

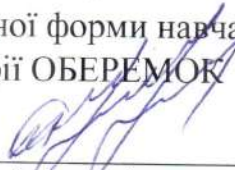
**МІНІСТЕРСТВО ОСВІТИ І НАУКИ
УКРАЇНИ КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ
УНІВЕРСИТЕТ ІМЕНІ ТАРАСА
ШЕВЧЕНКА**

Економічний факультет

Кафедра статистики, інформаційно-аналітичних систем і демографії

**КВАЛІФІКАЦІЙНА МАГІСТЕРСЬКА РОБОТА
РОЗРОБКА МОДЕЛІ КОРОТКОСТРОКОВОГО ПРОГНОЗУВАННЯ
КОТИРУВАНЬ КУРСІВ КРИПТОВАЛЮТ**

Студентки II курсу
спеціальності 051 Економіка
за освітньою програмою
«Економічна аналітика та
статистика»
денної форми навчання
Марії ОБЕРЕМОК



(підпис)

Науковий керівник:
к.е.н., доцент
Надія ІВАНЧЕНКО



(підпис)

Роботу допущено до захисту на засіданні ЕК рішенням кафедри
статистики, інформаційно-аналітичних систем і демографії,
протокол №10 від «04» травня 2023 року

Завідувач кафедри, д.е.н., проф.


Наталія КОВТУН

Київ – 2023

РЕФЕРАТ

Кваліфікаційна робота магістра містить 65 с., 34 рис., 2 табл., 28 джерел, 1 додаток.

Ключові слова: криптовалюта, курси валют, короткострокове прогнозування, часові ряди, ARIMA, мова програмування Python, Bitcoin.

Предмет дослідження: методи прогнозування курсів криптовалют на основі часових рядів.

Об'єкт дослідження: ринок криптовалют, часовий ряд курсу Bitcoin.

Мета дослідження: створення моделі короткострокового прогнозування котирувань курсів криптовалют.

Методи дослідження: графічний аналіз, аналіз часових рядів, метод згладжування ковзним середнім, інтегрована модель авторегресії ковзного середнього ARIMA.

Наукова новизна, практична значимість дослідження: у роботі було побудовано та проаналізовано статистичні моделі прогнозування часових рядів. Зібрано теоретичний матеріал щодо факторів, що впливають на зміну котирувань курсів криптовалют, удосконалено методологію короткострокового прогнозування курсів криптовалют за стандартом CRISP-DM (Cross-Industry Standard Process for Data Mining).

Практична цінність: побудована статистична модель може бути використана інвесторами та менеджерами з ризиків, які шукають інструменти для прогнозування цін на криптовалюти.

RESUME

Taras Shevchenko National University of Kyiv

Faculty of Economics, Department of Statistics and Demography

Key words: cryptocurrency, exchange rates, short-term forecasting, time series, ARIMA, Python, Bitcoin..

The purpose of graduation research of student Mariia Oberemok's "Development of a model for short-term forecasting of cryptocurrencies exchange rate" is creating a model for short-term forecasting of cryptocurrency exchange rates.

The author created a model for short-term forecasting of the exchange rate of cryptocurrencies. Collected theoretical material on the factors affecting the change in quotes of cryptocurrency rates and improved the methodology of short-term forecasting of cryptocurrency rates according to the CRISP-DM (Cross-Industry Standard Process for Data Mining) standard.

To construct optimal model are used graphical analysis, time series analysis, moving average smoothing method, and integrated ARIMA moving average autoregression model.

Research contains: 65 pages, 34 fig., 2 tables, 28 bibliog., 1 append.

ЗМІСТ

ВСТУП	5
РОЗДІЛ І. ДОСЛІДЖЕННЯ ТЕОРЕТИЧНИХ ЗАСАД ФУНКЦІОНУВАННЯ РИНКУ КРИПТОВАЛЮТ	7
1.1 Поняття криптовалюти	7
1.2 Фактори, що впливають на коливання курсів криптовалют	11
РОЗДІЛ 2. МЕТОДИ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ	15
2.1 Методологія короткострокового прогнозування котирувань курсів криптовалют	15
2.2 Статистичні методи для прогнозування обмінних курсів криптовалют	18
РОЗДІЛ 3. КОРОТКОСТРОКОВЕ ПРОГНОЗУВАННЯ КОТИРУВАНЬ КУРСУ КРИПТОВАЛЮТ	20
3.1 Опис бази даних дослідження курсу криптовалют.....	20
3.2 Побудова моделі для прогнозування	24
3.3 Оцінка отриманих результатів та практичне застосування.....	32
ВИСНОВКИ	48
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	50
ДОДАТКИ	53

ВСТУП

За останні роки криптовалюти набули величезної популярності. Децентралізована природа криптовалют дозволяє використовувати їх для багатьох цілей, починаючи від онлайн-покупок і закінчуючи міжнародними грошовими переказами. Крім того, технологія блокчейн, яка лежить в основі багатьох криптовалют, має потенціал для революції в різних галузях, включаючи фінанси, охорону здоров'я та управління ланцюгами поставок.

Прогнозування курсів криптовалют є надзвичайно важливим, адже це дозволяє виявити закономірності та встановити тенденції розвитку всього ринку. Своєчасна та точна інформація про короткострокові коливання валютного курсу дозволяє учасникам ринку впроваджувати торгові стратегії та керувати ризиками.

Актуальність статистичного аналізу та побудови моделей прогнозування курсів криптовалют обумовлені зростанням попиту на точні та надійні моделі прогнозування курсів криптовалют.

Мета даної роботи – створення моделі короткострокового прогнозування курсів криптовалют.

Об'єктом дослідження є ринок криптовалют у цілому.

Предметом дослідження є сукупність теоретичних, методологічно-організаційних та практичних аспектів прогнозування курсів криптовалют на основі часових рядів.

Основні завданнями, що вирішуються в роботі:

- визначення теоретичних засад функціонування ринку криптовалют;
- дослідження основних факторів, що впливають на коливання курсів криптовалют;
- визначення методології короткострокового прогнозування котирувань курсів криптовалют;
- визначення статистичних методів для прогнозування;
- визначення бази дослідження, інструментів розробки, аналізу та візуалізації;

- оцінювання сформованих часових рядів;
- побудова та порівняння статистичних моделей;
- визначення кращої моделі короткострокового прогнозування котирувань курсів криптовалют;
- оцінка отриманих результатів.

Дослідженням у сфері прогнозування курсів криптовалют займалися наступні вчені: М. Касі [19], П. Броквел [18], П. Бідюк [1,2], В. Дербенцев [6], Г. Великоіваненко [6], Н. Даценко [4, 5], та інші. Загалом ці вчені зробили значний внесок у сферу прогнозування курсів криптовалют завдяки своїм дослідженням.

Статистичне дослідження здійснювалось із використанням даних ресурсу Yahoo Finance [14]. З урахуванням необхідності використання широкого спектру математичних обчислень, а також використання бібліотек машинного навчання та реалізації окремих моделей, у тому числі ARIMA моделі, мовою програмування було обрано – Python, версія 3.11.

Результати дослідження були опубліковані у вигляді наукової статті на тему «Методологія короткострокового прогнозування котирувань курсів криптовалют» в Науковому віснику Полтавського університету економіки і торгівлі, серія «Економічні науки», випуск 2 (108) 2023.

Магістерська кваліфікаційна робота складається із вступу, трьох розділів, висновків, списку використаних джерел та додатку.

1 ДОСЛІДЖЕННЯ ТЕОРЕТИЧНИХ ЗАСАД ФУНКЦІОНУВАННЯ РИНКУ КРИПТОВАЛЮТ

1.1 Поняття криптовалюти

Криптовалюти набули величезної популярності у фінансовій індустрії, і їх важливість неможливо переоцінити. Це віртуальні валюти, які використовують криптографію для безпеки та є децентралізованими, вільними від контролю будь-якого центрального банку чи фінансової установи.

Загальна ринкова капіталізація ринку криптовалют у листопаді 2021 року досягла історичного максимуму в понад 2,9 трильйона доларів США порівняно з лише 19,7 мільярда доларів США в січні 2017 року. Щоденний обсяг торгів криптовалютами перевищив 300 мільярдів доларів у травні 2021 року порівняно з приблизно 1 мільярдом доларів у 2013 році [9].



Рис. 1.1 Загальна ринкова капіталізація криптовалюти
Джерело: створено ресурсом <https://www.coindesk.com/> [9]

Перша криптовалюта, Bitcoin, була створена в 2009 році, через деякий час були створені численні інші криптовалюти, такі як Ethereum, Ripple і

Litecoin. Хоча всі вони мають певні характеристики, кожна криптовалюта унікальна та має власний набір можливостей і функцій.



Рис. 1.2 Розподіл криптовалют за часткою ринкової капіталізації
Джерело: створено ресурсом <https://www.coingecko.com/> [11]

Криптовалюти відомі своєю високою волатильністю, коливання цін на деякі з них перевищує 100% протягом одного дня. Наприклад, у квітні 2021 року ціна Bitcoin досягла історичного максимуму в понад 64 000 доларів США, а протягом двох місяців впала нижче 30 000 доларів [9].

Серед найпопулярніших криптовалют можна виділити наступні:

1. Bitcoin (BTC) — це перша і найвідоміша криптовалюта. Bitcoin розроблений як цифрова валюта, яка забезпечує безпечні та прозорі транзакції без посередників. З обмеженою пропозицією в 21 мільйон монет Bitcoin отримав визнання як засіб збереження вартості та засіб обміну [8, 9].
2. Ethereum (ETH) - є платформою на основі блокчейну, яка полегшує розробку децентралізованих програм (DApps) і смарт-контрактів. Вона забезпечує програмоване середовище, де розробники можуть створювати та розгортати програми, не покладаючись на централізований орган. Криптовалюта Ethereum, Ether (ETH), використовується для забезпечення транзакцій і виконання смарт-контрактів у мережі [12].
3. Binance Coin (BNB)— це рідна криптовалюта біржі Binance, однієї з

найбільших бірж криптовалют у світі. Створена у 2017 році BNB працює на Binance Smart Chain (BSC). BNB можна використовувати для оплати комісії за транзакції на платформі Binance і для доступу до різних сервісів в екосистемі Binance.

4. Ripple (XRP) представлений у 2012 році, є одночасно цифровим платіжним протоколом і криптовалютою. Його мета – швидкі, недорогі міжнародні грошові перекази та покращення ліквідності у фінансовій галузі. Ripple співпрацює з банками та фінансовими установами. XRP, рідна криптовалюта мережі Ripple, використовується для полегшення транзакцій і забезпечення ліквідності на платформі.
5. Cardano (ADA) — це блокчейн-платформа, запущена в 2015 році. Cardano використовує механізм консенсусу proof-of-stake (PoS) під назвою Ouroboros і прагне забезпечити платформу для створення децентралізованих програм.

Таблиця 1.1 – Перелік найбільш популярних криптовалют

Криптовалюта	Рік створення	Ринкова капіталізація, % Станом на 07.05.2023	Ключові особливості
Bitcoin (BTC)	2009	45,24	- Перша та найвідоміша криптовалюта - Обмежена пропозиція (21 мільйон монет)
Ethereum (ETH)	2015	18,48	- Друга за величиною криптовалюта за ринковою капіталізацією
Binance Coin (BNB)	2017	4,12	- Власний токен біржі Binance - Створено на основі Binance Smart Chain (BSC)
Ripple (XRP)	2012	1,93	- Розроблений для швидких і недорогих міжнародних грошових переказів - Співпраця з фінансовими установами

Продовження таблиці 1.1

Cardano (ADA)	2015	1,08	- Блокчейн-платформа з акцентом на безпеку та сталість
---------------	------	------	--

Джерело: побудовано автором на основі даних <https://www.coindesk.com/> [9]

Криптовалюти пропонують кілька переваг перед традиційними методами оплати, зокрема швидші транзакції, нижчі комісії за транзакції та більшу прозорість. Крім того, криптовалютами може користуватися кожен, хто має доступ до Інтернету, що робить їх доступними для широкого кола людей у всьому світі.

Незважаючи на свої переваги, криптовалюти також мають ряд проблем і ризиків. Однією з найбільших проблем є їх висока волатильність, яка може призвести до значних коливань цін протягом короткого періоду. Ця волатильність робить криптовалюти ризикованими для інвесторів, а також обмежує їх використання як надійного засобу збереження вартості або засобу обміну.

Ще одна проблема— це питання регулювання. Оскільки вони не контролюються жодним центральним органом влади, криптовалюти часто використовуються для незаконної діяльності, відмивання грошей, ухилення від сплати податків. Уряди та фінансові установи в усьому світі намагаються створити нормативну базу, яка могла б ефективно вирішувати ці проблеми, водночас дозволяючи криптовалютам рости та процвітати.

Незважаючи на ці проблеми, популярність криптовалют продовжує зростати, і вони все частіше використовуються як методи оплати та інвестицій. Як наслідок, зростає попит на точні та надійні моделі прогнозування для прогнозування курсів криптовалюти.

1.2 Фактори, що впливають на коливання курсів криптовалют

Обмінні курси криптовалют залежать від низки факторів, зокрема економічних, технологічних, регуляторних і ринкових. Серед ключових факторів, які впливають на розрахунок курсу обміну криптовалют, можна виділити наступні (рис. 1.3):

Ринкові фактори	Нормативно-правові фактори:	Технологічні фактори:	Економічні фактори
Ринковий попит і пропозиція Настрої інвесторів Ринкова ліквідність Висвітлення в ЗМІ Інтеграція та прийняття Конкуренція Безпека мережі Маніпулювання ринком	Регуляторне середовище	Технологічні розробки	Економічні фактори (включаючи рівень інфляції, процентні ставки, економічну стабільність і геополітичні події)

Рис. 1.3 Узагальнююча таблиця факторів, що впливають на коливання курсів криптовалют

Джерело: створено автором

Як і традиційні валюти, курси криптовалют залежать від попиту та пропозиції. Пропозиція криптовалюти визначається швидкістю майнінгу та загальним лімітом пропозиції. Швидкість майнінгу означає швидкість, з якою генеруються нові одиниці криптовалюти, а загальний ліміт пропозиції — це максимальна кількість одиниць, яку можна створити. Якщо швидкість майнінгу висока, а загальний ліміт пропозиції низький, пропозиція криптовалюти буде обмеженою, що може призвести до зростання її ціни[20].

