns'].max()
median_lr = df['Log_Returns'].median()

stats = pd.DataFrame({
    'Mean': [mean_lr],
    'Std Dev': [std_lr],
    'Min': [min_lr],
    'Max': [max_lr],
})

```

```

    'Median': [median_lr]
})

print("==== Статистика Log Returns ====")
stats

# =====
# 5. Побудова графіка Log Returns
# =====
plt.figure(figsize=(14,6))
plt.plot(df['Date'], df['Log_Returns'], color='blue')
plt.title("Графік Log>Returns Bitcoin")
plt.xlabel("Дата")
plt.ylabel("Log Return")
plt.grid(True)
plt.show()

# =====
# 6. Гістограма Log Returns
# =====
plt.figure(figsize=(10,5))
plt.hist(df['Log_Returns'], bins=50, color='green', alpha=0.7)
plt.title("Гістограма розподілу Log>Returns Bitcoin")
plt.xlabel("Log Return")
plt.ylabel("Кількість спостережень")
plt.grid(True)
plt.show()

# =====
# 7. Кумулятивна дохідність
# =====
df['Cumulative_Returns'] = df['Log_Returns'].cumsum()

plt.figure(figsize=(14,6))
plt.plot(df['Date'], df['Cumulative_Returns'], color='red')
plt.title("Кумулятивна лог-дохідність Bitcoin")
plt.xlabel("Дата")
plt.ylabel("Сума Log Returns")
plt.grid(True)
plt.show()

```

Додаток 3

Value at Risk (VaR) .ipynb

```

# =====
# 1. Імпорт бібліотек
# =====
from google.colab import files
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

plt.style.use('seaborn-v0_8')

# =====
# 2. Завантаження даних
# =====
uploaded = files.upload()
file_name = list(uploaded.keys())[0]

df = pd.read_csv(file_name)
df['Date'] = pd.to_datetime(df['Date'])
df.sort_values('Date', inplace=True)

df.head()

# =====
# 3. Обчислення log-returns
# =====
df['Log_Returns'] = np.log(df['Close'] / df['Close'].shift(1))
df = df.dropna()

df[['Date', 'Close', 'Log_Returns']].head()

# =====
# 4. Історичний VaR (повернення)
# =====

alpha_95 = 0.95
alpha_99 = 0.99

# Для втрат беремо нижні квантілі розподілу log-returns
VaR_95 = np.quantile(df['Log_Returns'], 1 - alpha_95)
VaR_99 = np.quantile(df['Log_Returns'], 1 - alpha_99)

print("==== Історичний VaR у термінах log-returns (щоденний) ====")
print(f"VaR 95%: {VaR_95:.6f}")
print(f"VaR 99%: {VaR_99:.6f}")

# =====
# 5. VaR у грошовому вираженні
# =====

portfolio_value = 10000

```

```

VaR_95_money = -VaR_95 * portfolio_value
VaR_99_money = -VaR_99 * portfolio_value

print("==== Історичний VaR для портфеля", portfolio_value, "USD ====")
print(f"VaR 95%: {VaR_95_money:.2f} USD")
print(f"VaR 99%: {VaR_99_money:.2f} USD")

# =====
# 6. Гістограма log-returns з лініями VaR
# =====
plt.figure(figsize=(10,5))
plt.hist(df['Log_Returns'], bins=50, alpha=0.7, color='skyblue', edgecolor='black')
plt.axvline(VaR_95, color='red', linestyle='--', linewidth=2, label=f"VaR 95% = {VaR_95:.3f}")
plt.axvline(VaR_99, color='darkred', linestyle='--', linewidth=2, label=f"VaR 99% = {VaR_99:.3f}")
plt.title("Гістограма Log>Returns Bitcoin з лініями VaR")
plt.xlabel("Log Return")
plt.ylabel("Кількість спостережень")
plt.legend()
plt.grid(True)
plt.show()

