

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
ІМЕНІ ТАРАСА ШЕВЧЕНКА  
Факультет інформаційних технологій  
Кафедра інтелектуальних технологій

**КВАЛІФІКАЦІЙНА РОБОТА**  
**на здобуття освітнього ступеня «магістр»**  
НА ТЕМУ:

«Інтелектуальний модуль прогнозування криптовалют  
з використанням методів семантичного та технічного  
аналізів»

Галузь знань: 12 «Інформаційні технології»  
Спеціальність: 122 «Комп'ютерні науки»  
Освітньо-наукова програма «Технології штучного інтелекту»

Виконав:  
студент 2 курсу магістратури, групи ТШІ-21

\_\_\_\_\_ Бербер А. С. \_\_\_\_\_  
(ПБ)

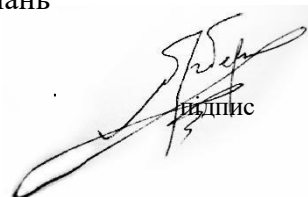
Науковий керівник:

\_\_\_\_\_ Гайна Г. А \_\_\_\_\_  
(ПБ)

професор, кандидат технічних наук  
(науковий ступінь, вчене звання)

Засвідчую, що в цій кваліфікаційній роботі  
немає запозичень з праць інших авторів без  
відповідних посилань

Студент



Відпис

Кваліфікаційна робота допущена до захисту  
рішенням кафедри *інтелектуальних технологій*

Протокол № 12 від «11» травня 2023 р.

Зав. кафедри \_\_\_\_\_ доц. Іларіонов О.Є.  
підпис

Київ 2023

## ЗМІСТ

ВСТУП	5
РОЗДІЛ 1 АНАЛІЗ ЗАДАЧІ ПРОГНОЗУВАННЯ КУРСУ КРИПТОВАЛЮТ	7
1.1	7
1.1.1 Загальні відомості	7
1.1.2 Bitcoin	8
1.2 Аналіз актуальності задачі прогнозування курсу криптовалют	9
1.3	11
1.4	11
1.5 Наявні підходи	13
1.5.1 Фундаментальний аналіз та його використання при роз’язанні задачі прогнозування курсу криптовалют	13
1.5.1.1 Прикладне використання фундаментального аналізу для задачі прогнозування курсу криптовалют	14
1.5.2 Технічний аналіз та його використання при роз’язанні задачі прогнозування курсу криптовалют	15
1.6 Наявні програмні рішення	18
До найбільш популярних рішень відносяться наступні:	18
1.6.1 TradingView	18
1.6.2	19
1.6.3 TensorCharts	20
1.6.4 LunarCRUSH	21
1.6.5 Coinigy	22
1.6.6 CryptoCompare	23

1.6.7 TradeSanta	24
1.7 Постановка задачі	25
1.8 Представлення системи у вигляді чорної скрині	26
1.9 Висновки до першого розділу	27
РОЗДІЛ 2 МАТЕМАТИЧНІ ОСНОВИ ДЛЯ ЗАДАЧІ ПРОГНОЗУВАННЯ КУРСУ КРИПТОВАЛЮТ	28
2.1 Архітектура інтелектуального модуля	28
2.2 Семантичний аналіз настроїв	31
2.2.1 Полярність та семантична оцінка	33
2.3 Задача регресії	34
2.4 Експоненційна ковзна середня	37
2.5 Використання моделі ARIMA для задач регресії	39
2.6 Узагальнена архітектура створюваної системи	41
2.7 Алгоритм скорінгу повідомлень	43
2.8 Проектні рішення	44
2.9 Область застосування та перспективність алгоритму	46
2.10 Практична цінність розробки	46
2.11 Висновки до другого розділу	47
РОЗДІЛ 3. АНАЛІЗ РЕАЛІЗОВАНОГО ІНТЕЛЕКТУАЛЬНОГО МОДУЛЮ ПРОГНОЗУВАННЯ КРИПТОВАЛЮТ З ВИКОРИСТАННЯМ МЕТОДІВ СЕМАНТИЧНОГО І ТЕХНІЧНОГО АНАЛІЗІВ	48
3.1 Інформаційна технологія роботи з системою	48
3.1.2 Підготовка даних	50
3.1.2.1 Підготовка історичних даних ціни криптовалюти	50
3.1.2.2 Підготовка даних повідомлень	50

3.1.3 Особливості пошуку даних для семантичної оцінки.	51
3.2 Тестування розробленого інтелектуального модулю	54
3.3 Порівняння отриманих результатів з результатами використання інших методів	57
3.4	60
ВИСНОВКИ	60
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	61
ДОДАТОК А Код модулю обробки вхідної інформації від користувача	65
ДОДАТОК Б Код модулю для семантичної оцінки повідомлень	67
ДОДАТОК В Код модулю для скорінгової оцінки повідомлень	68
ДОДАТОК Д Код модулю для перевірки репрезентативності скорінгових оцінок	69
ДОДАТОК Е Код модулю для побудови прогнозів ціни криптовалюти на історичних даних	73
ДОДАТОК Є Код модулю для побудови прогнозів ціни криптовалюти на історичних даних та скорінгових оцінках повідомлень	76
ДОДАТОК Ж Зразок вхідних даних про повідомлення	79

## ВСТУП

Сучасний світ фінансів 21 століття постійно розвивається, і криптовалюти стали важливою складовою сучасних фінансових ринків. Криптовалюти відрізняються від традиційних валют своєю децентралізованою структурою, а також високою волатильністю, що створює нові можливості для торгівлі та інвестицій. У зв'язку з цим, створення ефективних та надійних методів прогнозування криптовалют стає особливо актуальним та вимагає використання інноваційних підходів.

Тема кваліфікаційної роботи "Інтелектуальний модуль прогнозування криптовалют з використанням методів семантичного та технічного аналізів" є актуальною, оскільки об'єднує передові методи аналізу фінансових ринків з новітніми технологіями обробки масивів даних та штучного інтелекту.

Метою кваліфікаційної роботи є розробка інтелектуального модуля прогнозування криптовалют, який враховує як семантичний аналіз, так і технічний аналіз для створення більш точних та обґрунтованих прогнозів. Семантичний аналіз полягає у вивченні настроїв користувачів на основі їхніх публікацій у соціальних мережах, таких як Twitter, та виявленні закономірностей, які можуть впливати на динаміку цін криптовалют. Технічний аналіз використовує історичні дані про ціни та обсяги торгівлі для виявлення трендів та патернів, що можуть свідчити про майбутні зміни на ринку.

Для цього розв'язано наступні задачі:

1. Проведено аналіз існуючих систем прогнозування, досліджено їх основні етапи роботи;
2. Розроблено архітектуру системи прогнозування;
3. Реалізовано частини системи у вигляді програмного продукту

Кваліфікаційна робота складається з кількох розділів, де розглядаються теоретичні аспекти семантичного та технічного аналізів в контексті

прогнозування криптовалют. Окремий розділ присвячено практичній реалізації інтелектуального модуля, включаючи збір та попередню обробку даних, побудову моделі та аналіз її ефективності.

У результаті дослідження очікується отримання нових знань щодо можливості використання методів семантичного та технічного аналізів для задачі прогнозуванні криптовалют. Розроблений інтелектуальний модуль матиме практичне значення для зацікавлених сторін, які займаються роботою на ринках криптовалют та криптобіржах, сприяючи підвищенню ефективності їхньої діяльності та підтримці прийняття обґрунтованих рішень.

## РОЗДІЛ 1 АНАЛІЗ ЗАДАЧІ ПРОГНОЗУВАННЯ КУРСУ КРИПТОВАЛЮТ

### 1.1 Криптовалюти

#### 1.1.1 Загальні відомості

Криптовалюта – це цифрова або віртуальна валюта, яка використовує криптографію для забезпечення безпеки та контролю над створенням нових одиниць та проведенням транзакцій. Криптовалюти є альтернативою традиційним фіатним валютам, таким як долари, євро та інші національні валюти, але відрізняються рядом особливостей.[1]

Однією з ключових відмінностей криптовалют від традиційних валют є децентралізація. Більшість криптовалют функціонують на основі блокчейн технології, яка дозволяє створювати децентралізовані мережі без потреби у посередниках, таких як банки або урядові установи. Завдяки цьому, криптовалюти можуть забезпечити більшу фінансову свободу та зменшити потенційні ризики зловживань з боку централізованих установ.

Криптовалюти забезпечують певний рівень анонімності та прозорості для користувачів. Хоча транзакції в блокчейні є відкритими для перегляду, особиста інформація користувачів, така як імена чи адреси, залишається прихованою. Це допомагає забезпечити приватність користувачів та запобігти неправомірному використанню їх особистих даних.[2]

Криптовалюти характеризуються високою волатильністю, що означає значні коливання цін протягом короткого періоду часу. Це створює можливості для криптотрейдерів та інвесторів отримати високі прибутки від спекулятивних операцій, але також значно підвищує ризики збитків. Волатильність криптовалют зумовлена рядом факторів, в тому числі обмежена ліквідність, ринкові настрої, новини та регулятивні рішення.[2]

### 1.1.2 Bitcoin

Bitcoin (BTC) є першою та найбільш відомою криптовалютою, яка була створена у 2009 році. Він поклав початок розвитку криптовалютного ринку. Розробником Bitcoin вважається анонімний автор під псевдонімом Satoshi Nakamoto, який опублікував статтю "Bitcoin: A Peer-to-Peer Electronic Cash System" (Nakamoto, 2008). [1]

Bitcoin є децентралізованою платіжною системою, що дозволяє користувачам здійснювати безпосередні електронні платежі без потреби в посередниках, таких як банки або інші фінансові установи.

Основою функціонування Bitcoin є технологія блокчейн - децентралізований реєстр транзакцій, що зберігається на багатьох комп'ютерах (вузлах) мережі. Блокчейн представляє собою послідовність блоків, кожен з яких містить набір транзакцій. Блоки додаються до блокчейну через процес майнінгу. Завдяки криптографічним методам, блокчейн забезпечує високий рівень безпеки, надійності та прозорості транзакцій. [3]

Майнінг Bitcoin полягає в розв'язанні складних математичних завдань, що дозволяє додавати нові блоки з транзакціями до блокчейна. Майнери, які успішно розв'язують ці завдання, отримують винагороду у вигляді нових одиниць Bitcoin та комісій за проведення транзакцій.

Для забезпечення безпеки та контролю над монетами Bitcoin, користувачі використовують криптографічні ключі: приватний ключ і публічний ключ. Криптогаманці для Bitcoin можуть бути апаратними, програмними, або веб-сервісами. Вони забезпечують зберігання приватних ключів та доступ до балансу Bitcoin, а також дозволяють здійснювати транзакції.[4]

## 1.2 Аналіз актуальності задачі прогнозування курсу криптовалют

Ринок криптовалют був створений як альтернатива традиційним фінансовим ринкам і почав активно розвиватися останніми роками (рис. 1.1). Водночас виникла потреба у прогнозуванні курсу криптовалюти, яке є ключовим моментом для успішних торгів на цьому ринку.

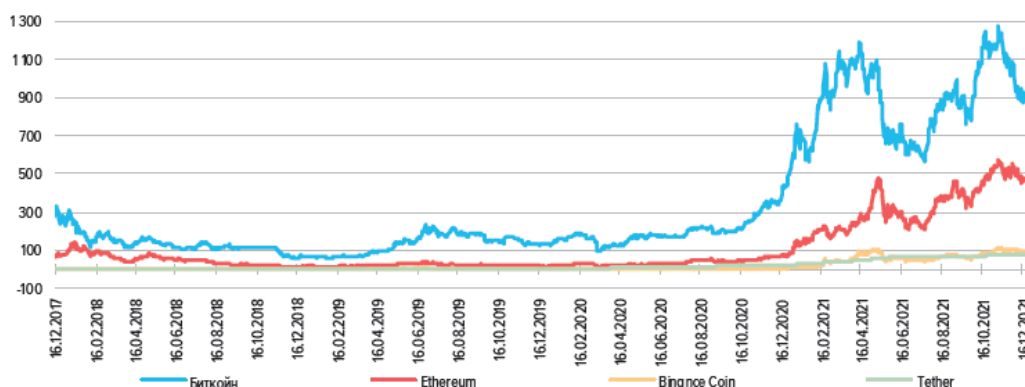


Рисунок 1.1 – Графік капіталізації криптовалют за період 2017-2021 роки

Актуальність завдання прогнозування курсу криптовалюти зумовлена кількома чинниками.