Максимальна кількість Bitcoin становить 21 мільйон одиниць. На 12 травня 2021 року було доступно 18,7 мільйонів монет, і як тільки кількість монет буде становити 21 мільйон одиниць, ціна може зрости через обмежену пропозицію [8].

На відміну від цього, Ethereum, друга за величиною криптовалюта, не має обмежень щодо пропозиції.

З іншого боку, попит на криптовалюту залежить від ще інших факторів, включаючи сприйняту цінність криптовалюти, швидкість впровадження та рівень підтримки криптовалюти. Наприклад, засновник Tesla Ілон Маск на початку 2021 року оголосив, що його компанія з виробництва електромобілів придбала Bitcoin на 1,5 мільярда доларів. Також, було додано опцію покупки електромобілей його компанії за криптовалюту. Це спричинило значне зростання попиту на Bitcoin.

Ринковий настрій передбачає загальний настрій або ставлення інвесторів до певної криптовалюти чи ринку в цілому. Позитивні ринкові настрої можуть призвести до збільшення попиту на криптовалюту, тоді як негативні ринкові настрої можуть спричинити зниження попиту.

На настрої можуть впливати фактичні фундаментальні показники криптовалюти. Інвесторів привабили Bitcoin та Ethereum, оскільки вони першими запропонували децентралізовані способи переказу грошей і створили технологію, яка допомагає зберігати та відстежувати транзакції.

На ринкові настрої також впливає кілька факторів, таких як події новин та соціальні мережі. Наприклад, якщо з'являться негативні новини щодо певної криптовалюти, наприклад порушення системи безпеки, настрої ринку щодо цієї криптовалюти, ймовірно, стануть негативними, що може спричинити падіння її ціни.

Немає послідовного глобального регулювання сектору криптовалют. Однією з причин є те, що криптовалюти не базуються в одному місці, тому національні органи влади, такі як фінансові регулятори та центральні банки, повинні встановлювати власні правила щодо цифрової готівки. Характер і обсяг цих правил можуть мати значний вплив на курси криптовалют. Наприклад, якщо країна заборонить використання криптовалют, це негативно вплине на обмінні курси. І навпаки, якби країна прийняла закони, які заохочують впровадження криптовалют, це могло б мати позитивний вплив на обмінні курси.

Технологічні інновації є одним з найважливіших факторів, які впливають на коливання курсів криптовалют. Нові технологічні рішення додають додаткову функціональність та поліпшують ефективність криптовалютних платформ, що привертає більше інвесторів та користувачів. Розширення можливостей блокчейну та впровадження смарт-контрактів є прикладами таких інновацій.

Наприклад, платформа Ethereum, завдяки впровадженню смарт-контрактів, стала привабливішою для децентралізованих додатків (DApps) та розумних контрактів. Це привело до зростання популярності Ethereum та підвищення попиту на його криптовалюту. Такі технологічні інновації надають користувачам більше можливостей і забезпечують широке використання криптовалют, що може позитивно вплинути на їх курси.

Цифрові валюти пов'язані з певним рівнем ризику, пов'язаним з кіберзлочинністю та порушеннями безпеки. Однак, розвиток нових технологій та заходів безпеки допомагає зменшити ці ризики і забезпечує більшу довіру до криптовалют. Наприклад, вдосконалені системи шифрування та захисту даних можуть залучати більше інвесторів, що може позитивно вплинути на попит та курси криптовалют.

Одним з додаткових факторів, що впливає на коливання курсів криптовалют, - це спекулятивна природа цього ринку. Багато інвесторів та трейдерів сприймають криптовалюту як інструмент для отримання прибутку від короткострокових торговельних операцій. Це може призводити до швидких змін курсів, оскільки великі обсяги купівлі та продажу впливають на ринкову ціну.

Крім того, курси криптовалют можуть бути під впливом маніпуляцій та фейкових новин. У світі криптовалют існує поняття "памп" і "дамп", коли група інвесторів штучно збільшує або занижує ціну певної криптовалюти, щоб отримати прибуток за рахунок інших учасників ринку. Такі маніпуляції можуть значно впливати на курси криптовалют і створювати нестабільність.

У кінцевому підсумку, коливання курсів криптовалют є комплексним явищем, що залежить від різних факторів, таких як економічні, технологічні, регуляторні та ринкові. Пропозиція та попит на криптовалюту, їх технологічний

прогрес, ринкові настрої, новини та геополітичні події — усе це має вплив на коливання курсів криптовалют.

2 МЕТОДИ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ

2.1 Методологія короткострокового прогнозування котирувань курсів криптовалют

Для створення методології короткострокового прогнозування курсів криптовалют було використано стандарт CRISP-DM (Cross-Industry Standard Process for Data Mining) - популярна методологія, яка використовується для проектів інтелектуального аналізу даних і машинного навчання [23]. Вона складається з шести основних етапів: розуміння бізнесу, розуміння даних, підготовка даних, моделювання, оцінка та розгортання. Цю методологію можна адаптувати для розробки методології короткострокового прогнозування курсів криптовалют наступним чином (рис. 2.1):

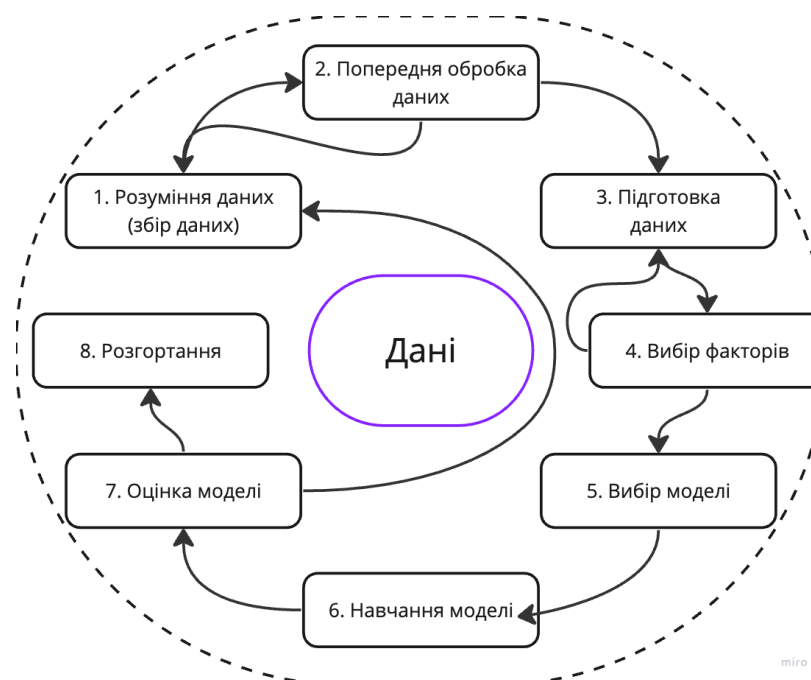


Рис. 2.1 - Адаптована схема для дослідження згідно стандарту CRISP-DM

Джерело: побудовано автором

1. Розуміння даних - на цьому етапі визначаються та збираються джерела та типи даних, включаючи історичні дані про ціну криптовалюти, що аналізується, дані про обсяги та іншу відповідну інформацію, яку можна використовувати для навчання та тестування моделей прогнозування. Цей

етап включає дослідження даних і оцінку їх якості. У якості джерел інформації можуть виступати біржі криптовалют, фінансові веб-сайти та API.

2. Попередня обробка даних, щоб усунути будь-які викиди, прогалини або помилки в даних. Це можна зробити за допомогою різних статистичних методів, включаючи інтерполяцію, регресійний аналіз і згладжування даних.
3. Підготовка даних: на цьому етапі дані очищаються, трансформуються та попередньо обробляються для створення набору даних, який можна використовувати для моделювання.
4. Вибір факторів: наступним кроком є вибір відповідних факторів, які можуть вплинути на вартість аналізованої криптовалюти. Це включає такі фактори, як обсяг торгів, ринкові настрої, мережева активність і технічні індикатори.
5. Вибір моделі: наступним кроком є вибір моделі для короткострокового прогнозування курсу криптовалюти. Найбільш релевантними з точки зору прогнозування упорядкованих в часі даних є моделі часових рядів, регресійні моделі.
6. Навчання моделі на історичних даних про ціни та вибраних факторів. Цей етап передбачає навчання моделі на відібраних даних та оптимізацію параметрів моделі за допомогою різних методів.
7. Оцінка моделі: наступним етапом є оцінка продуктивності навченої моделі на основі даних тестування. Це передбачає обчислення різних показників, включаючи середню абсолютну похибку та середню квадратичну похибку.
8. Розгортання: на цьому етапі вибрана модель розгортається для створення короткострокових прогнозів обмінних курсів криптовалюти. Цей етап передбачає розробку програми або програмного засобу, який може приймати нові дані та надавати оновлені прогнози.

Адаптована методологія короткострокового прогнозування обмінних

курсів криптовалют є вичерпною та охоплює всі необхідні етапи процесу прогнозування, від розуміння даних і попередньої обробки до вибору моделі, навчання, оцінки та розгортання. Вона забезпечує комплексну основу для короткострокового прогнозування курсів криптовалюти, підвищує якість прогнозів і надає цінну інформацію про ринок криптовалюти.

2.2 Статистичні методи для прогнозування обмінних курсів криптовалют

Одним із найбільш використовуваних статистичних методів для прогнозування обмінних курсів криптовалют є аналіз часових рядів. Аналіз часових рядів передбачає аналіз точок даних, зібраних протягом певного часу, щоб визначити закономірності та тенденції. Аналіз часових рядів дозволяє ідентифікувати закономірності в історичних даних, які потім можна використовувати для прогнозування майбутніх коливань обмінного курсу.

Існує кілька моделей, які можна використовувати для прогнозування криптовалюти, включаючи авторегресійну модель (AR), модель авторегресії та ковзної середньої (ARMA) та модель авторегресійної інтегрованої ковзної середньої (ARIMA) [2].

Таблиця 2.1 – Статистичні моделі для прогнозування

Модель	Опис	Використання
Авторегресійна модель (AR)	Модель, що враховує залежність між попередніми значеннями курсу та майбутніми значеннями з деяким часовим лагом.	Аналіз та прогнозування трендів та залежностей між попередніми та майбутніми значеннями курсу.
Модель авторегресії та ковзного середнього (ARMA)	Модель, яка поєднує в собі авторегресійну складову (AR) та складову ковзного середнього (MA).	Опис статистичних властивостей даних та прогнозування на основі залежностей між попередніми значеннями та випадковими шоками.
Модель авторегресійної інтегрованої ковзної середньої (ARIMA)	Модель, що враховує тренд, сезонність та автокореляцію. Базується на авторегресійній складовій (AR), моделі ковзного середнього (MA) та інтегруванні ряду даних.	Короткострокові прогнози, коли історичні дані доступні та враховуються тренд, сезонність та автокореляція.

Джерело: побудовано автором

Авторегресійна модель (AR) враховує, що майбутні значення курсу криптовалюти залежать від попередніх значень з певним часовим лагом. Перевагами цієї моделі є її простота, легкість інтерпретації та здатність виявляти та прогнозувати тренди. Проте, AR-модель не враховує сезонності та інших складні структури в даних, що може обмежувати її точність в прогнозуванні.

Модель авторегресії та ковзного середнього (ARMA) поєднує авторегресійні та ковзні середні компоненти, що дозволяє враховувати як тренд, так і шум в даних. ARMA-модель може бути ефективною в прогнозуванні криптовалютних курсів, якщо вони мають явно виражену автокореляцію та стаціонарність. Проте, недоліком цієї моделі є складність визначення оптимальних параметрів, особливо для складних часових рядів, а також вимога до стаціонарності даних.

Модель ARIMA є популярним методом для аналізу часових рядів, оскільки вона може обробляти як стаціонарні, так і нестаціонарні дані часових рядів. Точність прогнозів моделі ARIMA можна оцінити за допомогою таких показників, як середня абсолютна помилка (MAE) і середня квадратична помилка (RMSE). Серед переваг можна виділити врахування тренду, сезонності, автокореляції даних, можливість як короткострокового прогнозу, так і довгострокового. Серед недоліків можна виділити необхідність достатньої кількості історичних даних для побудови моделі.

3 КОРОТКОСТРОКОВЕ ПРОНОЗУВАННЯ КОТИРУВАНЬ КУРСУ КРИПТОВАЛЮТИ

3.2. Опис бази дослідження курсу криптовалют

Пошук і отримання інформації щодо поточного курсу криптовалют є доволі простим, ця інформація розміщена у вільному доступі на багатьох сайтах, наприклад, Coindesk або Finance.ua, рис 3.1 та 3.2.



Рисунок 3.1 – Котирування BTC

Джерело: створено ресурсом <https://www.coindesk.com/> [9]

Проблема таких ресурсів полягає в тому, що користувачу доступні якісні представлення котирувань криптовалют з вибором діапазону дат та інших параметрів, але відсутня можливість отримання даних у вигляді файлу даних для подальшого аналізу. З цієї причини для виконання поставленої задачі необхідно використовувати ресурси, що надають програмний інтерфейс (API) для отримання необхідних даних.