# =====
# 7. Ковзний (rolling) VaR 95% (вікно 100 днів)
# =====

window = 100

rolling_VaR_95 = df['Log_Returns'].rolling(window).quantile(1 - alpha_95)

plt.figure(figsize=(14,6))
plt.plot(df['Date'], df['Log_Returns'], alpha=0.3, label='Log Returns')
plt.plot(df['Date'], rolling_VaR_95, color='red', label=f"Rolling VaR 95% (вікно {window} днів)")
plt.title("Ковзний історичний VaR 95% для Bitcoin")
plt.xlabel("Дата")
plt.ylabel("Log Return / VaR")
plt.legend()
plt.grid(True)
plt.show()

```

Додаток И

Кореляційний аналіз між різними криптовалютами .ірунв

```

# =====
# 1. Імпорт бібліотек
# =====
from google.colab import files
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

plt.style.use('seaborn-v0_8')

# =====
# 2. Завантажуємо файли
# =====
uploaded = files.upload()
uploaded

# =====
# 3. Зчитуємо кожен файл, беремо Date та Close,
# називаємо колонку по Symbol (BTC, ETH, LTC)
# =====
crypto_dfs = []

for fname in uploaded.keys():
    df_tmp = pd.read_csv(fname)

    # Переконаємось, що є потрібні колонки
    # (у твоїх файлах: SNo, Name, Symbol, Date, High, Low, Open, Close, Volume, Marketcap)
    df_tmp['Date'] = pd.to_datetime(df_tmp['Date'])
    df_tmp.sort_values('Date', inplace=True)

    symbol = df_tmp['Symbol'].iloc[0] # BTC / ETH / LTC

    df_tmp = df_tmp[['Date', 'Close']]
    df_tmp.rename(columns={'Close': symbol}, inplace=True)

    crypto_dfs.append(df_tmp)

# Об'єднуємо всі три датафрейми по даті (перетин дат)
from functools import reduce
merged = reduce(lambda left, right: pd.merge(left, right, on='Date', how='inner'), crypto_dfs)

merged.head()

# =====
# 4. Кореляційна матриця для цін (Close)
# =====
price_cols = [col for col in merged.columns if col != 'Date']

```

```

corr_prices = merged[price_cols].corr()

print("==== Кореляційна матриця цін (Close) ====")
corr_prices
# Heatmap для кореляції цін
plt.figure(figsize=(6,5))
plt.imshow(corr_prices, cmap='coolwarm', vmin=-1, vmax=1)
plt.colorbar(label='Коефіцієнт кореляції')
plt.xticks(range(len(price_cols)), price_cols)
plt.yticks(range(len(price_cols)), price_cols)
plt.title("Кореляційна матриця цін (Close) BTC, ADA, XLM")
for i in range(len(price_cols)):
    for j in range(len(price_cols)):
        plt.text(j, i, f'{corr_prices.iloc[i, j]:.2f}',
                ha='center', va='center', color='black')
plt.tight_layout()
plt.show()
# =====
# 5. Log-returns для BTC, ETH, LTC
# =====
log_ret = merged.copy()
for col in price_cols:
    log_ret[col] = np.log(log_ret[col] / log_ret[col].shift(1))
log_ret = log_ret.dropna()
log_ret.head()

# =====
# 6. Кореляційна матриця для log-returns
# =====
corr_lr = log_ret[price_cols].corr()
print("==== Кореляційна матриця log-returns ====")
corr_lr
# Heatmap для кореляції log-returns
plt.figure(figsize=(6,5))
plt.imshow(corr_lr, cmap='coolwarm', vmin=-1, vmax=1)
plt.colorbar(label='Коефіцієнт кореляції')
plt.xticks(range(len(price_cols)), price_cols)
plt.yticks(range(len(price_cols)), price_cols)
plt.title("Кореляційна матриця log-returns BTC, ADA, XLM")
for i in range(len(price_cols)):
    for j in range(len(price_cols)):
        plt.text(j, i, f'{corr_lr.iloc[i, j]:.2f}',
                ha='center', va='center', color='black')
plt.tight_layout()
plt.show()
stats_lr = log_ret[price_cols].describe().T
print("==== Описова статистика log-returns BTC, ETH, LTC ====")
stats_lr

```