По-перше, криптовалюти стають дедалі популярнішими серед інвесторів і трейдерів, що призвело до зростання їхньої капіталізації та збільшення волатильності на ринку.

По-друге, криптовалюти є новим ринком, на якому немає встановлених правил та стандартів, що ускладнює прогнозування та вимагає від трейдерів глибшого аналізу.

Технічний аналіз є одним із найпоширеніших методів прогнозування курсу криптовалюти. Однак, наявність великої кількості новин, регулятивних змін та інших зовнішніх факторів, які можуть вплинути на курс криптовалюти, створюють складнощі для використання лише технічного аналізу. У цьому випадку семантичний аналіз може бути корисним доповненням для

прогнозування курсу криптовалюти, оскільки він може враховувати зовнішні фактори та зміни на ринку.

Загалом завдання прогнозування курсу криптовалюти залишається актуальним і важливим для трейдерів та інвесторів на цьому ринку. Прогнозування курсу криптовалюти може допомогти трейдерам приймати більш усвідомлені рішення та керувати ризиками своїх позицій. Але при цьому необхідно враховувати, що прогнозування курсу криптовалюти є складним завданням, і трейдери повинні використовувати різноманітні методи аналізу та залишатися настороже з огляду на можливі зміни на ринку.

Далі можна розглянути кілька аспектів, які можуть підтвердити актуальність цього завдання.

По-перше, криптовалюти продовжують привертати увагу як інвесторів, і широкої публіки. Деякі з них стали невід'ємною частиною фінансової екосистеми та використовуються для купівлі товарів та послуг, а інші досі залишаються невідомими для широкої аудиторії. Все це означає, що розуміння того, як зміниться курс криптовалюти у майбутньому, є важливим аспектом для багатьох людей.

По-друге, криптовалюти, як і інші активи, схильні до впливу економічних, політичних та технологічних факторів. Зміни у законодавстві, політичній обстановці та технологічних трендах можуть вплинути на попит на криптовалюту і, отже, на її курс. Використання технічного аналізу може допомогти у визначенні цих факторів та їх впливу на курс криптовалюти.

По-третє, останнім часом спостерігається збільшення кількості людей, які займаються торгівлею криптовалютами. Вони використовують різні технічні індикатори та інші інструменти для прогнозування курсу криптовалюти та прийняття рішень щодо купівлі та продажу. Це демонструє, що завдання прогнозування курсу криптовалюти не лише актуальне, а й має практичне застосування.

По-четверте, криптовалюти продовжують еволюціонувати та змінюватися. Нові технології та інновації в області блокчейну та криптографії можуть вплинути на вартість криптовалюти. Застосування технічного аналізу може допомогти у виявленні нових трендів та прогнозуванні курсу криптовалюти у світлі цих змін.[5]

Таким чином, прогнозування курсу криптовалюти є актуальним і важливим завданням, яке стає все більш затребуваним у зв'язку з популярністю криптовалют, що зростає, і збільшенням числа учасників ринку. Використання технічного аналізу, поєднаного з іншими методами, може допомогти у прогнозуванні курсу криптовалюти та прийнятті обґрунтованих інвестиційних рішень.

### 0.3 Аналіз предметної області, існуючих бізнес-процесів

Областю застосування розроблюваного інтелектуального модулю є прогнозування курсу криптовалюти в рамках торгів на крипторинках або криптобіржах.

Криптові біржі або крипторинки надають можливість користувачам купляти та продавати ту чи іншу обрану криптовалюту. Важливою складовою прийняття рішення на покупку чи продаж криптовалюти є розуміння чи піде ціна вгору чи вниз в майбутньому. Користувач аналізує поведінку ринку та приймає рішення на основі власних знань чи суджень. Після покупки користувач або отримує прибуток, якщо ціна криптовалюти піде вгору чи зазнає збитків, якщо ціна криптовалюти піде вниз. Після продажу користувач недоотримує вигоду, у випадку, якщо ціна зростатиме або убереже себе від втрат якщо ціна падатиме.

### 0.4 Опис профілів зацікавлених сторін

Проведемо аналіз профілів зацікавлених сторін (стейкхолдерів). Для цього виділими наступні 4 показники:

1. Здобута вигода
2. Неявні потреби
3. Основні інтереси
4. Обмеження

Результати аналізу наведені в таблиці 1.1

Таблиця 1.1 - Профілі зацікавлених сторін (стейкхолдерів )

Зацікавлені сторони (стейкхолдери)	Здобута вигода	Неявні потреби	Основні інтереси	Обмеження
Користувачі криптобірж	Прибуток.	Зменшення витрат ресурсів на прогнозування	Точність прогнозування на достойному рівні Простота експлуатації	Немає
Криптобіржі	Стабільність роботи Контроль навантаження на систему	Прибуток від махінацій	Широка сфера застосування Універсальність	Системні вимоги Наявність ресурсів
Криптоінвестори	Прибуток.	Зменшення витрат ресурсів на прогнозування	Точність прогнозування на достойному рівні Максимізація прибутків	Немає
Криптоаналітики	Полегшення праці	Економія часу Зменшення витрат ресурсів на прогнозування	Простота експлуатації Універсальність	Немає

Розробник	Підвищення кваліфікації Збільшення портфоліо	Наукове дослідження	Точність прогнозування не гірша ніж аналоги	Обмеженість часу розробки
-----------	--	---------------------	---	---------------------------

## 1.5 Наявні підходи

### 1.5.1 Фундаментальний аналіз та його використання при роз'язанні задачі прогнозування курсу криптовалют

Фундаментальний аналіз - це метод оцінки активів, який включає дослідження економічних, фінансових та інших чинників, які можуть вплинути на їх вартість. Метою фундаментального аналізу є визначення внутрішньої вартості активу та порівняння його з поточною ринковою вартістю. Якщо внутрішня вартість активу виявляється вищою за ринкову, актив може бути вважається недооціненим, а якщо нижчою - переоціненим.

Вартість віртуальних монет встановлюється ринковим шляхом. Чим більше попит на певну криптовалюту, тим вище її курс. Попит же в свою чергу залежить від тих переваг, які пропонує монета. Попит формується на тлі новин, нових розробок, анонсів компаній. Популярність криптовалют, реклама та хороші новини також є ознакою швидкого зростання курсу Bitcoin або будь-якого іншої криптовалюти. Чим більше людей знає про товар, тим більше людей захочуть вкласти в нього свої гроші або грати на біржі.

При аналізі криптовалют використовуються наступні основні компоненти:

Економічні чинники: включають аналіз ринкової капіталізації криптовалют, обсягу торгівлі та динаміки інфляції.

Фінансові чинники: аналізуються фінансові показники проекту, такі як прибутковість, витрати та прибуток на одну монету.

Технічні чинники: вивчаються технологічні аспекти, такі як масштабованість, швидкість транзакцій, рівень безпеки та консенсусний механізм.

Соціальні чинники: включають аналіз спільноти та активності розробників, партнерств та конкуренції.

Також бувають спекулятивні стрибки в курсі - так звані «дампи» і «пампи». Їх провокують великі трейдери на біржах з метою заробітку. Їх не особливо цікавить, скільки коштує монета. За рахунок великої кількості активів, вони здатні створювати короткостроковий вплив на курс.[6]

#### 1.5.1.1 Прикладне використання фундаментального аналізу для задачі прогнозування курсу криптовалют

Курс криптовалют залежить виключно від попиту людей, то одним зі способів прогнозування курсу є кількісна оцінка цього попиту та подальший аналіз впливу попиту на ціну криптовалюти.

Висновки про попит і популярність тієї чи іншої криптовалюти можна зробити використовуючи дані про те, як часто дана тема обговорюється у інтернеті, наприклад, один з таких показників – кількість повідомлень зі згадуванням відповідної криптовалюти у соціальній мережі.



Рисунок 1.2 – Приклад повідомлення з соціальної мережі Twitter зі згадуванням криптовалюти

З переваг цього показнику можна виділити наступні:

- є кількісним;
- технічно доступний для отримання чи обчислення;
- просто використовувати у математичних моделях.

Серед недоліків можна виділити наступні:

- доступ платний;
- доступ обмежений по сумарній кількості отриманої інформації та швидкості отримання (кількість запитів у хвилину).

Виходячи з наведеного аналізу і суттєвих недоліків аналогів у даній роботі використовується саме такий підхід.[44]

1.5.2 Технічний аналіз та його використання при роз’язанні задачі прогнозування курсу криптовалют

Криптовалюти є одним із найбільш волатильних ринків у світі. Це означає, що курси криптовалют можуть швидко змінюватись у відповідь на безліч факторів, таких як новини, регулювання та зміна попиту та пропозиції. У цій ситуації трейдери часто використовують технічний аналіз для прогнозування курсу криптовалют та прийняття обґрунтованих рішень щодо купівлі та продажу.

Технічний аналіз – це метод аналізу ринку, який ґрунтується на вивченні історичних цінових даних, щоб визначити тенденції на ринку та прогнозувати майбутні цінові рухи. Він використовує графічні та числові інструменти, такі як графіки цін, ковзні середні, індикатори та технічні осцилятори (рис. 1.3) [7,8,9]

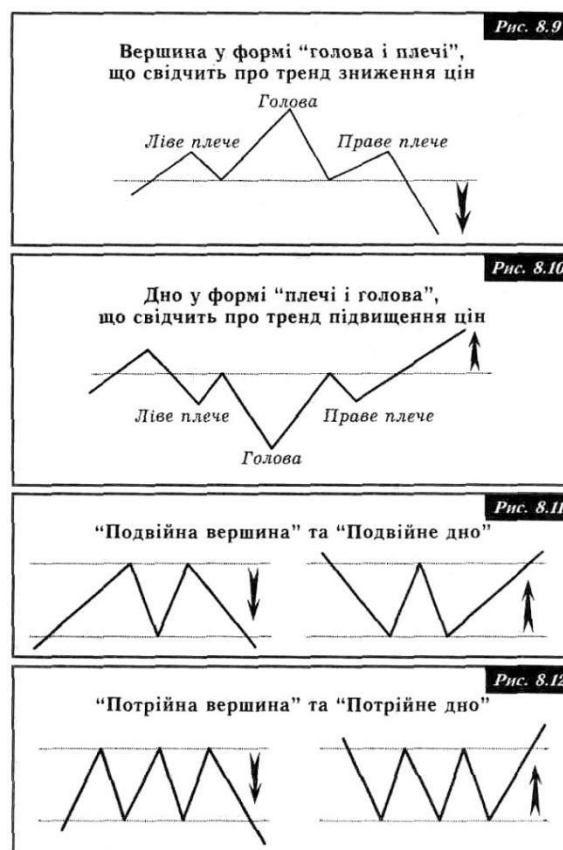


Рисунок 1.3 – Зображення прикладу формалізації курсу криптовалюти в технічному аналізі

Технічний аналіз може бути ефективним інструментом прогнозування курсу криптовалют. Одним із найпоширеніших методів технічного аналізу є

аналіз тренду, який використовується для визначення напрямку руху ціни криптовалюти. Він ґрунтується на вивченні графіка цін, щоб визначити, в якому напрямку рухається курс.[7]

Інший метод, який використовується в технічному аналізі - це аналіз підтримки та опору (рис. 1.4). Він використовується для визначення рівнів цін, на яких курс криптовалюти може змінити свій напрямок. Трейдери можуть використовувати ці рівні, щоб визначити точки входу та виходу з ринку. [10]

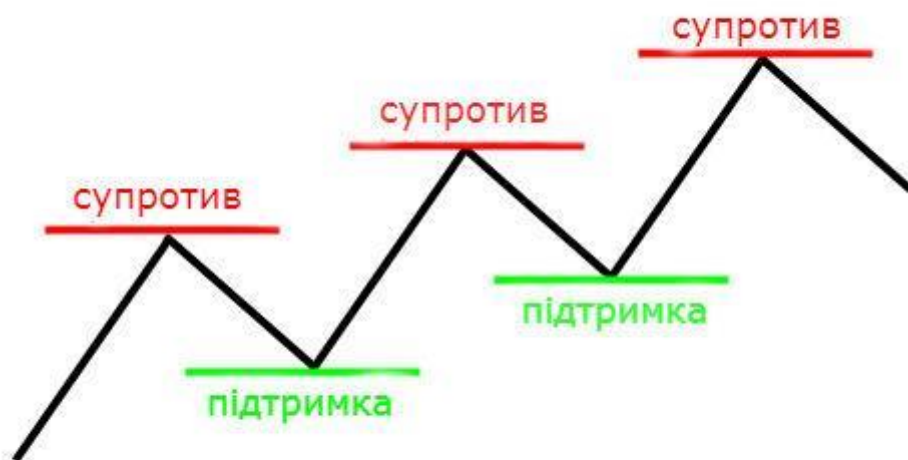


Рисунок 1.4 – Схематичне зображення рівнів підтримки та опору в технічному аналізі

Існує також ряд технічних індикаторів, таких як середні ковзаючі, індикатори сили тренду та індикатори обсягу, які можуть бути використані для прогнозування курсу криптовалют. Вони можуть допомогти трейдерам визначити моменти, коли курс криптовалюти може змінити свій напрямок і піти вгору або вниз.[11,12]

Інший важливий аспект технічного аналізу – це управління ризиками. Трейдери можуть використовувати технічний аналіз для визначення рівнів стоп-лосса та рівнів профіт-тейка, щоб керувати своїми ризиками та захистити свої позиції від великих втрат.