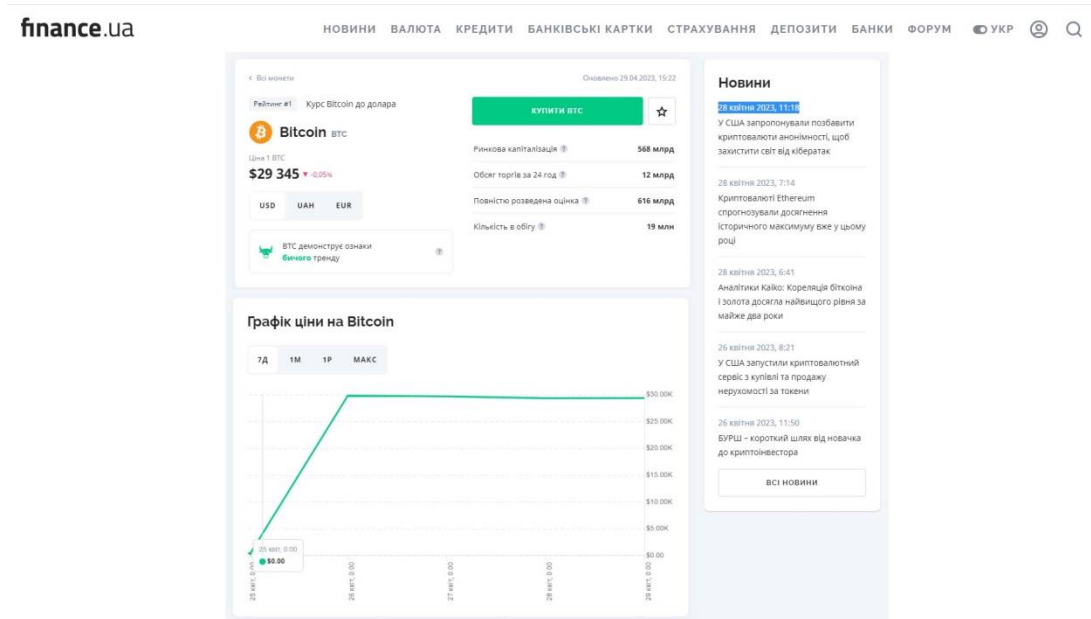


Рисунок 3.2 – Котирування BTC

Джерело: створено ресурсом <https://finance.ua/> [12]

Yahoo Finance- це ресурс, який стабільно функціонує протягом довгого терміну та надає необхідні дані з визначеною користувачем детальністю: можливо отримувати історичні дані як з інтервалом в одну хвилину так і за кожен день. Історія котирувань охоплює весь період існування криптовалют.

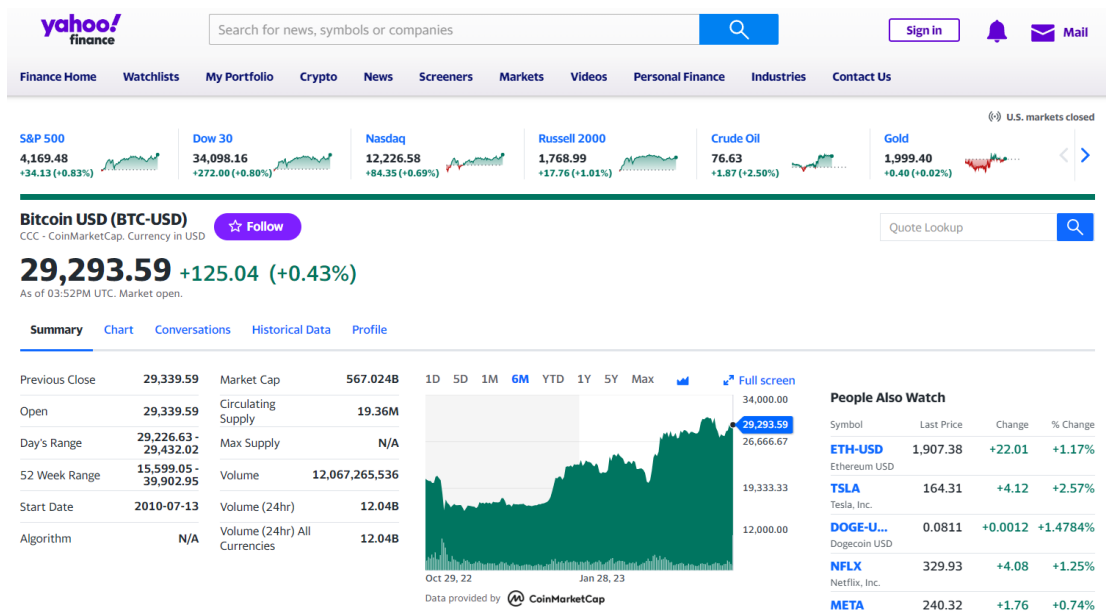


Рисунок 3.3 – Котирування BTC [13]

Джерело: створено ресурсом <https://finance.yahoo.com/> [13]

Python-Бібліотека уfinance, останнє оновлення якої з'явилося 16 квітня

2023 року, підтримує багатопотоковість, взаємодіє з протоколами безпеки Yahoo Finance та розповсюджується під ліцензією Apache License, version 2.0. Отже, являє собою зручний, швидкий та легальний спосіб отримання необхідних для подальшого аналізу даних з Yahoo Finance.

На рисунку 3.4 показано комірку Jupiter Notebook, в якій відбувається завантаження та форматування даних.

Завантаження даних

```
df = yf.download(tickers='BTC-USD', period = '1y', interval = '1d')
df = df[['Close']].copy()
df['date'] = df.index
df.rename(columns={'Close': 'rate'}, inplace=True)
df.reset_index(inplace=True, drop=True)
df.date = df.date.apply(lambda x: x.date())
df.head(7)
```

[*****100%*****] 1 of 1 completed

	rate	date
0	38469.093750	2022-05-01
1	38529.328125	2022-05-02
2	37750.453125	2022-05-03
3	39698.371094	2022-05-04
4	36575.140625	2022-05-05
5	36040.921875	2022-05-06
6	35501.953125	2022-05-07

Рисунок 3.4 – Завантаження та форматування даних
Джерело: побудовано автором на основі власної бази даних

Початково до таблиці завантажуються низка показників (open, high, low, close, adj close, volume). Усі вони окрім останнього тісно корелюють між собою, а останній є показником обсягу торгів за відповідний день. Тому в таблицю, що буде використана для моделювання, завантажені дані котирувань на закритті ціни (close) кожного дня за останній рік. Дані графічно представлені на рисунку 3.5.

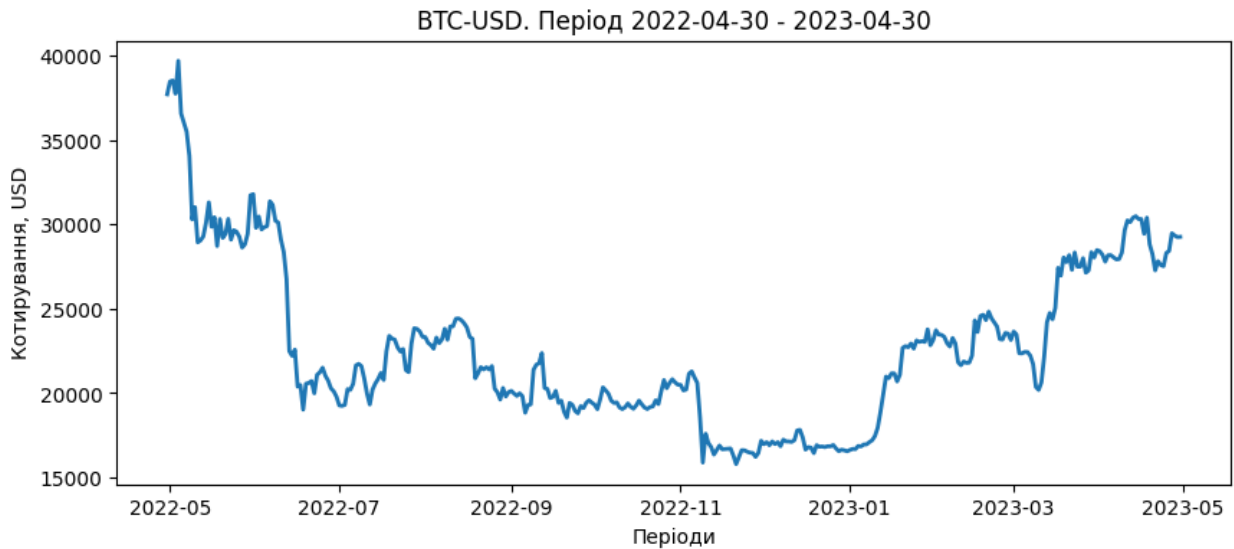


Рисунок 3.5 – Графічне представлення динаміки котирувань BTC-USD за період 30.04.2022 по 30.04.2023

Джерело: побудовано автором на основі власної бази даних

В подальшому для аналізу та побудови моделей використовуємо дані, що були завантажені у таблицю, представлену на рисунку 3.4.

Треба зауважити, що аналогічним чином можуть бути завантажені інші дані – котирування Ethereum, Gecko, або котирування акцій компаній з S&P 500. Але це не означає, що наведений нижче аналіз, статистичні висновки, або побудовані моделі можуть бути без жодних змін застосовані до котирувань інших активів.

3.2. Побудова моделі для прогнозування

Основна мова програмування – Python, версія 3.11. Python обраний як одна з трьох найпопулярніших мов програмування. З урахуванням необхідності використання широкого спектру математичних обчислень, а також використання бібліотек машинного навчання та реалізації окремих моделей, у тому числі ARIMA моделі, мова програмування Python для проведення аналізу стає практично безальтернативною.

Використовувані в програмній реалізації пакети:

- YFinance – завантаження необхідних для роботи даних про котирування в датафрейм Pandas.
- NumPy – перетворення даних
- Pandas – робота з датафреймами (таблиці даних)
- Matplotlib – основний пакет візуалізації даних
- Seaborn – обгортка Matplotlib для зручної візуалізації даних з датафреймів Pandas
- Statsmodels: модель ARIMA та інші статистичні інструменти, що допомагають у процесі налаштування моделей;
- Sklearn – бібліотеки машинного навчання, в поточній задачі використовується для зручного розрахунку необхідних метрик;
- Itertools, Os, Datetime, Warnings – допоміжні інструменти та налаштування Jupiter Notebook.

Pmdarima – бібліотека реалізації моделі ARIMA, яка буде застосовуватись для порівняльного аналізу. Це статистична бібліотека, призначена для заповнення порожнечі в можливостях аналізу часових рядів Python. Вона включає:

- еквівалент функціональності R `auto.arima`
- збірка статистичних тестів стаціонарності та сезонності

- утиліти часових рядів, такі як диференціювання та зворотня диференціація;
- численні ендогенні та екзогенні трансформатори та властивості, включаючи перетворення Бокса-Кокса та Фур'є;
- утиліти перехресної перевірки (cross validation)
- конвеєри (pipelines) Scikit-learn для консолідації.

Разом з `pmddarima` для аналізу та прогнозування часових рядів використовуємо `statsmodels`, що є доповненням `ScyPy`, та включає досить велику кількість моделей аналізу часових рядів: сезони ARIMA та ARIMAX, VARMA та VARMAX, моделі Маркова, прості авторегресійні моделі та ряд інших.

Усі етапи побудови, вивчення та порівняння моделей, що будуть створюватись, буде виконано в оточені `Jupyter`.

Задачею програмної реалізації є побудова апарату для дослідження часового ряду котирувань криптовалюти, зокрема BTC, моделей короткотермінового прогнозування котирувань, засобів оцінки побудованих моделей та візуалізації результатів аналізу та прогнозування. Також порівняння досліджених моделей за обраними метриками якості.

Для оцінки якості моделі будемо використовувати стандартні та специфічні для поставленої задачі метрики.

В якості стандартних метрик використовуємо середню абсолютну похибку (MAE) та середньоквадратичну похибку (RMSE).

Також доповнимо метрики показником, що визначається як кількість періодів прогнозування де коректне передбачення тренду, до загальної кількості періодів прогнозування. Означена метрика дозволить показати, в якій кількості періодів правильно передбачено тренд – зріст або падіння котирувань відносно останнього періоду тренувальних даних. Таким чином визначимо відсоток періодів, наступних за останнім тренувальним періодом, в яких інвестор отримає очікуване: дійсно отримає прибуток, можливо менший відносно прогнозованого, але уникне збитку.

Розглядаючи динаміку котирувань, показану на рисунку 3.5, можна

відзначити генеральну зміну тренду та дрібні осциляції котирувань навколо неї. Така поведінка через низку причин притаманна котируванням багатьох активів. В якості першого кроку проведемо згладжування кривої, результати запишемо в окремі стовпчики датафрейму, що використовується.

Графічно результати згладжування, що проведено з використанням ковзного середнього, показані на рисунку 3.6.

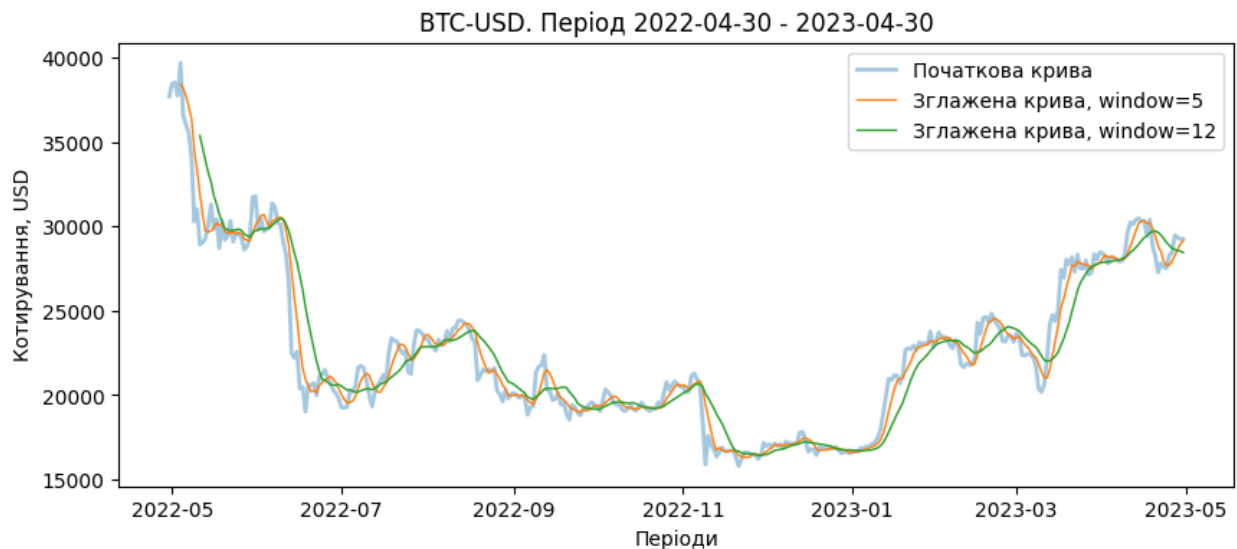


Рисунок 3.6 – Результати згладжування ковзним середнім

Джерело: побудовано автором на основі власної бази даних

На рисунку 3.6 можна бачити, що кожна з кривих повторює усі основні тренди глобальної кривої. Разом з цим усунуті дрібні осциляції навколо лінії тренду.

Наступним кроком є перевірка стаціонарності часового ряду, що розглядається. На перший погляд ряд є нестаціонарним, але підтвердимо це шляхом побудови додаткових кривих та аналітично, використавши доповнений тест Дікі – Фулера.

Якщо ковзна статистика демонструє чітку тенденцію (вгору чи вниз) і показує мінливу дисперсію (збільшення або зменшення амплітуди), можна зробити висновок, що ряд, швидше за все, не буде стаціонарним [2].

З графіків на рисунку 3.7 можна робити висновок, що початкові дані не є стаціонарними на відміну від перетворених даних.

Перейдемо до оцінки стаціонарності даних аналітично. Для цього створюємо метод `show_adfuller_results(data: pd.Series, message: str)`, куди в якості параметра `data` передаються дані, стаціонарність яких треба оцінити.

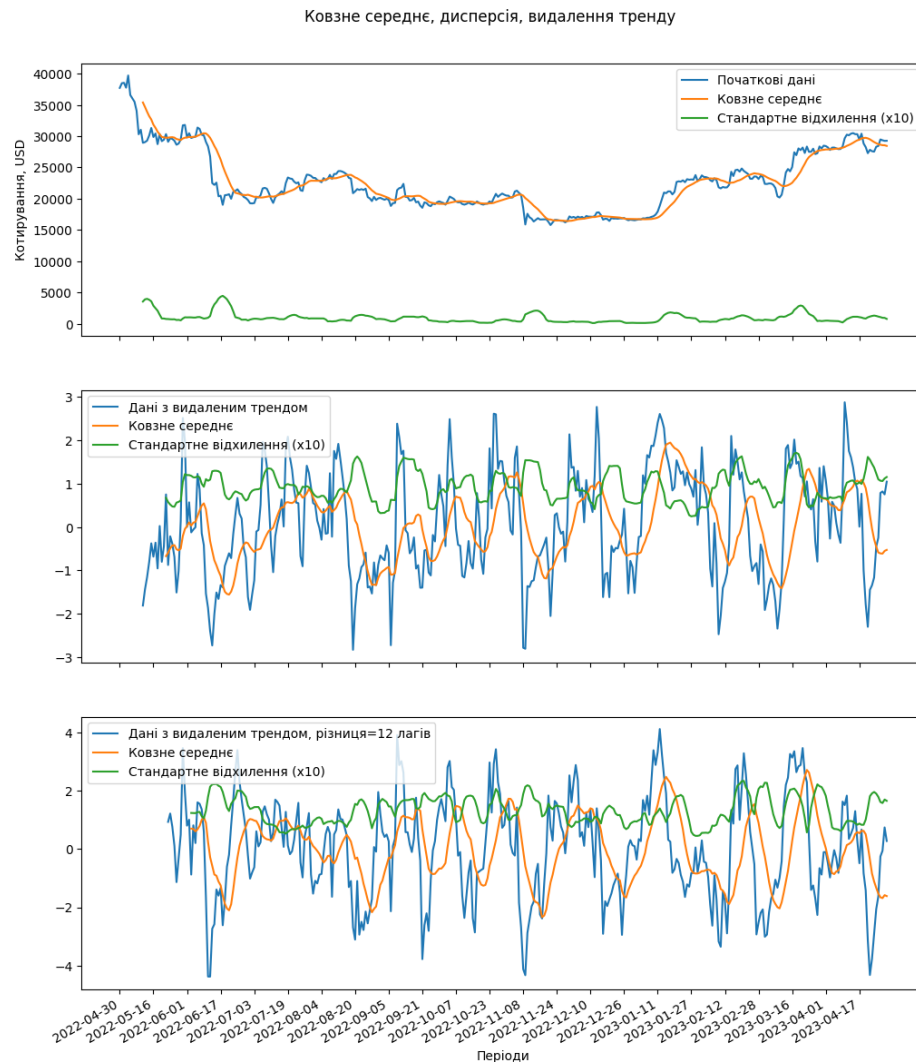


Рисунок 3.7 -Перевірка стаціонарності часового ряду графічно
Джерело: побудовано автором на основі власної бази даних

Результати виводу на початкових та перетворених раніше даних наведено на рисунку 3.8.

Початкові дані
 Значення статистики тесту = -2.852
 Р-значення = 0.051
 Критичні значення:
 1%: -3.448 - Дані не є стаціонарними з довірчою ймовірністю 99%
 5%: -2.869 - Дані не є стаціонарними з довірчою ймовірністю 95%
 10%: -2.571 - Дані є стаціонарними з довірчою ймовірністю 90%

Дані з видаленим трендом
 Значення статистики тесту = -6.663
 Р-значення = 0.000
 Критичні значення:
 1%: -3.449 - Дані є стаціонарними з довірчою ймовірністю 99%
 5%: -2.870 - Дані є стаціонарними з довірчою ймовірністю 95%
 10%: -2.571 - Дані є стаціонарними з довірчою ймовірністю 90%

Дані з видаленим трендом, різниця=12 лагів
 Значення статистики тесту = -5.816
 Р-значення = 0.000
 Критичні значення:
 1%: -3.450 - Дані є стаціонарними з довірчою ймовірністю 99%
 5%: -2.870 - Дані є стаціонарними з довірчою ймовірністю 95%
 10%: -2.571 - Дані є стаціонарними з довірчою ймовірністю 90%

Рисунок 3.8 – Вивід результатів розширеного тесту Дікі-Фулера

Джерело: побудовано автором на основі власної бази даних

З результатів що зображені на рисунку 3.8 можна зробити висновки, що співпадають з висновками, зробленими раніше на підставі графічного відтворення (рисунок 3.7). Початкові дані не є стаціонарними на відміну від перетворених даних.

ARIMA (Auto Regressive Integrated Moving Average) є узагальненням простішої моделі ARMA з додаванням складової інтегрованості. Integrated – процес що призводить часовий ряд до стаціонарності.

Знаходження параметру моделі d (differencing), що приводить ряд до стаціонарності є першим кроком в знаходженні коректних параметрів, що описують модель (p, d, q) : p – лінійна комбінація лагів, q - лінійна комбінація лагів прогнозованої помилки.

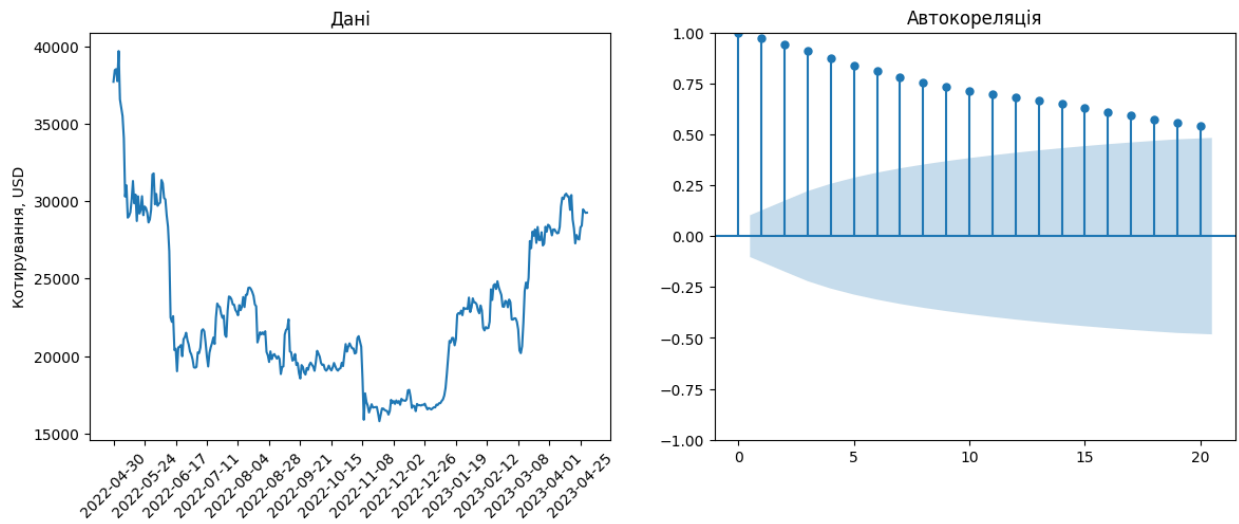


Рисунок 3.9 – Графік автокореляції – початкові дані

Джерело: побудовано автором на основі власної бази даних

На графіку видно, що дані мають тенденцію, відповідно нестационарність часового ряду, що розглядається, підтверджується.

Розглянемо різниці першого та другого порядку для визначення, який порядок різниць буде використовуватись при побудові моделі.

Тепер переходимо до вивчення графіків автокореляції та часткової автокореляції, що показані на рисунках 3.10 та 3.11. Визначення термінів моделі за допомогою цих діаграм є складним і схильним до помилок: на перший погляд графіки виглядають таким чином, що це модель $ARIMA(0, 1, 0)$. Якщо це справді модель $ARIMA(0, 1, 0)$, наші відмінні дані будуть так звані «білим шумом», а вихідні дані — це «випадкове блукання».

Якщо змінні є незалежними та однаково розподіленими із середнім нульовим значенням, часовий ряд є білим шумом. Крім того, випадкове блукання — це ряд чисел, де кожне значення є функцією попереднього значення. Ряд білого шуму залишається, коли береться перша різниця випадкового блукання.

Отже, під час вивчення часового ряду може статися, що він є випадковим блуканням, а значить, його статистичний аналіз на великому інтервалі неможливий.

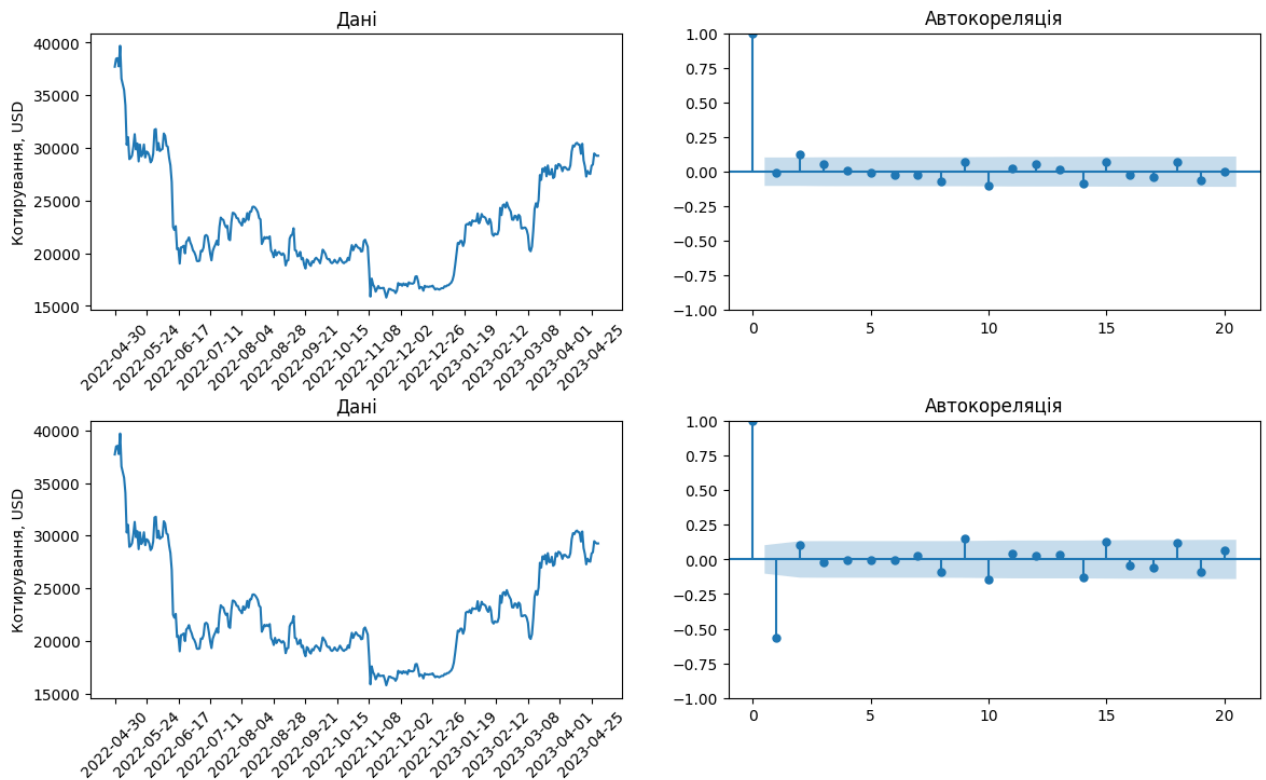


Рисунок 3.10 – Графік автокореляції – перший та другий порядки

Джерело: побудовано автором на основі власної бази даних

При $d = 1$ бачимо на графіку 3.10, що дотримується стаціонарність ряду.

При $d = 2$ ми бачимо на графіку 3.10, що другий лаг йде далеко в від'ємну зону.

Отже, попередньо візьмемо порядок обчислення різниць $d = 1$.

Побудуємо графік часткової автокореляції (PACF).

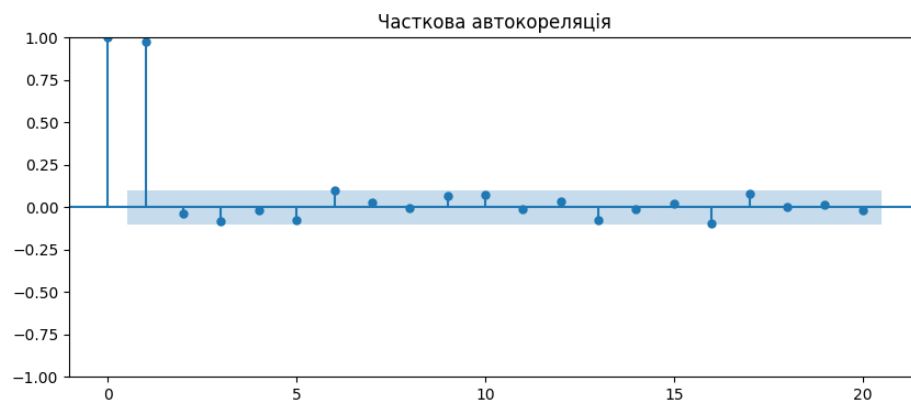


Рисунок 3.11 – Часткова автокореляція даних

Джерело: побудовано автором на основі власної бази даних

Аналізуючи графік можна зупинитись на значеннях порядку авторегресійної моделі від 1 до 3. Спочатку будемо проводити тестування моделей з порядком 1, а за необхідності збільшуватимемо порядок моделей.