Використання технічного аналізу для прогнозування курсу криптовалют може бути ефективним, якщо він застосовується правильно та з обережністю. Трейдери повинні враховувати безліч факторів, таких як новини та регулятивні зміни, які можуть вплинути на курс криптовалюти.[13,14,15,16,17]

## 1.6 Наявні програмні рішення

Існує досить багато програмних рішень, які вирішують представлену проблему, проте більшість з них мають недоліки як от платне використання, неуніверсальність, складність систем, ресурсозатратність програмного забезпечення.

До найбільш популярних рішень відносяться наступні:

1. TradingView
2. Santiment
3. TensorCharts
4. LunarCRUSH
5. Coinigy
6. CryptoCompare
7. TradeSanta

В наступних підрозділах будуть виділені особливості платформи, основні переваги та недоліки.

### 1.6.1 TradingView

TradingView — це популярна платформа для огляду ринку активів, яка надає розширені інструменти технічного аналізу та можливості прогнозування для різних фінансових ринків, у тому числі криптовалют. Вона містить велику

бібліотеку технічних індикаторів, інструментів для малювання та налаштування макетів діаграм.[20]

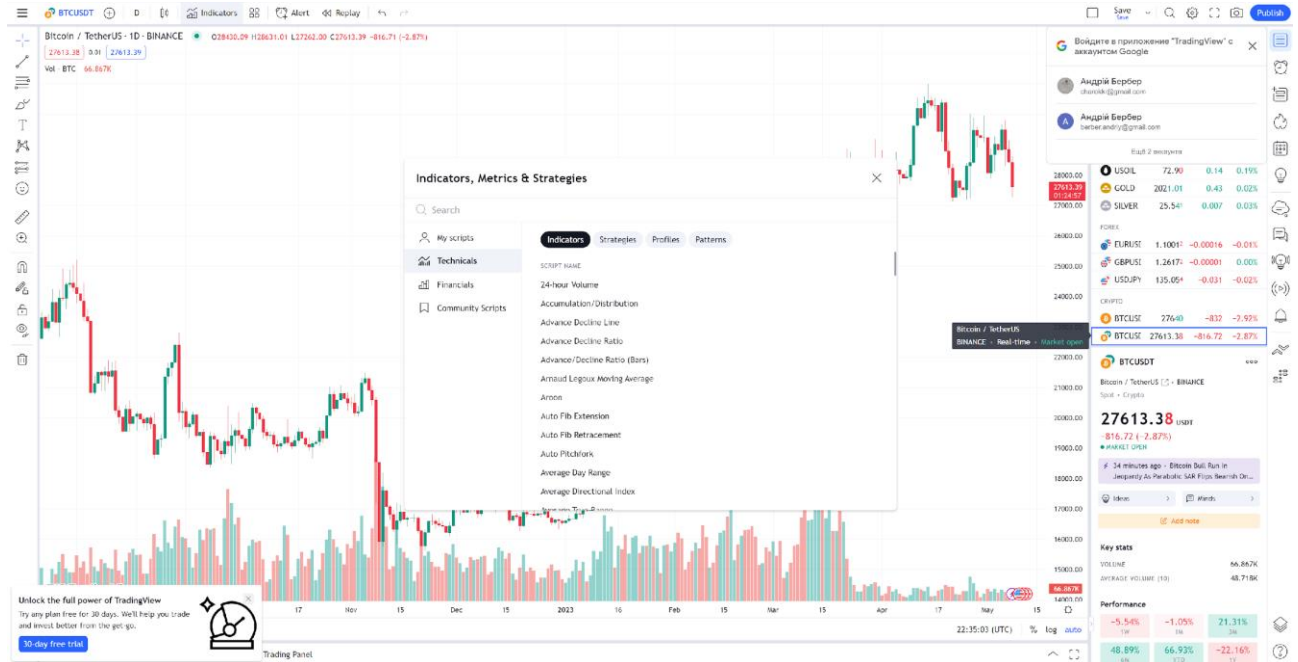


Рисунок 1.5 – Ілюстрація інтерфейсу платформи TradingView

Переваги:

1. Найбільша спільнота користувачів.
2. Розширені інструменти для побудови діаграм, аналізу та прогнозування зі зручним інтерфейсом.
3. Підтримує велику кількість крипто та фінансових ринків.

Недоліки:

1. Обмежена кількість методів прогнозування, відсутність спеціалізованих методів.
2. Платна.

## 1.6.2 Santiment

Santiment — це платформа для аналізу ринку, яка поєднує великий спектр даних, для надання інформації та прогнозів для ринків криптовалют. Вона

пропонує різноманітні інструменти, у тому числі аналіз настроїв, відстеження зростання криптомережі та моніторинг активності торгів.[21]



Рисунок 1.6 – Ілюстрація інтерфейсу платформи Santiment

Переваги:

1. Комплексні джерела даних для поглибленого аналізу.
2. Унікальні метрики та індикатори, яких немає на інших платформах.
3. Активна спільнота та ресурси підтримки.

Недоліки:

1. Складний інтерфейс.
2. Більшість інструментів прогнозування платні, а ті що не платні мають погану точність.

### 1.6.3 TensorCharts

TensorCharts — це відносно нова платформа для створення діаграм і аналізу, яка використовує алгоритми машинного навчання для прогнозування

курсу цін криптовалют. Вона пропонує широкий спектр інструментів, в тому числі торгові сигнали та технічні індикатори з можливістю тонкого налаштування.[22]

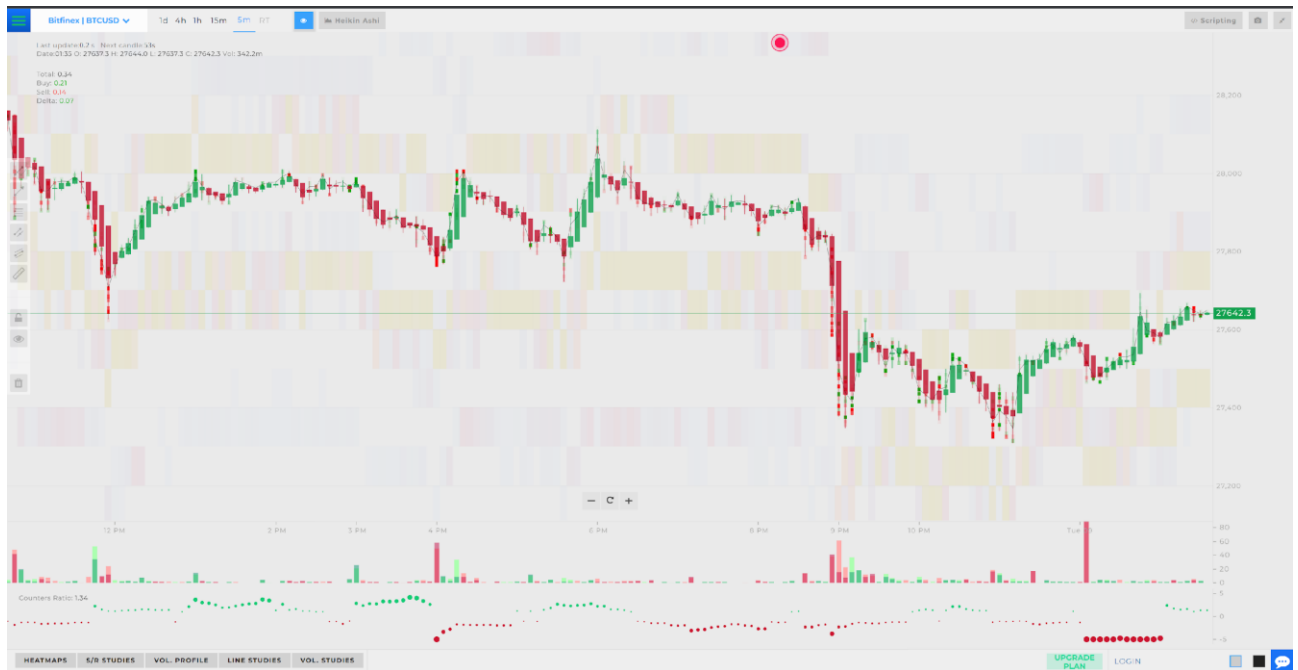


Рисунок 1.7 – Ілюстрація інтерфейсу платформи TensorCharts

Переваги:

1. Прогнозування на основі машинного навчання.
2. Налаштовані інструменти та індикатори для створення діаграм.
3. Ринкові дані в реальному часі для точного прогнозування.

Недоліки:

1. Високий поріг входу
2. Платна.
3. Відсутність використання аналізу настроїв, тобто повна відірваність від зовнішнього світу.

#### 1.6.4 LunarCRUSH

LunarCRUSH — це платформа, яка поєднує аналітику соціальних мереж з традиційними методами технічного аналізу, щоб надати унікальну інформацію та прогнози для курсу криптовалют. Вона досліджує соціальний вплив на різні криптовалюти для передбачення руху цін.[23]

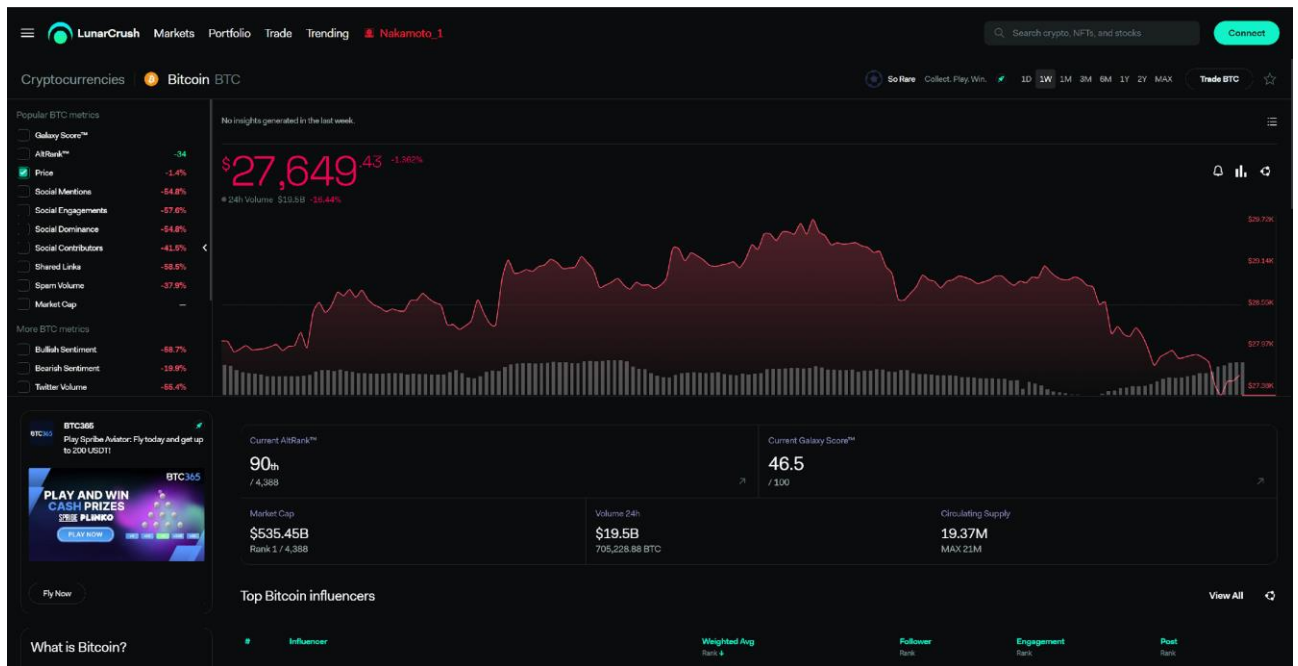


Рисунок 1.8 – Ілюстрація інтерфейсу платформи LunarCRUSH

Переваги:

1. Підхід до аналізу ринку з використанням даних із соціальних мереж.
2. Простий інтерфейс.