Повернемося до графіку автокореляції 3.9 для визначення порядку ковзного середнього (МА). З технічної точки зору, МА - помилка запізненого прогнозу. На рисунку 3.9 не бачимо суттєвих відхилень, тому початково беремо $q = 1$.

3.3. Оцінка отриманих результатів та практичне застосування

З метою тестування моделей набір даних, що вивчаються, повинен бути поділений на тренувальний та тестовий датасети.

Дані за останні 90 днів відносимо до тестової частини датасету, а решту – до частини, що використовується для навчання моделей. Розподіл даних зображено на рисунку 3.12.



Рисунок 3.12 – Розподіл даних на тренувальні та тестові

Джерело: побудовано автором на основі власної бази даних

Будуючи першу авторегресійну модель першого порядку ($p = 1$), отримуємо результати, що візуалізовано на рисунку 3.13

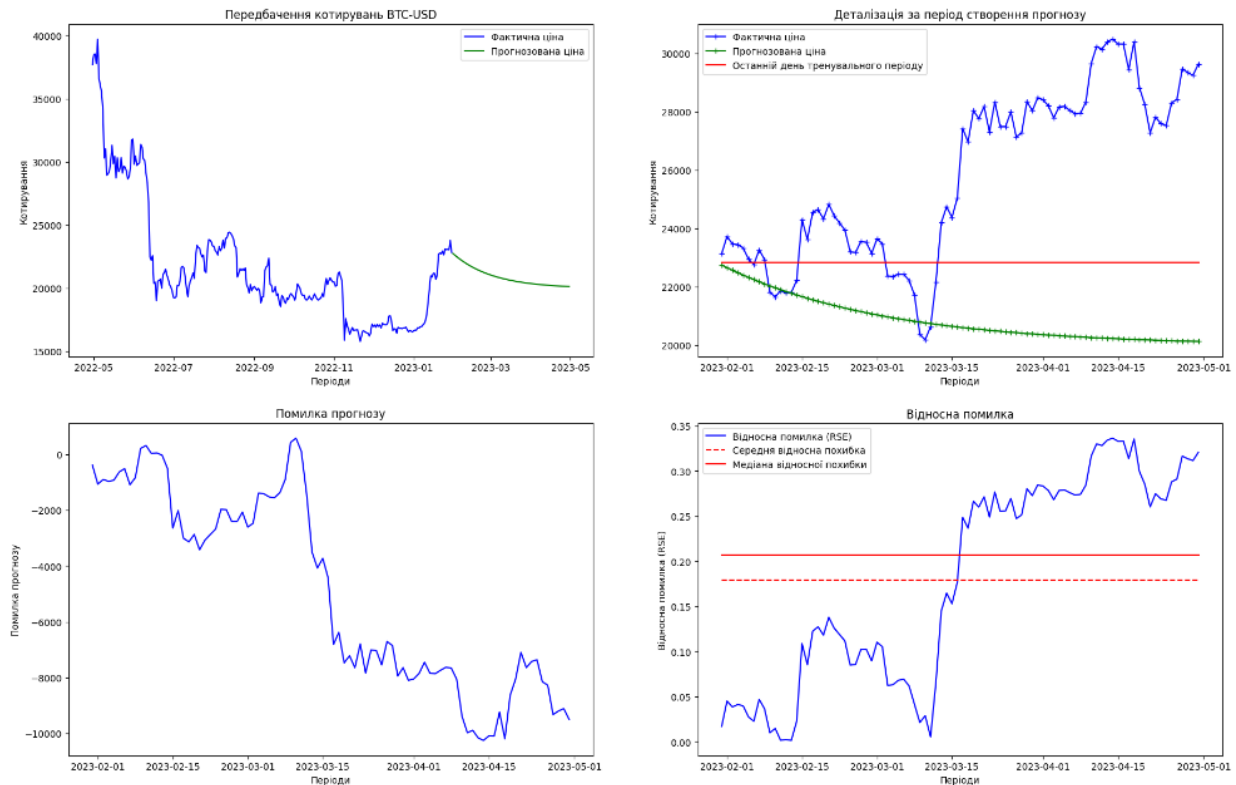


Рисунок 3.13 – Авторегресійна модель першого порядку

Джерело: побудовано автором на основі власної бази даних

За допомогою вищевизначених метрик оцінюємо отримані прогнозовані значення ціни, як показано на рисунку 3.14.

Середня абсолютна похибка: 4942.642
 Середньоквадратична помилка: 6021.078
 Коефіцієнт детермінації моделі: -3.225
 Метрика інвестора: 0.189

Рисунок 3.14 – Оцінка отриманої моделі першого порядку

Джерело: побудовано автором на основі власної бази даних

Оцінюючи візуально, робимо висновок, що модель може лише правильно прогнозувати початкову частину періоду прогнозування, що характеризується зниженням котирувань. Про це свідчить й поступове зростання помилки прогнозу при просуванні до більших дат періоду прогнозування.

Зміна порядку авторегресійної моделі $p = 2,3$ не призводить до покращення прогнозування – середньоквадратична помилка прогнозів не

знижується.

Шляхом подальшої зміни порядку моделі та оптимізації її інших гіперпараметрів вдається досягнути зменшення середньоквадратичної похибки та отримати результати, що показані на рисунку 3.15.

Як можна бачити, модель значно краще відтворює тренди у порівнянні з першою авторегресійною моделлю, результати прогнозування для якої були показані на рисунку 3.13. Моделі вдається досить якісно відтворити тренди перших 20 днів тестового періоду, де відносна похибка не перевищує 0.06; а також з 40 по 85 дні, де відносна похибка не перевищує 0.08. Але різкий спад котирувань в інтервалі 20 – 40 день модель не відтворює, чим суттєво погіршена якість прогнозування. З практичної точки зору це означає потенційні втрати інвестора в період 30 – 40 днів, оскільки котирування падають нижче рівня останнього дня, що належить до тренувального періоду.

Середня абсолютна похибка: 1690.230
 Середньоквадратична помилка: 2496.853
 Коефіцієнт детермінації моделі: 0.274
 Метрика інвестора: 0.822

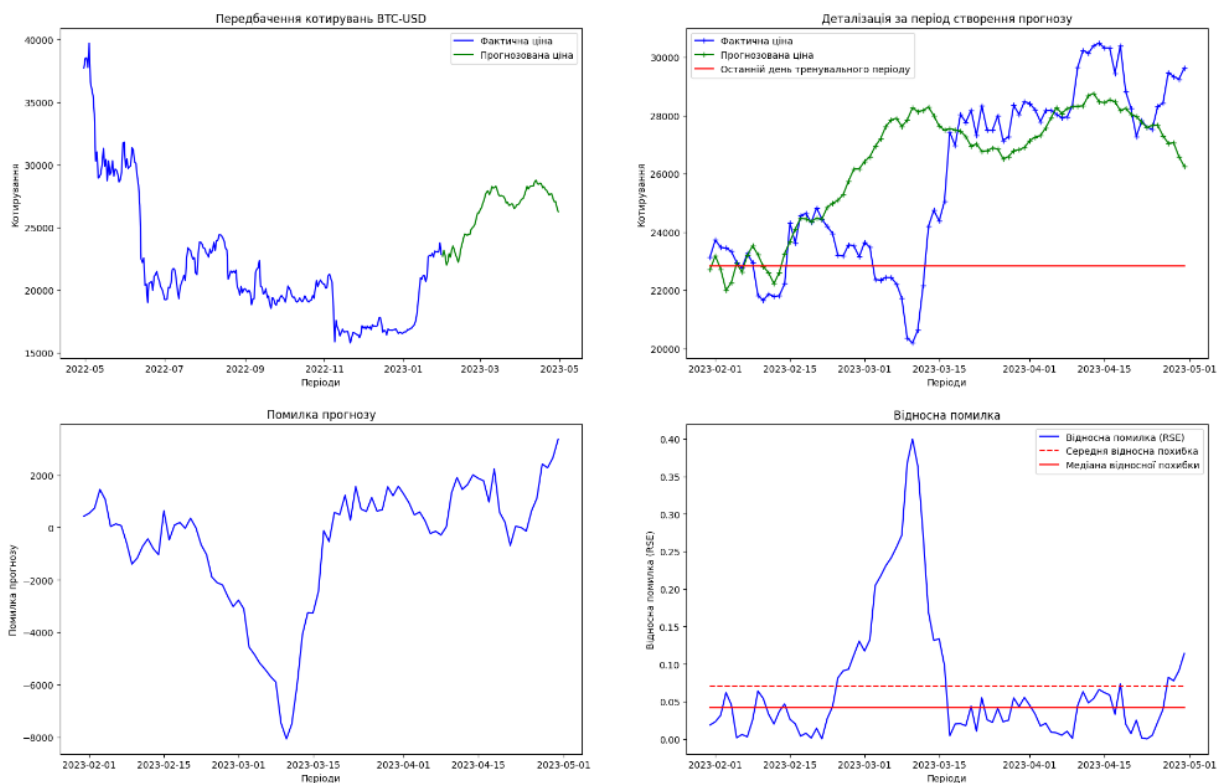


Рисунок 3.15 – Результати авторегресійної моделі після оптимізації гіперпараметрів

Джерело: побудовано автором на основі власної бази даних

Використання попередньо згладжених даних та моделі з оптимізованими гіперпараметрами дозволяє незначно покращити результати прогнозування у випадку використання меншого розміру вікна, рисунок 3.16.

Середня абсолютна похибка: 1780.821
 Середньоквадратична помилка: 2372.380
 Коефіцієнт детермінації моделі: 0.287
 Метрика інвестора: 0.779

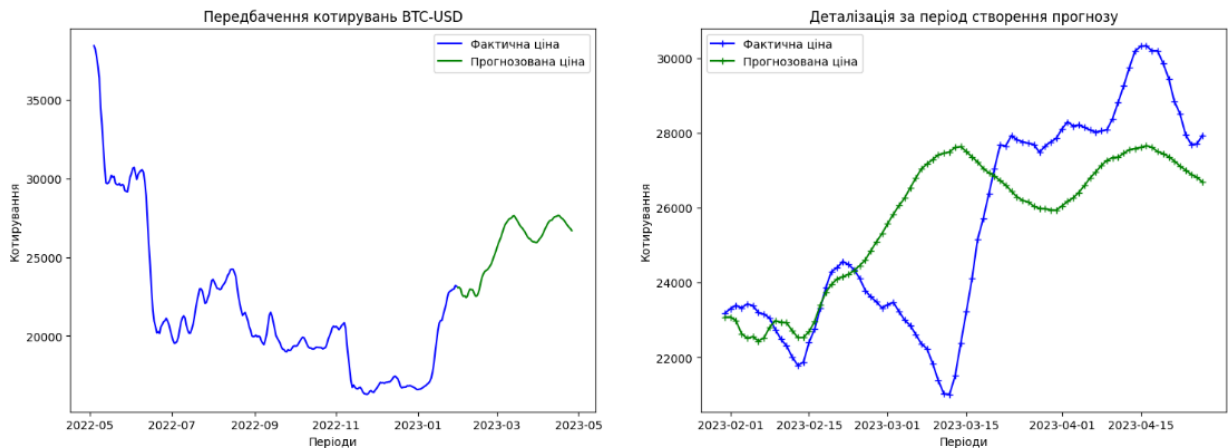


Рисунок 3.16 – Застосування моделі на згладжених даних з розміром вікна 5

Джерело: побудовано автором на основі власної бази даних

Середня абсолютна похибка: 1968.940
 Середньоквадратична помилка: 2448.288
 Коефіцієнт детермінації моделі: 0.024
 Метрика інвестора: 0.722

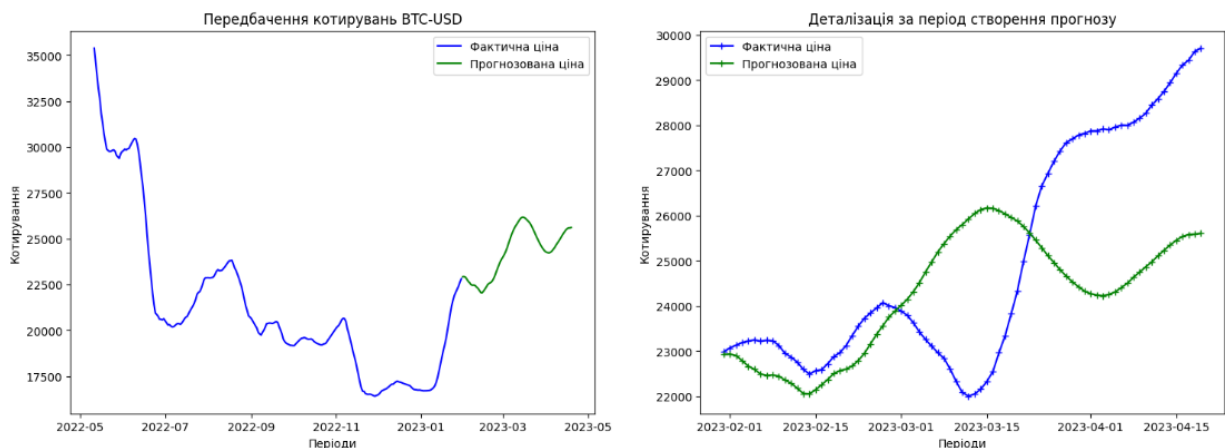


Рисунок 3.17 – Застосування моделі на згладжених даних з розміром вікна 12

Джерело: побудовано автором на основі власної бази даних

Використання згладжування з розміром вікна, що дорівнює 12, призводить

до меншого покращення моделі.

Якщо виміряти покращення прогнозування шляхом порівняння середньоквадратичних помилок, то в першому випадку вдалося досягти покращення біля 5%, а в другому – 2%. При цьому порівняння прогнозу з тестовими даними проводилось для відповідно згладжених даних, розбіжність прогнозу на згладжених даних з початковими (не згладженими) буде більше.

Загальною характеристикою отриманих простих авторегресійних моделей є те, що в достатній мірі справляються з короткостроковим прогнозуванням – перші 20-25 періодів (днів), а також передбачають зміну тренду при незначних коливаннях на більших періодах – до 90 днів. Але жодна з побудованих на цьому етапі моделей виявилась неспроможною передбачити різке та суттєве падіння котирувань в період 25 – 40 інтервал прогнозу.

Перейдемо до побудови моделі ARIMA. Спочатку для побудови використовуємо, як і раніше, statsmodels.

Отримуємо прогноз, що показано на рисунку 3.18.

Середня абсолютна похибка: 3226.532
Середньоквадратична помилка: 4000.243
Коефіцієнт детермінації моделі: -0.870
Метрика інвестора: 0.211

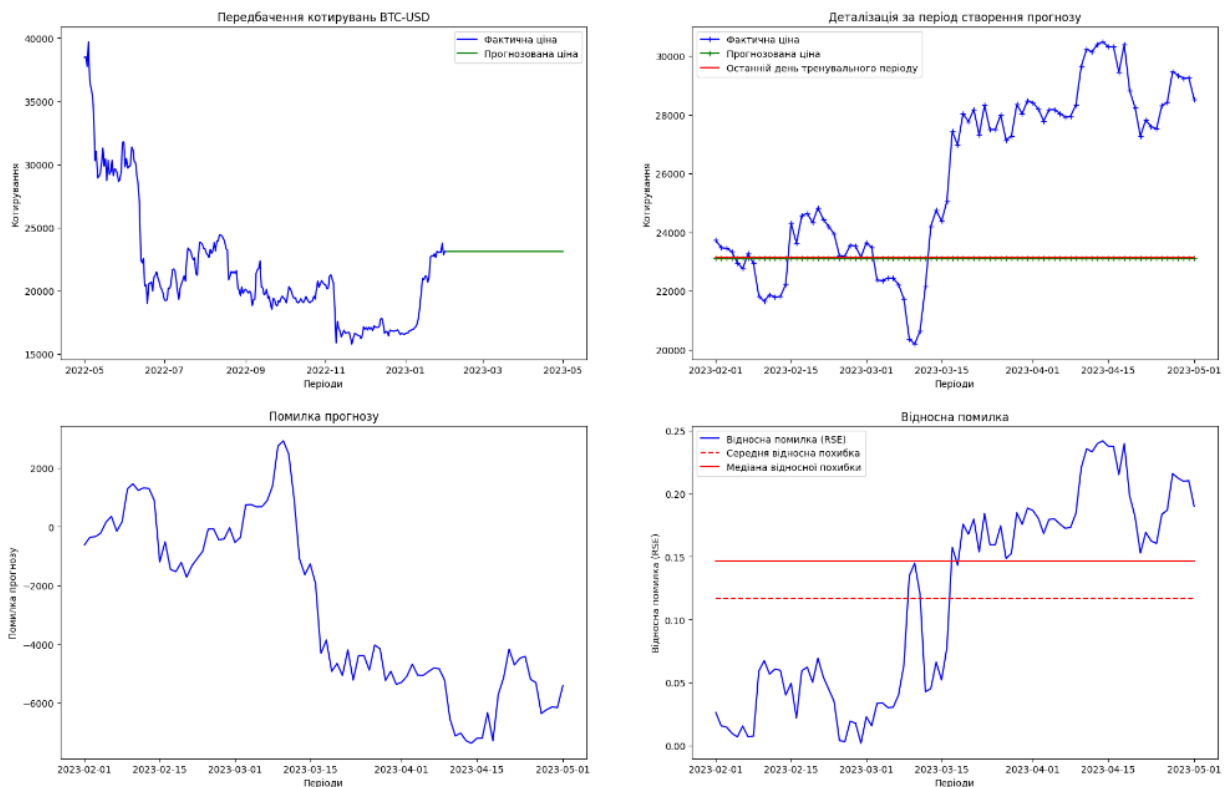


Рисунок 3.18 – Перший з прогнозів statsmodels.ARIMA

Джерело: побудовано автором на основі власної бази даних

За таких налаштувань, побудована модель ARIMA не справляється з прогнозуванням через те, що в параметрах моделі не задано параметр trend.

Додавши до налаштувань моделі детерміністський тренд виду

$$dt^4 + ct^3 \quad (3.1);$$

тренуємо модель ще раз та отримуємо результати, що показані на рисунку 3.21.

На підставі 3.21 можна зробити висновок, що урахування детерміністського тренду суттєво покращує результати прогнозування. З одного боку метрики для отриманого результату краще за будь-який з результатів, отриманих вище з регресійними моделями (AR). З іншого боку, це обумовлено перед усім за рахунок використання детерміністського тренду, до якого модель оптимізується теж з використанням методу найменших квадратів. Жодне з коливань котирувань, які мають місце, модель не враховує. Й це стосується не лише різкого падіння в інтервалі 25 – 40 періоду.

Середня абсолютна похибка: 1343.276
 Середньоквадратична помилка: 1786.469
 Коефіцієнт детермінації моделі: 0.627
 Метрика інвестора: 0.778

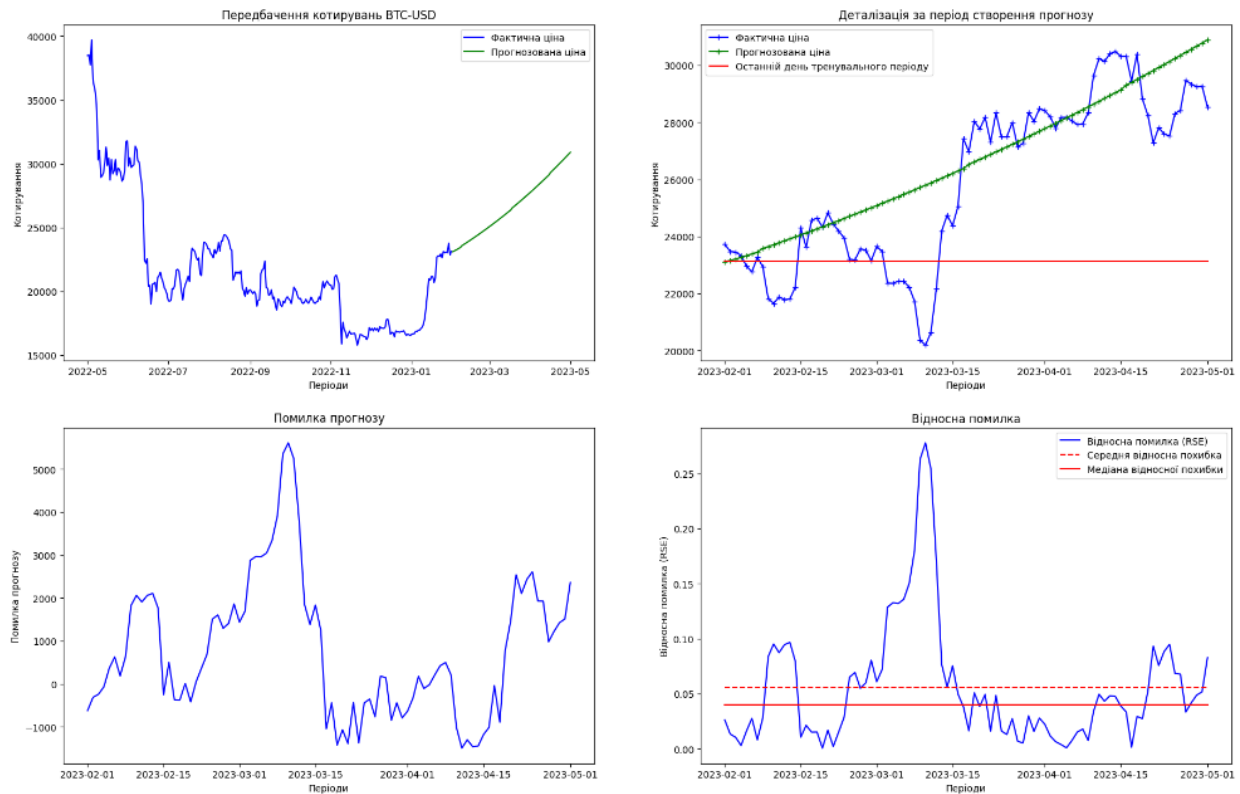


Рисунок 3.19 – Модель ARIMA з визначенням детерміністським трендом

Джерело: побудовано автором на основі власної бази даних

Для уточнення проведемо повторну оцінку максимального порядку різниці. Тут разом із розширеним тестом Дікі-Фулера використаємо так званий тест (критерій) KPSS (Квятковський – Філіпс – Шмідт – Шін).

За результатами вибору максимального значення цих означених двох тестів маємо максимальний порядок різниці, що дорівнює 2.

```

: kpsd_diffs = ndiffs(y_train, alpha=0.05, test='kpsd', max_d=5)
  adf_diffs = ndiffs(y_train, alpha=0.05, test='adf', max_d=5)
  n_diffs = max(adf_diffs, kpsd_diffs)
  print(f'Оцінка порядку різниці складає: {n_diffs}')

```

Оцінка порядку різниці складає: 2

Рисунок 3.20 – Визначення порядку різниці: максимум двох критеріїв

Джерело: побудовано автором на основі власної бази даних

При подальшій оптимізації параметрів моделі ARIMA врахуємо цей

результат.

В результаті оптимізації параметрів вдається покращити модель, що побудована на 17% за метрикою RMSE. Графічно результати відображено на рисунку 3.21.

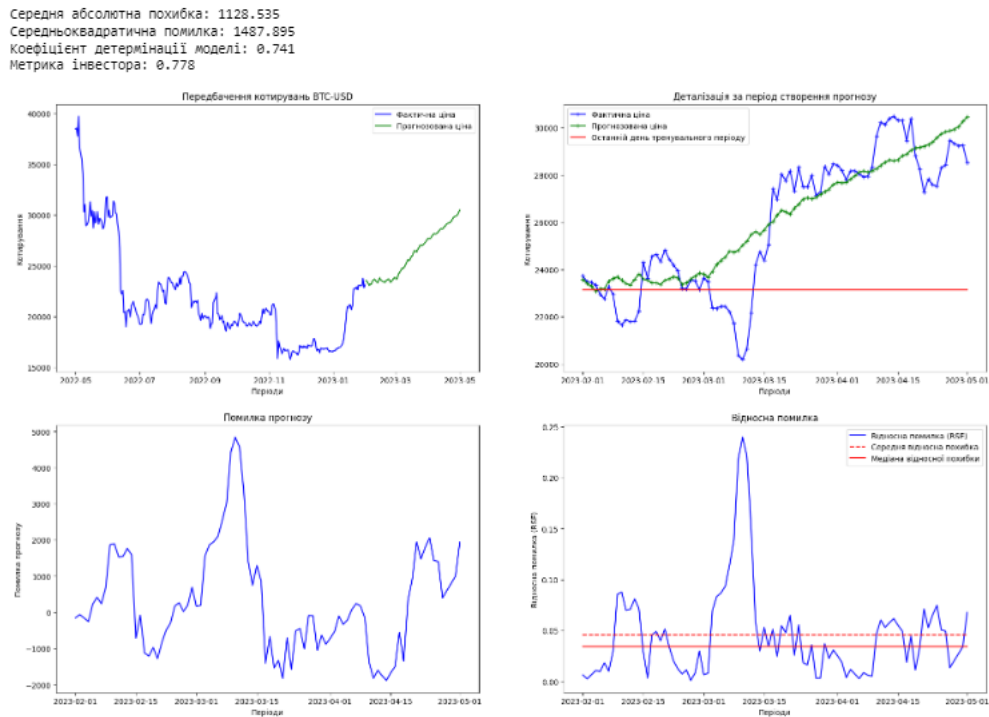


Рисунок 3.21 – Модель після оптимізації гіперпараметрів

Джерело: побудовано автором на основі власної бази даних

Отримана модель, як і раніше, не передбачає падіння в періоді 25 – 40 інтервал, але відтворює незначні осциляції навколо лінії тренду. Для згладжених даних з меншим вікном згладжування отримуємо результат, що показаний на рисунку 3.22.

Середня абсолютна похибка: 1260.139
 Середньоквадратична помилка: 1969.604
 Коефіцієнт детермінації моделі: 0.510
 Метрика інвестора: 0.756

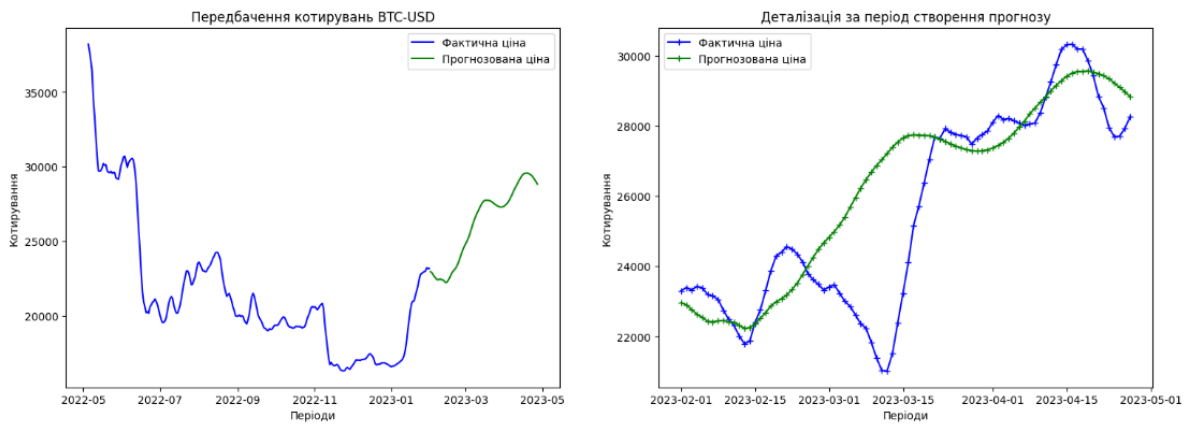


Рисунок 3.22 – Застосування моделі на згладжених даних

Джерело: побудовано автором на основі власної бази даних

Розглянувши 3.22, можна відмітити, що незважаючи на більше значення середньоквадратичної похибки, отриманий прогноз краще повторює зміни в динаміці котирувань, хоча й знов таки жодним чином не враховує суттєвий спад котирувань.