Недоліки:

1. Використання в основі даних з соціальних мереж може бути менш надійним, ніж традиційні ринкові дані.
2. Обмежена функціональність порівняно з більш комплексними платформами.
3. Платна

## 1.6.5 Coinigy

Coinigy — це торгова платформа професійного рівня для управління портфелем, яка пропонує вдосконалені інструменти для створення діаграм, технічного аналізу та прогнозування для ринків криптовалют. Вона інтегрується з понад 45 біржами та забезпечує уніфікований досвід торгівлі.[24]

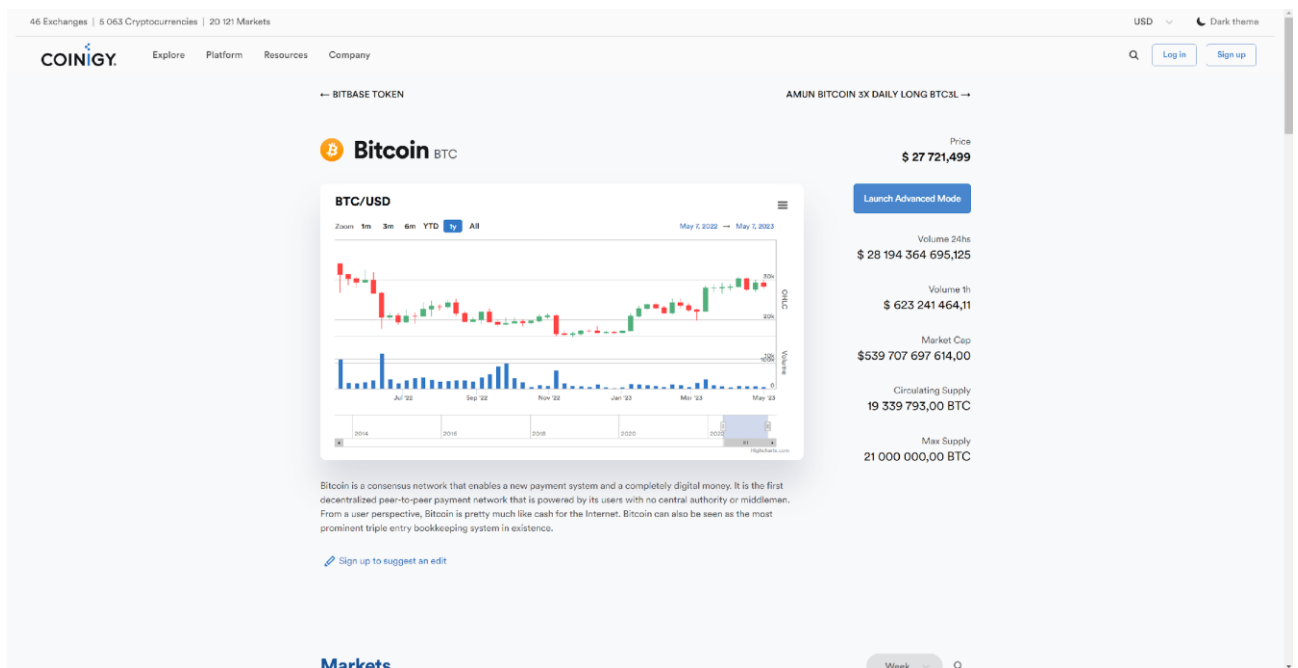


Рисунок 1.9 – Ілюстрація інтерфейсу платформи Coinigy

Переваги:

1. Комплексний технічний аналіз і інструменти для створення діаграм.
2. Повна інтеграція з багатьма біржами.
3. Кросплатформенна

Недоліки:

1. Платна, одна з найдорожчих.
2. Дуже обмежений доступ до розширених функцій у планах нижчого рівня, в тому числі до модулів прогнозування.

## 1.6.6 CryptoCompare

CryptoCompare — це комплексна платформа аналізу ринку, яка надає дані про криптовалюту, аналітику та інструменти дослідження. Вона збирає інформацію з різних джерел, для генерації прогнозів щодо широкого спектру криптовалют.[25]

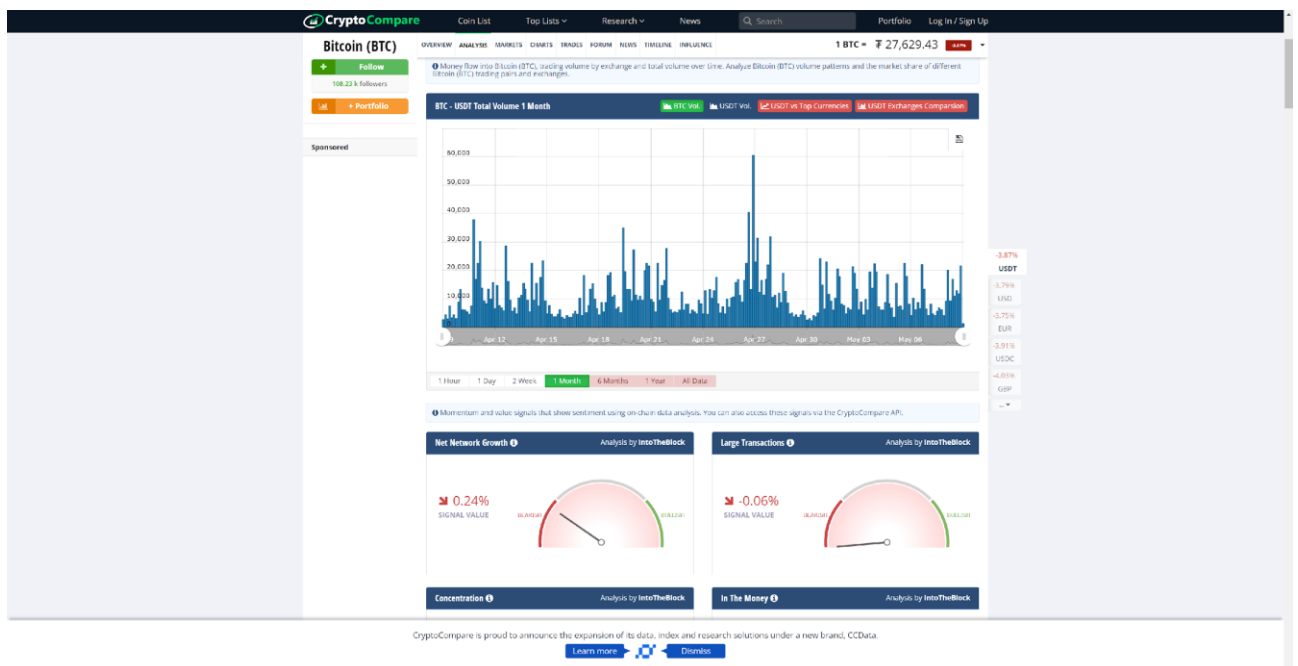


Рисунок 1.10 – Ілюстрація інтерфейсу платформи CryptoCompare

Переваги:

1. Велике охоплення даних з багатьох джерел.
2. Налаштовані та інтерактивні діаграми.
3. Є достойний безкоштовний пакет інструментів.

Недоліки:

1. Деякі просунуті функції доступні лише в платних планах.
2. Мала кількість методів прогнозування, низька точність

## 1.6.7 TradeSanta

TradeSanta — це хмарна платформа, яка пропонує автоматизовані торгові рішення для ринків криптовалют. Вона містить різноманітні технічні індикатори та настроювані торгові стратегії, що дозволяє користувачам створювати індивідуальні прогнози на основі історичних даних і даних у реальному часі.[26]

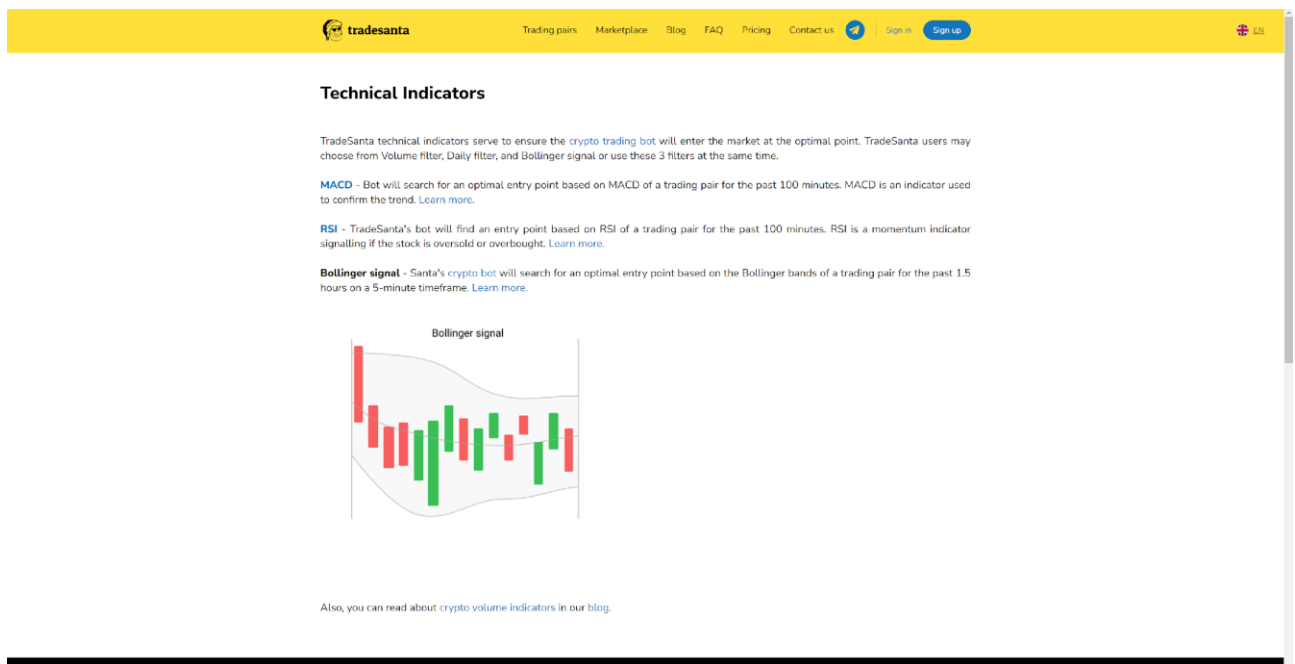


Рисунок 1.11 – Ілюстрація інтерфейсу платформи TradeSanta

Переваги:

1. Простий інтерфейс.
2. Настроювані стратегії та торгові боти.
3. Підтримує основні біржі криптовалют.

Недоліки:

1. Обмежені функції поглибленого аналізу.
2. Платна.
3. російський виробник

## 1.7 Постановка задачі

Об'єкт дослідження кваліфікаційної роботи – процес прогнозування криптовалют на прикладі біткоїна.

Предмет дослідження кваліфікаційної роботи – методи прогнозування криптовалют з застосуванням технологій штучного інтелекту.

Мета: розробка інтелектуального модулю для прогнозування криптовалют з поєднанням методів технічного та семантичного аналізів.

Функціональні вимоги:

1. Обробка повідомлень, що стосуються обраної користувачем криптовалюти
2. Оцінка повідомлень на основі семантичних показників та супутньої метаінформації.
3. Прогнозування курсу криптовалюти, обраної користувачем.

Нефункціональні вимоги:

1. Інтелектуальний модуль повинен давати досить точні дані для прогнозування для більш точних торгів.

Задача, яка розв'язується у випускній кваліфікаційній роботі – це дослідження методів прогнозування криптовалют на основі історичних даних і семантичної оцінки повідомлень по темі в соціальних мережах та підвищення точності прогнозування шляхом створення моделі, на базі кількох існуючих. Створена модель буде оцінена за критеріями, ці оцінки будуть використані для аналізу її ефективності та порівняння цих показників для різних моделей.

Має бути програмний продукт, у результаті роботи якого користувач зможе надати на вхід історичні дані про курс криптовалюти та дані про повідомлення на тему обраної криптовалюти і отримати результат прогнозування для обраної криптовалюти в заданий проміжок часу.

## 1.8 Представлення системи у вигляді чорної скрині

Далі на рисунку 1.5 наведене представлення системи у вигляді чорної скрині



Рисунок 1.12 – Представлення системи у вигляді чорної скрині

## 1.9 Висновки до першого розділу

В ході аналізу задачі прогнозування курсу криптовалют було оглянуто наукові роботи на науково-популярні статті присвячені підходам до прогнозування криптовалют. Спершу було проведено аналіз актуальності задачі прогнозування курсу криптовалют, в ході якого було доведено актуальність досліджуваної задачі. Було проведено огляд фундаментального аналізу, виділено його переваги й недоліки. Було проаналізовано засади технічного аналізу, сфери його застосування, особливості використання в задачі прогнозування курсу криптовалют. .

Базуючись на проведеному дослідженні було сформульовано об'єкт дослідження, предмет дослідження, мету, функціональні та нефункціональні вимоги, а також представлено досліджувану систему у вигляді чорної скрині.

## РОЗДІЛ 2 МАТЕМАТИЧНІ ОСНОВИ ДЛЯ ЗАДАЧІ ПРОГНОЗУВАННЯ КУРСУ КРИПТОВАЛЮТ

### 2.1 Архітектура інтелектуального модуля

Інтелектуальний модуль складається з наступних підмодулів:

1. Модуль обробки вхідної інформації від користувача, який отримує на вхід дані, обробляє їх, викликає наступні модулі
2. Модуль для семантичної оцінки повідомлень, який аналізує вхідні повідомлення, оцінює кожне з них.
3. Модуль для скорінгової оцінки повідомлень, який аналізує вхідні повідомлення та їх метадані, будує комплексну оцінку значимості повідомлень.
4. Модуль для перевірки репрезентативності скорінгових оцінок, який шукає кореляцію скорінгових оцінок і історичних даних ціни криптовалюти.
5. Модуль для побудови прогнозів ціни криптовалюти на історичних даних.
6. Модуль для побудови прогнозів ціни криптовалюти на історичних даних та скорінгових оцінках повідомлень.

Схема зв'язків системи відображено на рисунку 2.1.

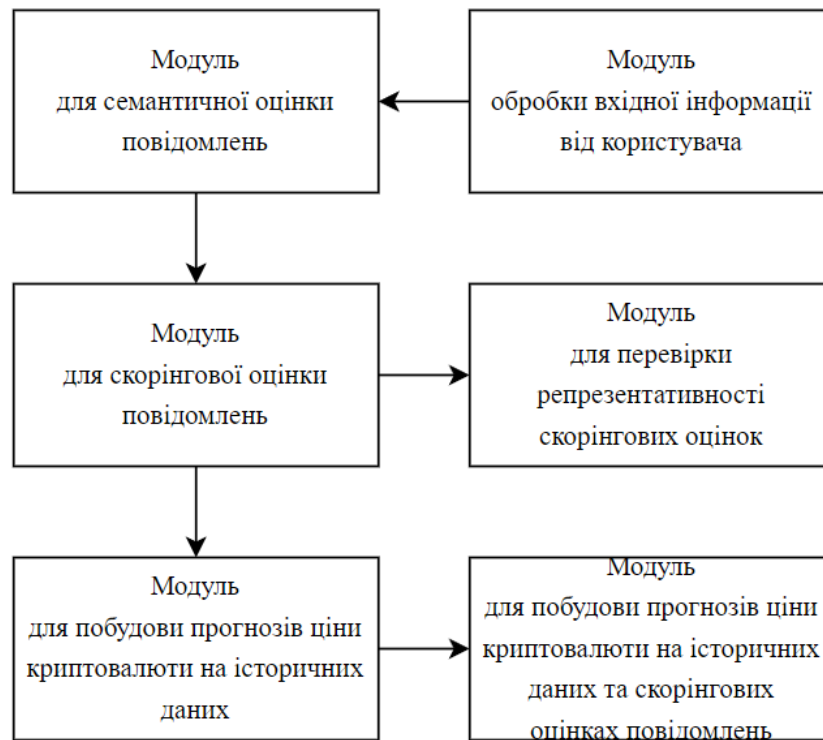


Рисунок 2.1 – Схема зв'язків системи

Нижче наведено основні функції системи та їх короткий опис.

1. Функція обробки історичних даних ціни криптовалюти

Функція отримує файл з історичними даними ціни криптовалюти, чистить та будує структуру даних датафрейм для подальшої обробки.

Вхідні дані: файл з історичними даними ціни криптовалюти

Вихідні дані: очищений датафрейм

2. Функція обробки даних про повідомлення на тему вибраної криптовалюти

Функція отримує файл з даними про повідомлення на тему вибраної криптовалюти, чистить та будує структуру даних для подальшої обробки.

Вхідні дані: файл з даними про повідомлення на тему вибраної криптовалюти

Вихідні дані: очищений датафрейм

3. Функція підгонки вхідних даних

Функція отримує датасети з попередніх функцій, задає проміжок часу, підганяє дані по часу

Вхідні дані: датафрейми з попередніх функцій

Вихідні дані: об'єднаний датафрейм з даними для прогнозування

4. Функція надання семантичних оцінок повідомленням

Функція надає семантичні оцінки повідомленням

Вхідні дані: датафрейм з повідомленнями

Вихідні дані: датафрейм з семантичними оцінками повідомлень

5. Функція надання скорінгової оцінки повідомленням на основі семантичних оцінок та метаданих про повідомлення

Функція надає скорінгову оцінку повідомленням на основі семантичних оцінок та метаданих про повідомлення

Вхідні дані: датафрейм з семантичними оцінками повідомлень

Вихідні дані: датафрейм з скорінговими оцінками повідомлень

6. Функція нормалізації скорінгових оцінок

Функція нормалізує скорінгові оцінки повідомлень

Вхідні дані: датафрейм з скорінговими оцінками повідомлень

Вихідні дані: датафрейм з нормалізованими скорінговими оцінками повідомлень

7. Функція перевірки кореляції семантичних оцінок повідомлень і історичних даних ціни криптовалюти

Функція будує графіки кореляцій семантичних оцінок повідомлень і історичних даних ціни криптовалюти для надання інформації про репрезентативність семантичних оцінок повідомлень

Вхідні дані: датафрейм з скорінговими оцінками повідомлень

Вихідні дані: графіки кореляцій

8. Функція прогнозування на основі історичних даних

Функція займається прогнозуванням ціни криптовалюти на основі історичних даних використовуючи авторегресійну модель ARIMA

Вхідні дані: датафрейм історичними даними ціни криптовалюти, заданий проміжок часу

Вихідні дані: прогнози ціни криптовалюти на основі історичних даних

9. Функція прогнозування на основі поєднання скорінгових оцінок і прогнозування на історичних даних.

Функція займається прогнозуванням ціни криптовалюти на основі історичних даних використовуючи авторегресійну модель ARIMA

Вхідні дані: прогнози ціни криптовалюти на основі історичних даних, скорінгові оцінки повідомлень

Вихідні дані: прогнози ціни криптовалюти на основі історичних даних та скорінгових оцінок

## 2.2 Семантичний аналіз настроїв

Семантичний аналіз (також відомий як аналіз настроїв) — це використання обробки природної мови, аналізу тексту, обчислювальної лінгвістики та біометричних даних для систематичного виявлення, виділення, кількісної оцінки та вивчення афективних станів та суб'єктивної інформації. Аналіз настроїв широко застосовується для озвучування матеріалів клієнтів, таких як огляди та відповіді на опитування, онлайн- та соціальні медіа, а також матеріали для охорони здоров'я для додатків, які варіюються від маркетингу до обслуговування клієнтів і клінічної медицини. [27]

Основним завданням аналізу настроїв є класифікація полярності даного тексту на рівні документа, речення або ознаки/аспекту — незалежно від того, чи є висловлена думка в документі, реченні чи ознакою/аспектом сутності позитивною, негативною чи нейтральною. Розширена класифікація настроїв «за межами полярності» розглядає, наприклад, емоційні стани, такі як насолода, гнів, огида, смуток, страх і здивування.

Незважаючи на те, що в більшості методів статистичної класифікації нейтральний клас ігнорується за припущення, що нейтральні тексти лежать поблизу кордону бінарного класифікатора, кілька дослідників припускають, що, як і в кожній задачі полярності, необхідно ідентифікувати три категорії. Більше того, можна довести, що конкретні класифікатори, такі як Max Entropy і SVMs, можуть отримати користь від введення нейтрального класу та покращити загальну точність класифікації. В принципі існує два способи роботи з нейтральним класом. Алгоритм або спочатку визначає нейтральну мову, відфільтровуючи її, а потім оцінюючи решту з точки зору позитивних і негативних настроїв, або створює тристоронню класифікацію за один крок. Цей другий підхід часто передбачає оцінку розподілу ймовірностей за всіма категоріями (наприклад, наївні байєсівські класифікатори, реалізовані NLTK). Чи і як використовувати нейтральний клас, залежить від природи даних: якщо дані чітко групуються на нейтральну, негативну та позитивну мову, має сенс відфільтрувати нейтральну мову та зосередитися на полярності між позитивними та негативними настроями. Якщо, навпаки, дані переважно нейтральні з невеликими відхиленнями в сторону позитивного та негативного впливу, ця стратегія ускладнить чітке розмежування двох полюсів.[28][29]

Іншим методом визначення настроїв є використання системи масштабування, за допомогою якої слова, які зазвичай асоціюються з негативним, нейтральним або позитивним настроєм, отримують відповідне число за шкалою від  $-1$  до  $1$  (від найбільш негативного до найбільш позитивного). Це дає змогу коригувати настрої даного терміна щодо його середовища (зазвичай на рівні речення). Коли фрагмент неструктурованого тексту аналізується за допомогою обробки природною мовою, кожному концепту у зазначеному середовищі присвоюється оцінка на основі того, як слова сентименту пов'язані з поняттям

та пов'язаною з ним оцінкою. Це дозволяє рухатися до більш витонченого розуміння настроїв, оскільки тепер можна налаштувати значення настрою концепції відносно модифікацій, які можуть його оточувати. Наприклад, слова, які посилюють, розслаблюють або заперечують почуття, виражені поняттям, можуть вплинути на його оцінку. Крім того, тексти можуть отримати позитивні та негативні оцінки сили настроїв, якщо мета полягає в тому, щоб визначити настрої в тексті, а не загальну полярність і силу тексту.[30]

Існують різні типи аналізу настроїв, як-от аналіз настроїв на основі аспектів, аналіз оцінок (позитивний, негативний, нейтральний), багатомовний аналіз настроїв та виявлення емоцій.

### 2.2.1 Полярність та семантична оцінка

Полярність - це міра ступеня, до якого певний текст виражає позитивне чи негативне почуття. Зазвичай воно представлене як безперервне значення від -1 до 1, де -1 вказує на різко негативні настрої, 1 вказує на сильно позитивні настрої, а 0 позначає нейтральні настрої. Оцінка полярності тексту може бути обчислена за допомогою різних математичних моделей і алгоритмів. Одним із простих підходів до розрахунку полярності є наступна формула [31]:

$$Polarity = \frac{N_{positive} - N_{negative}}{N_{total}} \quad (2.1)$$

Де  $N_{positive}$  – кількість позитивних слів,

$N_{negative}$  – кількість негативних слів,

$N_{total}$  – загальна кількість слів

Однак ця формула може не врахувати нюанси тексту, оскільки вона покладається виключно на кількість слів і не враховує контекст.