Використаємо інший інструментарій для створення прогнозу – бібліотеку `pytdarima`. Суттєвою відмінністю від `statsmodels`, що є збіркою великої кількості моделей, є те, що `pytdarima` сфокусована саме на моделях сімейства ARIMA. Бібліотека включає клас `AutoARIMA`, який є ARIMA моделлю, що шляхом повного перебору варіантів або процесу Гіндмана – Хандакара автоматично налаштовує параметри моделі.

Створимо модель із налаштуваннями, які показані на рисунку 3.23.

Використання pmdarima

```

step_wise=pm.auto_arima(
    y_train,
    exogenous=x_train,
    start_p=1,
    start_q=1,
    max_p=20,
    max_q=7,
    d=1,
    max_d=2,
    trace=True,
    error_action='ignore',
    suppress_warnings=True,
    stepwise=True,
    stationary=False,
    n_jobs=6
)
print(step_wise.summary())

```

Рисунок 3.23 – Інстанціювання AutoARIMA

Джерело: побудовано автором на основі власної бази даних

Передивляючись підсумки налаштування моделі, спостерігаємо наступний результат, що показаний на рисунку 3.24.

```

Performing stepwise search to minimize aic
ARIMA(1,1,1)(0,0,0)[0] intercept : AIC=4452.944, Time=0.08 sec
ARIMA(0,1,0)(0,0,0)[0] intercept : AIC=4448.239, Time=0.01 sec
ARIMA(1,1,0)(0,0,0)[0] intercept : AIC=4450.227, Time=0.00 sec
ARIMA(0,1,1)(0,0,0)[0] intercept : AIC=4450.230, Time=0.08 sec
ARIMA(0,1,0)(0,0,0)[0]          : AIC=4447.636, Time=0.00 sec

Best model: ARIMA(0,1,0)(0,0,0)[0]
Total fit time: 0.166 seconds

```

Рисунок 3.24 – Результати тренування моделі AutoARIMA

Джерело: побудовано автором на основі власної бази даних

Підібрана найкраща модель є випадковим блуканням, бо $p = 0, q = 0$. Прогноз, отриманий з використанням цієї моделі, також цілком підтверджує цей результат.

Середня абсолютна похибка: 3208.970
 Середньоквадратична помилка: 3979.609
 Коефіцієнт детермінації моделі: -0.851
 Метрика інвестора: 0.211

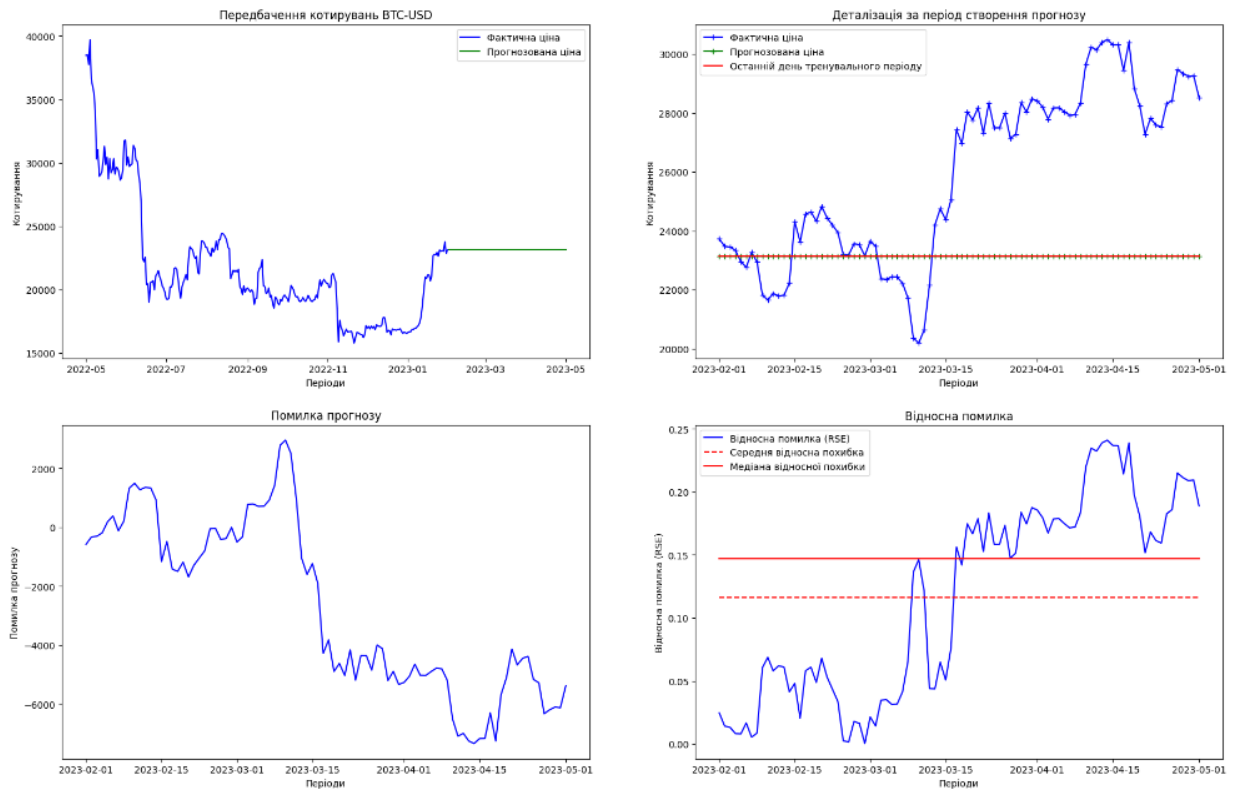


Рисунок 3.25 – Прогноз $ptdarima(0, 1, 0)$

Джерело: побудовано автором на основі власної бази даних

Використаємо Q-тест Льюнг — Бокса. Значення р-значення для різної кількості лагів показано на рисунку 3.26.

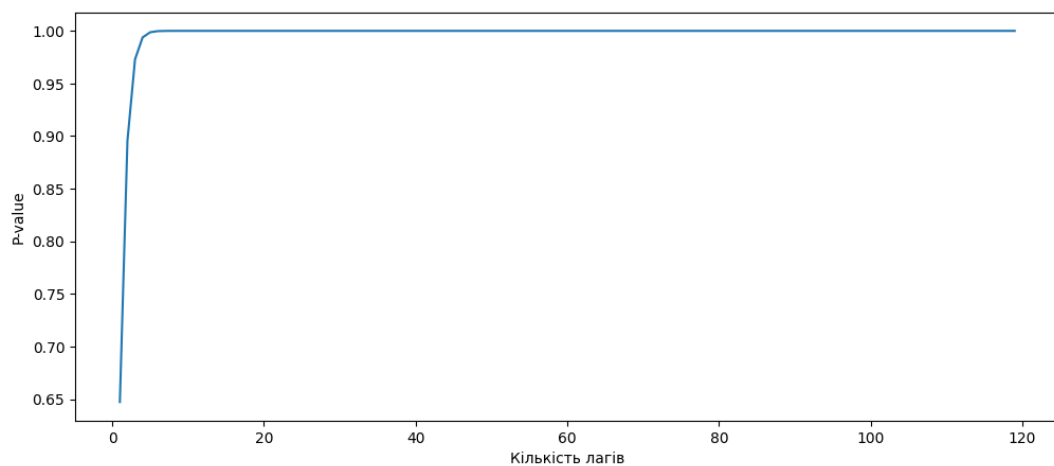


Рисунок 3.26 – P-значення тесту Льюнг — Бокса для різної кількості лагів

Джерело: побудовано автором на основі власної бази даних

Для кількості лагів 5 та більше значення p-value стрімко наближається до 1. Таким чином, гіпотеза про випадкове блукання підтверджується для кількості періодів, що перевищує 5.

Для згладжених з невеликим вікном даних крива рухається вправо – кількість лагів, при яких p-value не наближується до 1 збільшується. Але крива для більшого розміру вікна згладжування знаходиться лівіше.

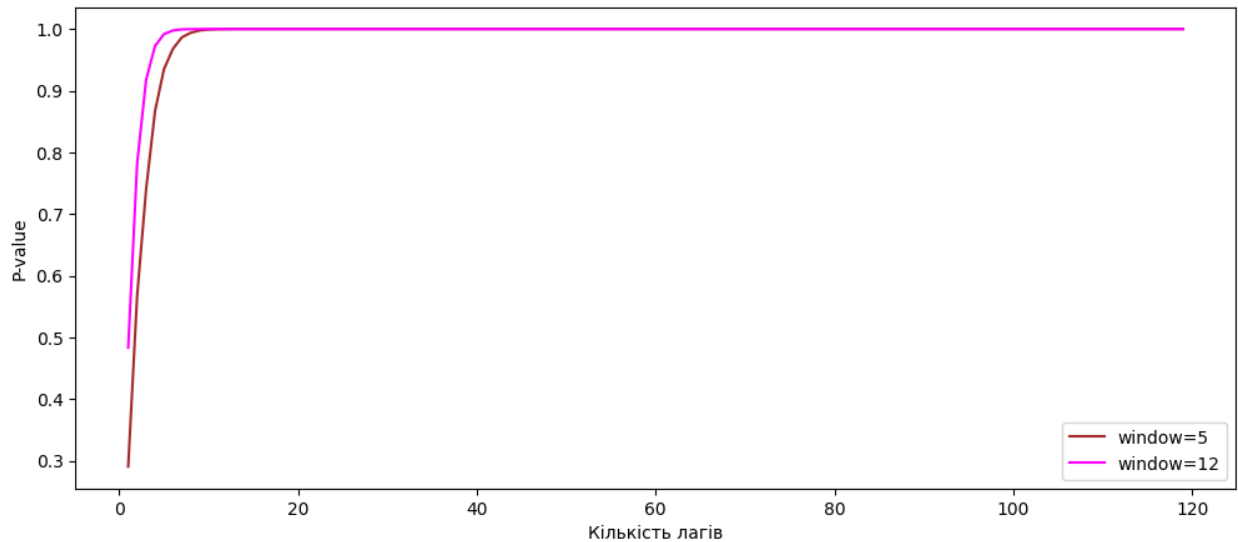


Рисунок 3.27 - P-значення тесту Льюнг — Бокса для різної кількості лагів на згладжених даних

Джерело: побудовано автором на основі власної бази даних

Для побудови короткострокового прогнозу обмежимо періоди (кількість) тренувальних даних моделі та період прогнозування, обмеживши останній 5 інтервалами. Для кількості тестових даних протестуємо різноманітні набори розмір яких не перевищує 30 значень. Використовувати будемо попередньо згладжені дані, що знизить компоненту білого шуму в розкладені часового ряду, що розглядається.

Задіявши повний перебір замість процесу Гіндмана – Хандакара, отримуємо, що найкращою з моделей буде ARIMA(3,1,2) з відсутньою сезонністю.

Графічно результати прогнозування в нових умовах показані на рисунку 3.28.

Середня абсолютна похибка: 12.810
 Середньоквадратична помилка: 16.216
 Коефіцієнт детермінації моделі: 0.870
 Метрика інвестора: 1.000

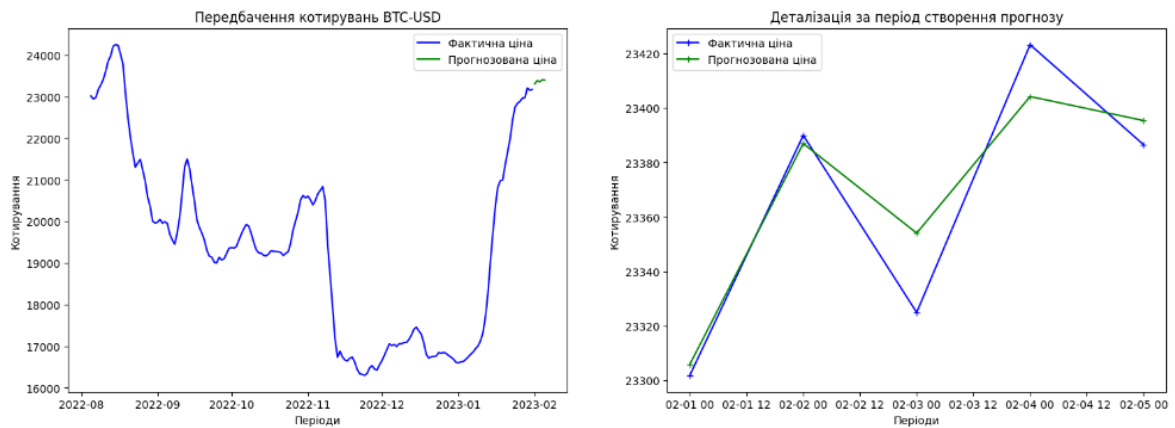


Рисунок 3.28 – Прогнозування на 5 періодах

Джерело: побудовано автором на основі власної бази даних

Розглядаючи результати прогнозування на 5 періодах на базі невеликої кількості спостережень отримано результат, що суттєво перевершує прогнозування на 90 періодів. На 5 періодів наперед вдається передбачити обидві: позитивну та негативну динаміку коливань, а в перших двох періодах фактичне та передбачене значення майже співпадають.

Про якість такого короткострокового прогнозу свідчать й отримані метрики, значення яких знаходяться на дуже хорошому рівні.

Перевіримо, чи справляється побудована модель з прогнозування на більший період. Як можна бачити з рисунку 3.29 – модель не справляється з прогнозуванням із збільшенням інтервалу прогнозування.