Семантична оцінка (оцінка настроїв) – це більш повний показник, який враховує як полярність, так і інтенсивність настроїв, виражених у певному тексті. Зазвичай і вона коливається від -1 до 1, подібно до полярності.

Поширеним підходом до обчислення оцінки настрою є використання лексиконів настрою, які призначають оцінку настрою кожному слову в тексті на основі його сприйнятої сили настрою. Оцінку настрою тексту можна розрахувати за такою формулою[32][33][34]:

$$\textit{Sentiment score} = \frac{\sum_{k=0}^N S_k}{N} \quad (2.2)$$

Де  $S_k$  – семантична оцінка слова,

$N$  – кількість слів у тексті

### 2.3 Задача регресії

У статистичному моделюванні регресійний аналіз являє собою набір статистичних процесів для оцінки співвідношень між змінними. Він включає в себе багато методів для моделювання та аналізу кількох змінних, коли фокус робиться на взаємозв'язку між залежною змінною та однією чи більше незалежними змінними. Більш конкретно, регресійний аналіз допомагає зрозуміти, як змінюється типове значення залежної змінної (або "змінної критерію"), коли будь-яка з незалежних змінних змінюється, тоді як інші незалежні змінні фіксуються.

Найчастіше регресійний аналіз оцінює умовне очікування залежної змінної за незалежними змінними, тобто середнє значення залежної змінної, коли незалежні змінні фіксуються. Рідше, основна увага приділяється квантилю або

іншому параметру розташування умовного розподілу залежної змінної за незалежними змінними. У всіх випадках повинна бути оцінена функція незалежних змінних, що називаються функцією регресії. У регресійному аналізі також представляє інтерес для характеристики варіації залежної змінної навколо передбачення функції регресії з використанням розподілу ймовірності. Приклад схематичного зображення регресійного аналізу зображений на рисунку 2.1.

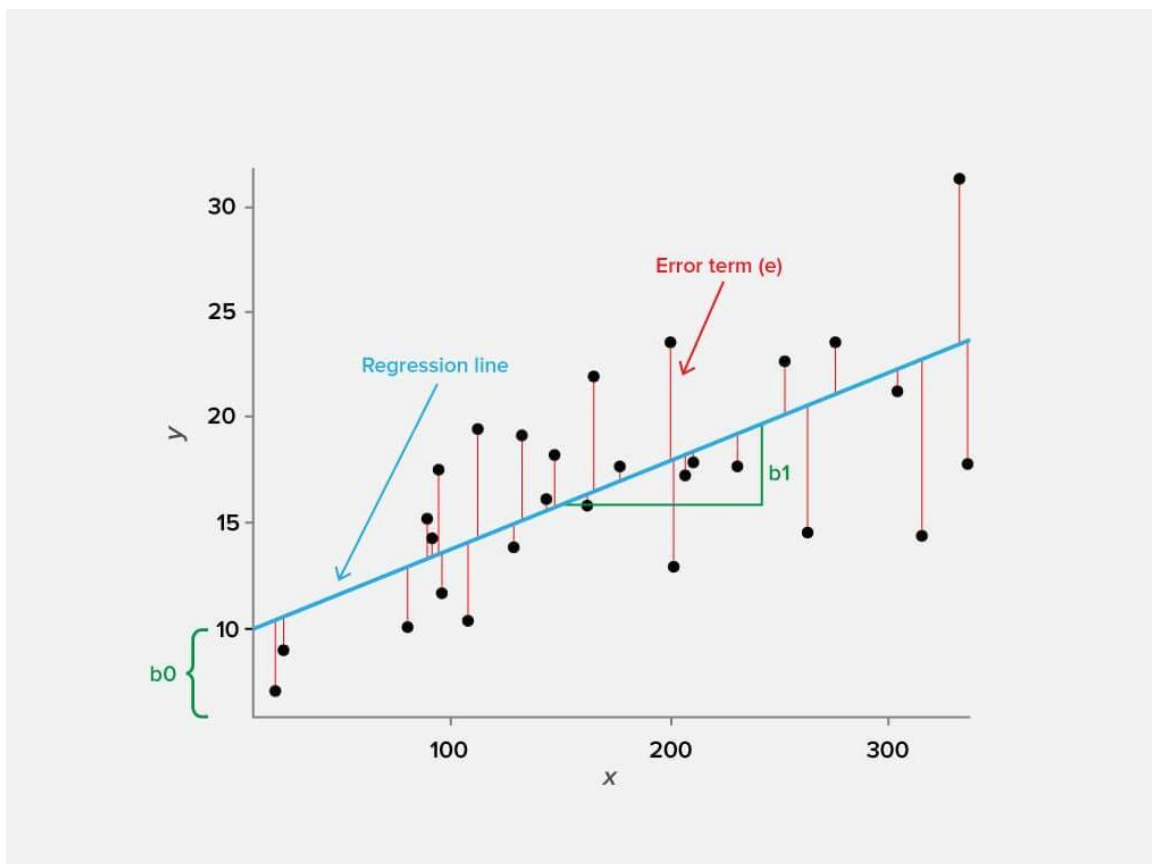


Рисунок 2.2 – Схематичне зображення елементів регресійного аналізу

Регресійний аналіз широко використовується для прогнозування, де його використання суттєво збігається з областю машинного навчання. Аналіз регресії також використовується для того, щоб зрозуміти, які з незалежних змінних пов'язані із залежною змінною, а також досліджувати форми цих відносин. У обмежених умовах регресійний аналіз може бути використаний для виявлення причинних зв'язків між незалежними та залежними змінними. Однак це може

призвести до ілюзій або помилкових відносин, наприклад, співвідношення не підтверджує причинності.

Формула регресії для кількох незалежних змінних виглядає наступним чином:

$$y = \beta_0 + \sum_{i=1}^n \beta_i * x_i + \varepsilon \quad (2.3)$$

Де  $y$  – залежна змінна,

$x_i$  – незалежні змінні,

$\beta_0$  – точка перетину,

$\beta_i$  – коефіцієнти регресії, що представляють вплив незалежних змінних на залежну,

$\varepsilon$  – випадкова помилка

Метою регресійного аналізу є оцінка коефіцієнтів регресії  $\beta_i$ , які мінімізують суму квадратів різниць між спостережуваними та прогнозованими значеннями залежної змінної  $y$ .

Розроблено багато методів проведення регресійного аналізу. Знайомі методи, такі як лінійна регресія та звичайна регресія найменших квадратів, є параметричними, оскільки функція регресії визначається в термінах кінцевого числа невідомих параметрів, які оцінюються за даними. Непараметрична регресія відноситься до методів, що дозволяють функціонувати регресійні функції в заданому наборі функцій, які можуть бути нескінченновимірними.[35]

## 2.4 Експоненційна ковзна середня

Експоненційна ковзна середня (ЕМА) - це статистичний інструмент, який використовується для аналізу та прогнозування цін на криптовалюти, а також на інші активи. Вона працює на основі усереднення цінових даних за певний період часу та враховує свіжіші дані, ніж проста ковзна середня.

Проста ковзна середня (SMA) обчислюється шляхом усереднення цінових даних за певний період часу і може бути корисною для визначення тренду на ринку. Однак вона має недолік - вона однаково зважує всі дані за певний період часу, включаючи старіші дані, які можуть бути менш релевантними для поточного стану ринку. Приклад використання ЕМА зображений на рисунку 2.2.

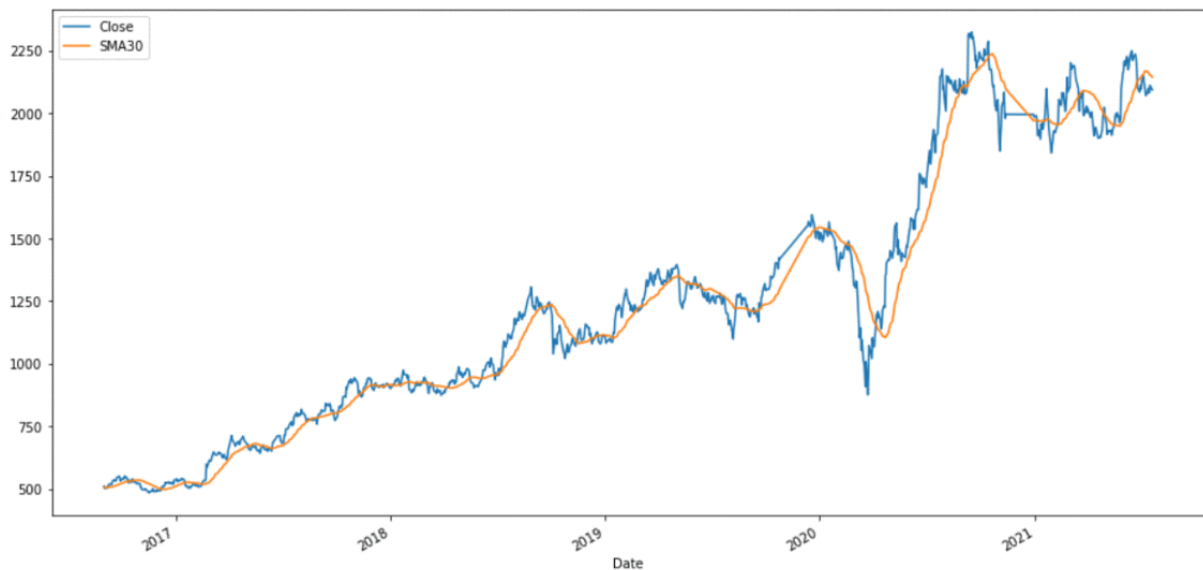


Рисунок 2.3 – Приклад використання ЕМА

ЕМА вирішує цю проблему, ввівши коефіцієнт згладжування, який віддає більшу вагу свіжішим даним, що робить її більш придатною для прогнозування курсу криптовалют.

ЕМА обчислюється шляхом усереднення цінових даних за певний період часу, при цьому свіжішим даними присвоюється більша вага. Коефіцієнт

згладжування обчислюється на основі періоду ЕМА і може бути змінений залежно від того, який ступінь важливості ви хочете надати свіжішим даним. Формула для обчислення ЕМА наведена нижче:

$$EMA_t = P_t * \alpha + EMA_{t-1} * (1 - \alpha) \quad (2.4)$$

Де  $EMA_t$  – ЕМА в момент часу  $t$ ,

$P_t$  – ціна в момент часу  $t$ ,

$EMA_{t-1}$  – ЕМА в момент часу  $t-1$ ,

$\alpha$  - коефіцієнт згладжування, що обчислюється за наступною формулою:

$$\alpha = \frac{2}{N + 1} \quad (2.5)$$

Де  $N$  – обраний період для ЕМА

Застосування ЕМА для прогнозування курсу криптовалют може бути корисним інструментом для інвесторів, оскільки вона дозволяє побачити поточну тенденцію на ринку та передбачити, куди може попрямувати курс криптовалюти. [36]

## 2.5 Використання моделі ARIMA для задач регресії

Модель авторегресійної інтегрованої ковзної середньої ARIMA (Autoregressive Integrated Moving Average) — це широко використовуваний метод прогнозування часових рядів, який поєднує компоненти авторегресії (AR), диференціювання (I) і ковзної середньої (MA) для моделювання основної структури часового ряду та створення.

Огляд компонентів моделі ARIMA:

1. Авторегресія (AR): Авторегресійна складова моделі ARIMA базується на припущенні, що поточне значення часового ряду можна передбачити за допомогою лінійної комбінації його попередніх значень. Порядок компонента AR ( $p$ ) представляє кількість попередніх моментів часу, використаних для прогнозу.
2. Диференціювання (I): компонент розрізнення використовується, щоб зробити нестационарний часовий ряд стаціонарним шляхом віднімання попереднього значення точки даних із поточного значення. Цей процес можна застосовувати кілька разів ( $d$ ), доки часовий ряд не стане стаціонарним. Стаціонарний часовий ряд – це ряд із постійним середнім значенням, дисперсією та автокореляцією в часі.
3. Ковзна середня (MA): компонент ковзної середньої моделі ARIMA припускає, що поточне значення часового ряду можна передбачити за допомогою лінійної комбінації минулих умов помилок (білий шум). Порядок компонента MA ( $q$ ) представляє кількість минулих термінів помилок, які використовуються для прогнозу.

Модель ARIMA зазвичай позначається як  $ARIMA(p, d, q)$ ,

де  $p$  — порядок компонента AR,

$d$  — ступінь різниці,

а  $q$  — порядок компонента MA.

Модель адаптується до даних часових рядів за допомогою таких методів, як оцінка максимальної правдоподібності (MLE) або умовних найменших квадратів (CLS) для оцінки коефіцієнтів компонентів AR і MA.

Моделі ARIMA широко використовуються для прогнозування у фінансах, економіці та інших галузях, що мають справу з даними часових рядів. Вони забезпечують гнучкий підхід до фіксації основних закономірностей і залежностей у даних, дозволяючи робити точні та надійні прогнози майбутніх спостережень.

Загальну модель ARIMA( $p, d, q$ ) можна виразити таким чином[]:

(2.6)

Де  $Y_t$  - значення часового ряду в момент  $t$ ,

$b$  – оператор кроку назад,

$p$  - порядок компонента AR,

$q$  — порядок компонента MA,

$c, a_i, b_j$  – параметри моделі,

$\Delta d$  – оператор різності часового ряду порядку  $d$

$\varepsilon_t$  – стаціонарний часовий ряд

При застосуванні моделі ARIMA до часового ряду, першим кроком потрібно переконатися, що часовий ряд є стаціонарним, що означає, що його середнє значення, дисперсія та структура автокореляції не змінюються з часом. Якщо часовий ряд є нестаціонарним, можна застосувати різницю (I), щоб перетворити його на стаціонарний ряд.

Необхідно визначити відповідні значення для  $p, d$  і  $q$ , використовуючи такі методи, як функція автокореляції (ACF), функція часткової автокореляції (PACF) і критерії вибору моделі, такі як інформаційний критерій Акаїке (AIC) або байєсівський інформаційний критерій (BIC).

Далі оцінити коефіцієнти AR і MA підігнавши модель ARIMA до стаціонарного часового ряду, використовуючи статистичні методи, такі як оцінка максимальної правдоподібності (MLE) або умовні найменші квадрати (CLS).

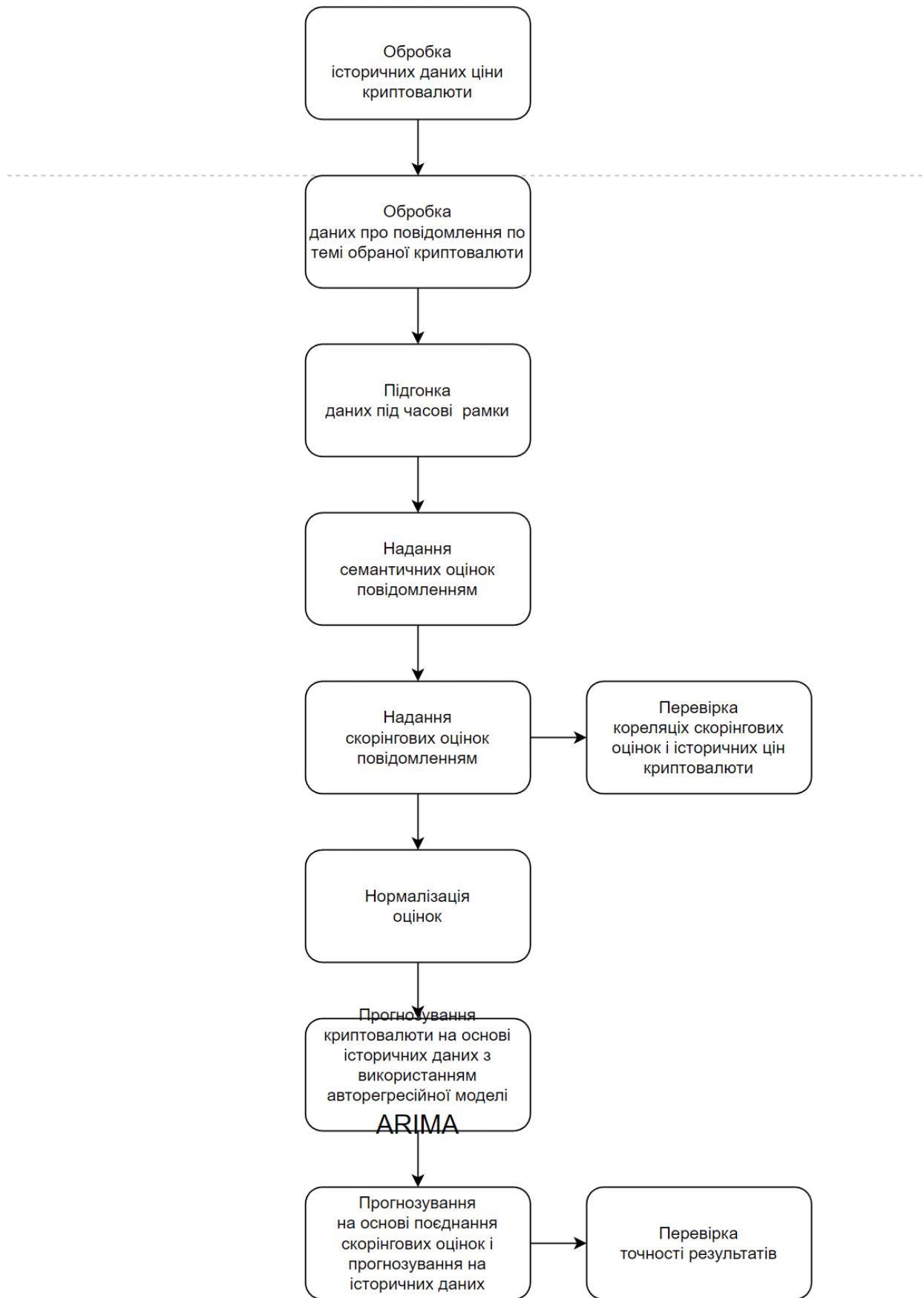
Після того як модель ARIMA підігнана, її можна використовувати для створення прогнозів на майбутні періоди часу.[37,38]

## 2.6 Узагальнена архітектура створюваної системи

Алгоритм роботи інтелектуального модулю прогнозування курсу криптовалют з використанням методів семантичного та технічного аналізів складається з наступних етапів:

1. Обробка історичних даних ціни криптовалюти
2. Обробка даних про повідомлення по темі обраної криптовалюти
3. Підгонка даних під часові рамки
4. Надання семантичних оцінок повідомленням
5. Надання скорінгових оцінок повідомленням
6. Нормалізація оцінок
7. Перевірка кореляцій скорінгових оцінок і історичних цін криптовалюти
8. Прогнозування криптовалюти на основі історичних даних з використанням авторегресійної моделі ARIMA
9. Прогнозування на основі поєднання скорінгових оцінок і прогнозування на історичних даних
10. Перевірка точності результатів

Алгоритм схематично зображений на рисунку 2.3.



## Рисунок 2.4 – Схематичне зображення алгоритму

### 2.7 Алгоритм скорінгу повідомлень

Семантичні оцінки повідомлень відображають настрої людей щодо ринкової ситуації, проте вони не враховують і інших факторів, наприклад кількість підписників у автора повідомлення, кількість ретвітів і т.д. Для підвищення кореляції і точності прогнозування було прийнято рішення побудувати власну формулу підрахунку впливовості повідомлень на базі метаданих. Далі описані фічі, які беруть участь у підрахунку скорінгової оцінки.

Семантична оцінка: (від -1 до 1) оцінка настрою, розрахована за допомогою методів обробки природної мови (NLP), представляє ступінь позитивності чи негативності повідомлення. Вища оцінка вказує на більш оптимістичний погляд на курс криптовалюти, тоді як нижча оцінка означає негативну перспективу.

Кількість підписників: кількість підписників користувача соціальної мережі, який опублікував повідомлення. Більша кількість підписників свідчить про більший потенційний вплив на настрої ринку.

Статус пересилання: двійкова функція, яка вказує, чи повідомлення є пересланим (1) чи оригінальним (0).

Кількість осіб, у яких користувач є в закладках: Відображає авторитетність висловлювача. Більша кількість закладок може означати більшу аудиторію та потенційно сильніший вплив на настрої ринку.

Враховуючи ці фічі, тепер можна визначити формулу для обчислення кінцевої скорінгової оцінки (S) для певного повідомлення.

$$S = \frac{\textit{compound} * \textit{followers} * (\textit{favourites} + 1) * (\textit{retweet} + 1)}{\textit{followers} + 1} \quad (2.7)$$

Де *compound* – семантична оцінка повідомлення,

*followers* – кількість підписників

*favourites* - кількість осіб, у яких користувач є в закладках

*retweet* – статус пересилання

Остаточна скорінгова оцінка (S) служить індикатором загального впливу повідомлення на настрої ринку криптовалюти. Маючи оцінки кількох повідомлень і аналізуючи їх зв'язок із історичними даними про ціни, була розроблена модель прогнозування, яка використовує скорінгову оцінку для підвищення точності прогнозування на майбутні зміни цін

Після цього остаточна скорінгова оцінка нормалізується та домножується на прогноз ціни за день по формулі:

$$\textit{final\_result}_t = \textit{result}_t * S_t \quad (2.8)$$

Де *result<sub>t</sub>* – прогноз побудований на історичних даних в момент часу t,

*S<sub>t</sub>* – нормалізована скорінгова оцінка в момент часу t

## 2.8 Проектні рішення

Оскільки програмні засоби, які вирішують поставлену задачу є платними або мають суттєві недоліки, було прийнято рішення про створення власного інтелектуального модулю.

Для цього було необхідно визначити мову програмування, базовий функціонал, додаткові інструменти та середовища розробки.

Мовою програмування було обрано Python, оскільки він чудово підходить для аналітичних задач, має багато вбудованого функціоналу, кросплатформенний, має багато готових рішень, які допоможуть при виконанні роботи. В роботі було використано багато сторонніх бібліотек. Серед основних варто виділити наступні:

1. Matplotlib – бібліотека для побудови графіків і візуалізацій. Є простою у використанні та є повноцінною для виконання необхідних в роботі задач.
2. NumPy – головна бібліотека для виконання математичних розрахунків. Має всі необхідні формули та є нересурсозатратною.
3. Pandas – бібліотека для роботи з CSV файлами. Допомогає обробляти сирі вхідні дані, виконує основні задачі по перетворенню даних.
4. vaderSentiment – бібліотека з моделями для аналізу настроїв текстових повідомлень. Найкраще серед аналогів працює з людською мовою.
5. Sklearn – включає в себе безліч моделей, а також має функції нормалізації даних.
6. Pmdarima – спеціальна бібліотека для роботи з авторегресивною моделлю ARIMA. Поміж інших її виділяє можливість автоматичного підбору параметрів моделі.

Середовищем для розробки було обрано PyCharm, оскільки воно є зручним, дозволяє тонко налаштовувати оточення, має інструменти для підтримки чистоти коду.

## 2.9 Область застосування та перспективність алгоритму

Запропонований алгоритм має великий потенціал для впровадження в системи прогнозування курсу криптовалют, оскільки використовує поєднання методів з різних аналізів, увібравши переваги кожного з них. Інтелектуальний модуль є безкоштовним і використовує нересурсозатратний алгоритм. Також його можна покращувати для підвищення точності результатів прогнозування, що визначає його перспективність використання в більш складних системах.

Завдяки використанню семантичного аналізу в алгоритмі враховуються зовнішні чинники, такі як ситуація в світі, настрої гравців ринку чи інші викиди, які не можуть бути враховані при використанні технічного аналізу, що базується лише на історичних даних. Модуль vaderSentiment дозволяє на достатньо точному рівні визначати настрої по повідомленнях в соціальних мережах.