Середня абсолютна похибка: 245.309
 Середньоквадратична помилка: 388.218
 Коефіцієнт детермінації моделі: -0.762
 Метрика інвестора: 0.600

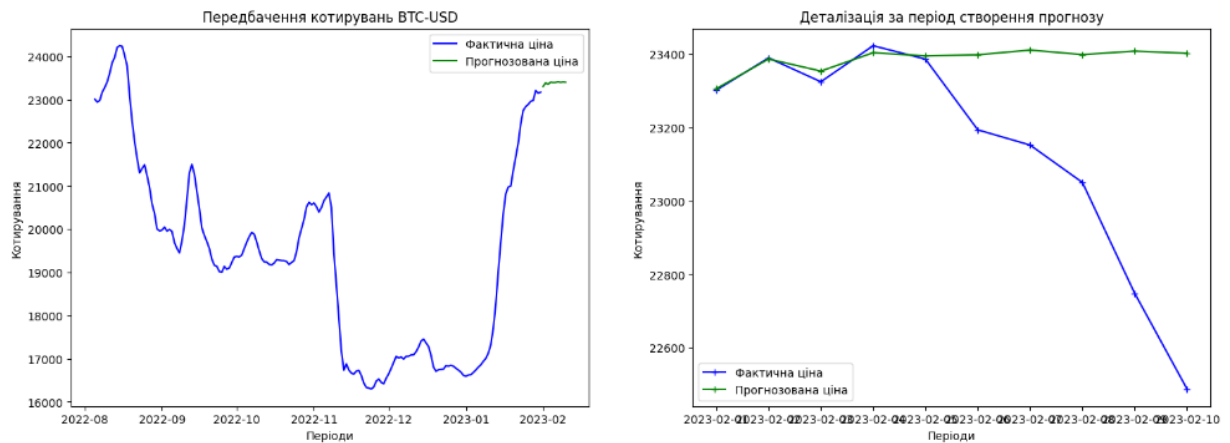


Рисунок 3.29 – Збільшений до 10 періодів інтервал прогнозування
 Джерело: побудовано автором на основі власної бази даних

Модель, що справляється на перших періодах із прогнозуванням зростаючого тренду, повністю провалює прогнозування зниження котирувань, яке відбувається з початку 6 інтервалу прогнозування.

Повертаючись до моделі ARIMA з statsmodels також будемо моделі невеликих порядків, спираючись на меншу кількість даних. Типовим прикладом вивчених при різних значеннях параметрів моделей буде представлена на рисунку 3.30 модель.

Середня абсолютна похибка: 218.656
 Середньоквадратична помилка: 259.938
 Коефіцієнт детермінації моделі: 0.210
 Метрика інвестора: 0.600

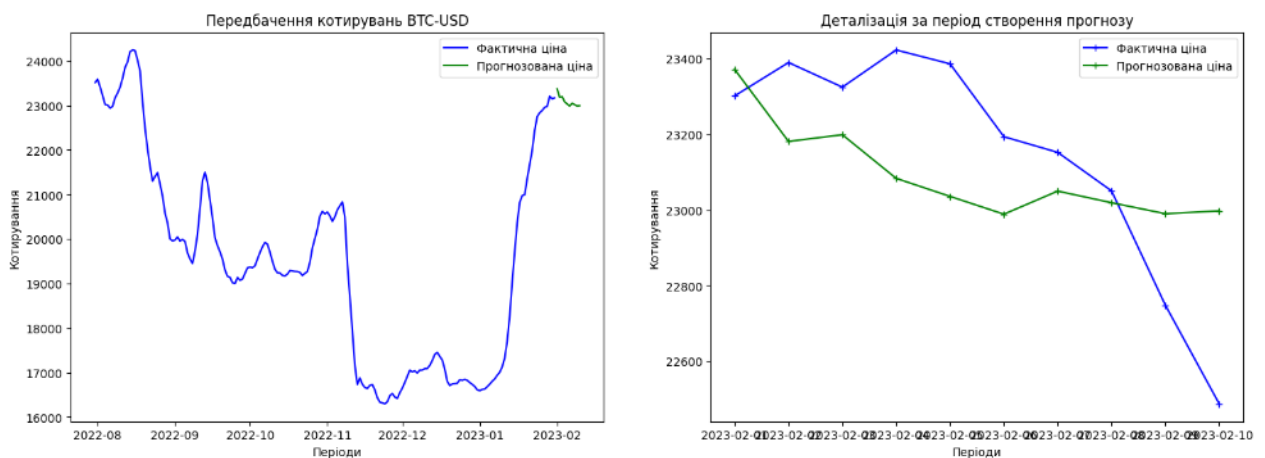


Рисунок 3.30 – Модель statsmodels.ARIMA

Джерело: побудовано автором на основі власної бази даних

Побудована на згладжених даних та невеликому тренувальному датасеті модель частково справляється з прогнозування трендів на інтервалі прогнозування в 10 днів, але слабо прогнозує збільшення котирувань на першій його частині. Окрім цього, прогнозоване спадання котирувань набагато менше, аніж фактичне.

Це також є підтвердженням того, що можливе прогнозування лише на суттєво обмежених інтервалах. З іншого боку, можливий погляд на результати прогнозування як на індикатор висхідного або низхідного, а не як на точний числовий прогноз. В такому сенсі результати прогнозування, що отримуються з використанням моделей сімейства ARIMA, мають певну практичну цінність.

З урахування результатів Q-теста Льюнга – Бокса, такий розгляд є доцільним.

В ході дослідження була побудована певна кількість моделей прогнозування часових рядів.

При попередньому вивченні часового ряду виявилось, що він не є стаціонарним з довірчою ймовірністю 95%. Різниці першого та більш високих порядків цього ряду є стаціонарними. Також в рамках дослідження виявлено, що детерміністський тренд процесу, що розглядається, найкраще визначається поліномом четвертого ступеню (3.1), в якого нульові коефіцієнти членів другого та меншого ступенів. Сезонний тренд в досліджуваних даних відсутній.

При детальному вивченні результатів прогнозування на порівняно великій кількості (270+ спостережень) тренувальних даних з використанням тесту Льюнга-Бокса виявилось, що зміна котирувань є випадковим блуканням, отже, якісний опис статистичними методами з подальшим прогнозуванням не є можливим. З іншого боку це підтвердилося тим, що найкращі налаштування моделі ARIMA виявились $(0, 1, 0)$, що також відповідає випадковому блуканню.

При переході до короткострокового прогнозування – 5-10 інтервалів (днів) виявилось, що можливо побудувати модель прогнозування на 4-6 періодів з

високою якістю. Але ця модель виявляється неспроможною до якісного прогнозування вже із збільшенням інтервалу прогнозування до 10 періодів.

На базі дослідження можна зробити висновок, що модель ARIMA при прогнозування продемонструвала кращі результати у порівнянні з простою регресійною моделлю (AR).

Ще одним висновком є те, що згладжування даних дозволяє підвищити якість тренування моделей за рахунок усунування дрібних осциляцій, що притаманні розглянутим котируванням Bitcoin. При цьому невелике вікно згладжування ($n=5$) виявляється достатнім, а підвищення його розміру до 12 спостережень у певних випадках виявляється надмірним.

Оптимізацію параметрів моделі найкраще проводити з використанням бібліотеки `pmдаріма`, що дозволяє проводити налаштування гіперпараметрів моделі в заданих діапазонах автоматично шляхом повного перебору або процесу Гіндмана – Хандакара.

ВИСНОВКИ

В даній роботі було досліджено ринок криптовалют та визначено основні фактори, що впливають на коливання курсів криптовалют. Результатом роботи є створення моделі короткострокового прогнозування котирувань курсів криптовалют.

В результаті розробки статистичних моделей короткострокового прогнозування курсів криптовалют можна зазначити, що побудовані моделі:

- дозволяють працювати з нестационарними часовими рядами, яким не притаманна виражена сезонність;
- спроможні передбачити загальний тренд на підвищення/зниження котирувань;
- прогнозують певні коливання котирувань навколо лінії тренду;
- на великих проміжках часу зміна котирувань, у тому числі згладжених даних є випадковим блуканням. Через це прогнозування котирувань на великому інтервалі часу з використанням статистичних моделей є майже неможливим;
- при переході до короткострокового прогнозування котирувань (4 - 6 періодів) вдається досягти помітно кращих результатів у порівнянні з прогнозуванням на 60-90 періодів. Це пояснюється тим, що для прогнозування використовується коротша база навчання моделі, в якій можна виявити статистичні закономірності;
- для короткострокового прогнозування отримано коефіцієнт детермінації моделі $R^2 = 0.87$. Для моделей з більшим інтервалом прогнозування коефіцієнт детермінації не перевищував 0.5;
- у значній кількості випадків отриманий прогноз слід розглядати як керівництво з трендів (зростання, падіння, знову зростання), аніж як конкретні цифрові значення котирувань у відповідний день.

Моделі прогнозування динаміки часового ряду певного показника

неспроможні передбачити зміни, що не є наслідками статистичних закономірностей. Котирування криптовалют, так само як котирування інших вартісних активів, схильні до впливу зовнішніх факторів. Так, різке падіння котирувань, що спостерігалось на початку березня 2023 року пов'язано з закриттям та подальшою ліквідацією банку Silvergate, одного з двох найбільших банків США, що працювали на той момент з криптовалютами. Звісно, цю подію не можна передбачити, виходячи виключно з статистичного аналізу історії котирувань Bitcoin.

Технічний трейдинг, та прогнозування котирувань є одним з досить розповсюджених застосувань машинного навчання. Зокрема, для прогнозування котирувань використовуються моделі прогнозування часових рядів.

Відповідно до висновків, побудовану модель можна використовувати для прогнозування тренду, а також локальних коливань навколо цієї лінії тренду. У той же час існує загроза суттєвих коливань котирувань, що не будуть передбачені.

Підводячи підсумок, можна стверджувати, що побудовані статистичні моделі є одним з корисних інструментів в технічному аналізі трейдинга, але не можуть бути використані в якості єдиного інструменту для аналізу.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Бідюк П. І. Часові ряди: моделювання та прогнозування. / Бідюк П. І., Савенков О. І., Баклан І. В. - Київ: ЕКМО, 2003. 82 с.
2. Бідюк П. І. Аналіз часових рядів (навчальний посібник) / П. І. Бідюк, В. Д. Романенко, О. Л. Тимошук. – Київ: Політехніка, 2010. – 317 с.
3. Галушка Є. О. Сутність криптовалют та перспективи їх розвитку / Є. О. Галушка, О. Д. Пакон. // Молодий вчений. – 2017. – №4. –634–638 с.
4. Даценко Н. В. Алгоритм моделювання часових рядів криптовалют / Н. В. Даценко, Ю. В. Ігнатова // Моніторинг, моделювання та менеджмент емерджентної економіки: матеріали 7-ої Міжнар. наук.-практ. конф., 23-25 травня 2018 р. Одеса-Черкаси, 2018. 99- 102 с.
5. Даценко Н. В. Застосування дерев класифікації та регресії до прогнозування часових рядів фінансових інструментів. Вчені записки. Сер. Економіко-математичні методи. 2018. № 19. 169–181 с.
6. Дербенцев В. Застосування методів машинного навчання до прогнозування часових рядів криптовалют. / Дербенцев В., Великоіваненко Г., Даценко Н.// Нейро-нечіткі технології моделювання в економіці. 2019. № 8. 65-93 с.
7. Сайт Alpha Vantage. [Електронний ресурс] Режим доступу: <https://www.alphavantage.co> (дата звернення: 26.03.2023)
8. Сайт Bitcoin BTC. [Електронний ресурс]. Режим доступу: <https://bitcoin.org/> (дата звернення: 26.03.2023)
9. Сайт Coin Desk. [Електронний ресурс] Режим доступу: <https://coindesk.com> (дата звернення: 20.03.2023)
10. Сайт Coin Market Cap. [Електронний ресурс] Режим доступу: <https://coinmarketcap.com/currencies/bitcoin> (дата звернення: 20.03.2023)
11. Сайт CoinGecko. [Електронний ресурс]. Режим доступу: <https://www.coingecko.com/en/global-charts> (дата звернення: 20.04.2023)

12. Сайт Ethereum ЕТН. [Электронный ресурс]. Режим доступа: <https://ethereum.org/> (дата звернення: 26.03.2023)
13. Сайт Finance.ua. [Электронный ресурс]. Режим доступа: <https://finance.ua/> (дата звернення: 20.04.2023)
14. Сайт Yahoo Finance. [Электронный ресурс]. Режим доступа: <https://finance.yahoo.com/cryptocurrencies/> (дата звернення: 20.03.2023)
15. Armstrong S. J. Principles of Forecasting: A Handbook for Researchers and Practitioners / Armstrong. – New York: Springer, 2002. – 850 p.
16. Bisgaard S. Time Series Analysis and Forecasting by Example / S. Bisgaard, M. Kulahci. – New Jersey: John Wiley & Sons, 2011. – 400 p.
17. Brauneis, A. Price discovery of cryptocurrencies: Bitcoin and beyond. / Brauneis, A., Mestel, R. // Economics Letters, 2018, 165, 58-61 p.
18. Brockwell P. J. Time Series: Theory and Methods / P. J. Brockwell, R. A. Davis. – New York: Springer, 2009. – 580 p.
19. Casey M.J. The Age of Cryptocurrency: How Bitcoin and the Blockchain Are Challenging the Global Economic Order – London: St. Martin's Press, 2015. – 368 p.
20. Ciaian, P. The economics of bitcoin price formation. / Ciaian, P., M. Rajcaniova, and d. Kancs // Applied Economics, 2016, 48 (19), 1799–1815 p.
21. Cryer J. D. Time Series Analysis: With Applications in R / J. D. Cryer, K. Chan. – New York: Springer, 2010. – 491 p.
22. Hamilton J. D. Time Series Analysis / James Douglas Hamilton. – Princeton: Princeton University Press, 1994. – 820 p.
23. Hotz N. What is CRISP DM?. [Электронный ресурс]. – 2023. – Режим доступа до ресурсу: <https://www.datascience-pm.com/crisp-dm-2/>
24. Hyndman R. J. Forecasting: principles and practice / R. J. Hyndman, G. Athanasopoulos. – Melbourne: OTexts, 2013. – 292 p.
25. Makridakis S. G. Forecasting: Methods and Applications / S. G. Makridakis, S. C. Wheelwright, R. J. Hyndman. – New York: John Wiley & Sons, 2011. – 656 p.

- 26.Sawant R. Bitcoin price prediction using Machine Learning. [Электронный ресурс]. – 2021. – Режим доступа до ресурсу: <https://medium.com/@rohansawant7978/forecasting-of-bitcoin-price-using-machine-learning-deep-learning-techniques-93bf662f46ab>
- 27.Shumway R. H. Time Series Analysis and Its Applications: With R Examples / R. H. Shumway, D. S. Stoffer. – New York: Springer, 2011. – 596 p.
- 28.Wei, W. C. Liquidity and market efficiency in cryptocurrencies. Economics Letters, 2018, 168. 21-24 p.