З іншого боку використання одного лише семантичного аналізу настроїв має досить слабку кореляцію зі змінами курсу, тому в якості основи використовуються класичні методи регресії, в нашому випадку авторегресивна модель ARIMA.

Тому поєднання цих двох підходів і дає більш повну оцінку ситуації на ринку, що в свою чергу підвищує точність прогнозування.

В подальшому алгоритм може бути покращений при інтеграції з більшими системами та збільшенні ресурсів.

## 2.10 Практична цінність розробки

Створення інтелектуального модуля для прогнозування курсу криптовалют має велику практичну цінність. Вона різна для кожної з зацікавлених сторін, описаних в першому розділі.

По-перше це обґрунтоване прийняття рішень: програма, яка прогнозує ціни на криптовалюту, може допомогти користувачам приймати більш

обґрунтовані інвестиційні та торгові рішення, надаючи їм уявлення про майбутні цінові тенденції та потенційні зміни ринку.

По-друге це управління ризиками. Аналізуючи та прогнозуючи коливання цін, програма може допомогти користувачам ефективніше керувати інвестиційними ризиками, дозволяючи їм розробляти стратегії пом'якшення та диверсифікації ризиків.

По-третє аналіз ринкових настроїв, використовуючи дані соціальних медіа, новинні статті чи інші відповідні джерела інформації, може надати цінну інформацію про загальні ринкові настрої.

По-четверте – універсальність, алгоритм описаний в роботі дозволяє адаптувати програму для прогнозування ціни будь якого токена. Обираючи конкретні криптовалюти, часові рамки та набір повідомлень з соціальних мереж, користувач може отримати персоналізовані прогнози, які відповідають його індивідуальним потребам.

## 2.11 Висновки до другого розділу

У даному розділі було описано основні математичні засади та формули створюваного програмного продукту.

Було проведено аналіз використання семантичного аналізу настроїв повідомлень, розглянуто основні параметри та формули, які можуть бути використані. Було досліджено задачу регресії для прогнозування криптовалют, досліджено використання ковзної середньої. Було прийнято рішення використовувати авторегресійну модель ARIMA для прогнозування курсу криптовалюти на історичних даних, розглянуто основні поняття та особливості її застосування.

Було проведено функціональний аналіз розроблюваного програмного продукту, виділення основних етапів алгоритму, їх схематичне зображення.

## РОЗДІЛ 3. АНАЛІЗ РЕАЛІЗОВАНОГО ІНТЕЛЕКТУАЛЬНОГО МОДУЛЮ ПРОГНОЗУВАННЯ КРИПТОВАЛЮТ З ВИКОРИСТАННЯМ МЕТОДІВ СЕМАНТИЧНОГО І ТЕХНІЧНОГО АНАЛІЗІВ

### 3.1 Інформаційна технологія роботи з системою

В даному розділі будуть описані основні положення роботи з системою. Буде описано системні вимоги, описано як організована робота з даними, їх структура, вимоги, а також особливості пошуку та підготовки даних для семантичної оцінки.

#### 3.1.1 Системні вимоги

Інтелектуальний модуль виконує складні математичні обчислення, тому було проведено тестування різних обчислювальних машин, для визначення мінімальних системних вимог. Результати швидкодії для прогнозування одного дня з історичними даними за 250 днів та різною кількістю повідомлень представлено в таблиці 3.1.

Програму було протестовано на трьох пристроях:

1.

- процесор: Intel Core i5-1135G7;
- оперативна пам'ять: 8Gb;
- місце на диску: 256 Gb;
- відеокарта Mesa Intel XE Graphics;
- ОС: Ubuntu 20.04.

2.

- процесор: Intel Core i5-9600K;
- оперативна пам'ять: 16Gb;

- місце на диску: 1024 Gb;
- відеокарта AMD RX570 8GB;
- ОС: Windows 10.

3.

- процесор: Intel Core i7-13700K;
- оперативна пам'ять: 16Gb;
- місце на диску: 1024 Gb;
- відеокарта Nvidia RTX 4070ti 12GB;
- ОС: Windows 10.

Таблиця 3.1 – Порівняння часу роботи на різних пристроях

Номер пристрою	1000 повідомлень	10000 повідомлень	20000 повідомлень	30000 повідомлень
1	721ms	1940ms	18965ms	42074ms
2	120ms	698ms	1767ms	4651ms
3	111ms	296ms	889ms	1360ms

По результатам дослідження часу видно, що показники виконання суттєво відрізняються в залежності від потужності обчислювальної машини. На їх основі було запропоновано оптимальні системні вимоги, які дозволили б виконувати розрахунки при великих обсягах даних за прийнятний час.

- процесор: Intel Core i5-9600K;
- оперативна пам'ять: 16Gb;
- місце на диску: 1024 Gb;
- відеокарта: неважливо;
- ОС: неважливо.

### 3.1.2 Підготовка даних

#### 3.1.2.1 Підготовка історичних даних ціни криптовалюти

Система отримує на вхід файл у форматі CSV. До обов'язкових колонок відносяться:

1. Date – містить дату історичної ціни
  - a. Тип даних: string
  - b. Формат дати: YYYY-MM-DD
2. Close – містить значення історичної ціни
  - a. Тип даних: float

Приклад:

Date	Open	High	Low	Close	Adj Close	Volume
2021-05-04	57214.1796875	57214.1796875	53191.42578125	53333.5390625	53333.5390625	68564706967
2021-05-05	53252.1640625	57911.36328125	52969.0546875	57424.0078125	57424.0078125	69241316747
2021-05-06	57441.30859375	58363.31640625	55382.5078125	56396.515625	56396.515625	69523285106
2021-05-07	56413.953125	58606.6328125	55321.84765625	57356.40234375	57356.40234375	68434023376
2021-05-08	57352.765625	59464.61328125	56975.2109375	58803.77734375	58803.77734375	65382980634
2021-05-09	58877.390625	59210.8828125	56482.00390625	58232.31640625	58232.31640625	65906690347
2021-05-10	58250.87109375	59519.35546875	54071.45703125	55859.796875	55859.796875	71776546298
2021-05-11	55847.2421875	56872.54296875	54608.65234375	56704.57421875	56704.57421875	61308396325
2021-05-12	56714.53125	57939.36328125	49150.53515625	49150.53515625	49150.53515625	75215403907
2021-05-13	49735.43359375	51330.84375	46980.01953125	49716.19140625	49716.19140625	96721152926
2021-05-14	49682.98046875	51438.1171875	48868.578125	49880.53515625	49880.53515625	55737497453
2021-05-15	49855.49609375	50639.6640625	46664.140625	46760.1875	46760.1875	59161047474
2021-05-16	46716.63671875	49720.04296875	43963.3515625	46456.05859375	46456.05859375	64047871555
2021-05-17	46415.8984375	46623.55859375	42207.2890625	43537.51171875	43537.51171875	74903638450
2021-05-18	43488.05859375	45812.45703125	42367.83203125	42909.40234375	42909.40234375	56187365084
2021-05-19	42944.9765625	43546.1171875	30681.49609375	37002.44140625	37002.44140625	126358098747

Рисунок 3.1 – Приклад історичних даних ціни Bitcoin

#### 3.1.2.2 Підготовка даних повідомлень

Система отримує на вхід файл у форматі CSV. До обов'язкових колонок відносяться:

1. date – містить дату історичної ціни
  - a. Тип даних: string

b. Формат дати: YYYY-MM-DD HH:MM:SS

2. user\_followers – містить кількість підписників висловлювача
  - a. Тип даних: int
3. user\_favourites – містить кількість людей в яких висловлювач є в закладках
  - a. Тип даних: int
4. is\_retweet – містить інформацію про те, чи є повідомлення пересланим
  - a. Тип даних: bool

Приклад:

index	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verified	date	text	hashtags	source	is_retweet
										nice project			
195760	محمد	esfahan	خودت رو به خودت	2021-01-02 19:0 616.0		1683	1712	ЛОЖЬ	2021-06-22 5:22	ortcoin ort okrate	['ortcoin', 'ort', 'oi	Twitter for Andro	ЛОЖЬ
1135874	Crypto Revolution Merchandise		Crypto Revolutic	2019-10-13 6:16 451.0	2045.0		45077.0	ЛОЖЬ	2021-08-25 7:06	Long Bitcoin shc	['Bitcoin']	Twitter Web App	ЛОЖЬ
332828	STRATOFITZ	USA	Digital Token En	2014-04-25 19:4 468.0	111.0		1262.0	ЛОЖЬ	2021-07-02 19:3	Top Trending Cr	['Cryptocurrency IFTTT		ЛОЖЬ
581946	U.Today	New York, USA	https://t.co/Tt6M	2018-01-17 9:28 35990.0	4447.0		682.0	ЛОЖЬ	2021-07-24 14:1	Can one expect	['BTC']	get shares	ЛОЖЬ
										We will see...			
108170	Mr Fulcanelli	Argentina	be decentralized	2010-08-23 20:4 164.0	100.0		9295.0	ЛОЖЬ	2021-05-29 15:5	bitcoin btc bnb b	['bitcoin', 'btc', 'bi	Twitter for Mac	ЛОЖЬ
										This is definitely			
1125573	Yigit.bey12		#BNB #BINANC Borsa : ~Sabirsızların pi	2020-11-24 20:2 65.0	1222.0		2067.0	ЛОЖЬ	2021-08-25 15:0	Airdrops Bitcoin	['Airdrops', 'Bitco	Twitter for Andro	ЛОЖЬ
			Follow me to										
2128151	Bitcoin Price Hourly		Get Bitcoin a	2020-12-13 5:58 229.0		0	1	ЛОЖЬ	2022-01-22 22:5	Current Bitcoin	['Bitcoin']	Coin Report	ЛОЖЬ
										One fact to keep SatoshiNakamot to align with the The FED discon Nuff said...			
2114699	PlanBTC	Earth	Obsession with : Financial planner Also love to trav	2021-03-02 0:56 29.0		102	275	ЛОЖЬ	2022-01-11 20:0	BuyBitcoin BTC	['SatoshiNakamc	Twitter for Andro	ЛОЖЬ
35276	Douglas Borthwi	New York, Cana	Chief Marketing	2010-07-25 14:5 6621.0		713	3070	ЛОЖЬ	2021-02-19 13:3	analysts stated t	['Bitcoin', 'BTC']	Twitter for Andro	ЛОЖЬ
245069	Thomas Dunne	Miami Beach, Fl	I post useful #Bil	2014-05-18 5:11 7725.0	36.0		1462.0	ЛОЖЬ	2021-06-21 0:15	Bitcoin: Exchang	['Bitcoin']	Twitter Web App	ЛОЖЬ
1178115	U.Today	New York, USA	https://t.co/Tt6M	2018-01-17 9:28 36867.0	4440.0		682.0	ЛОЖЬ	2021-08-23 16:0	Is seeking to giv	['Bitcoin']	get shares	ЛОЖЬ
2229858	AcornHoarder		Professional	2012-05-24 20:5 349.0		1888	15779	ЛОЖЬ	2022-01-21 3:44	You really think t	['bitcoin', 'btc']	Twitter Web App	ЛОЖЬ

Рисунок 3.2 – Приклад даних про повідомлення з соціальної мережі Twitter

### 3.1.3 Особливості пошуку даних для семантичної оцінки.

Залежність точності прогнозування від кількості і якості повідомлень є важливим фактором для моделі прогнозування, яка використовує аналіз настроїв для виявлення тенденцій в ціні криптовалют або інших фінансових інструментів.

Під час виконання роботи було проведено дослідження впливу кількості якості повідомлень на точність прогнозування. У якості індикаторів використовували графік крос-кореляції Пірсона [39][40].

Коефіцієнт кореляції Пірсона розраховують по формулі:

$$r_{xy} = \frac{\sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^m (x_i - \bar{x})^2 \sum_{i=1}^m (y_i - \bar{y})^2}} = \frac{cov(x, y)}{\sqrt{s_x^2 s_y^2}}, \quad (3.1)$$

Де  $\bar{x}, \bar{y}$  – вибіркові середні  $x^m, y^m$ ,

$s_x^2, s_y^2$  – вибіркові дисперсії,

$r \in [-1, 1]$

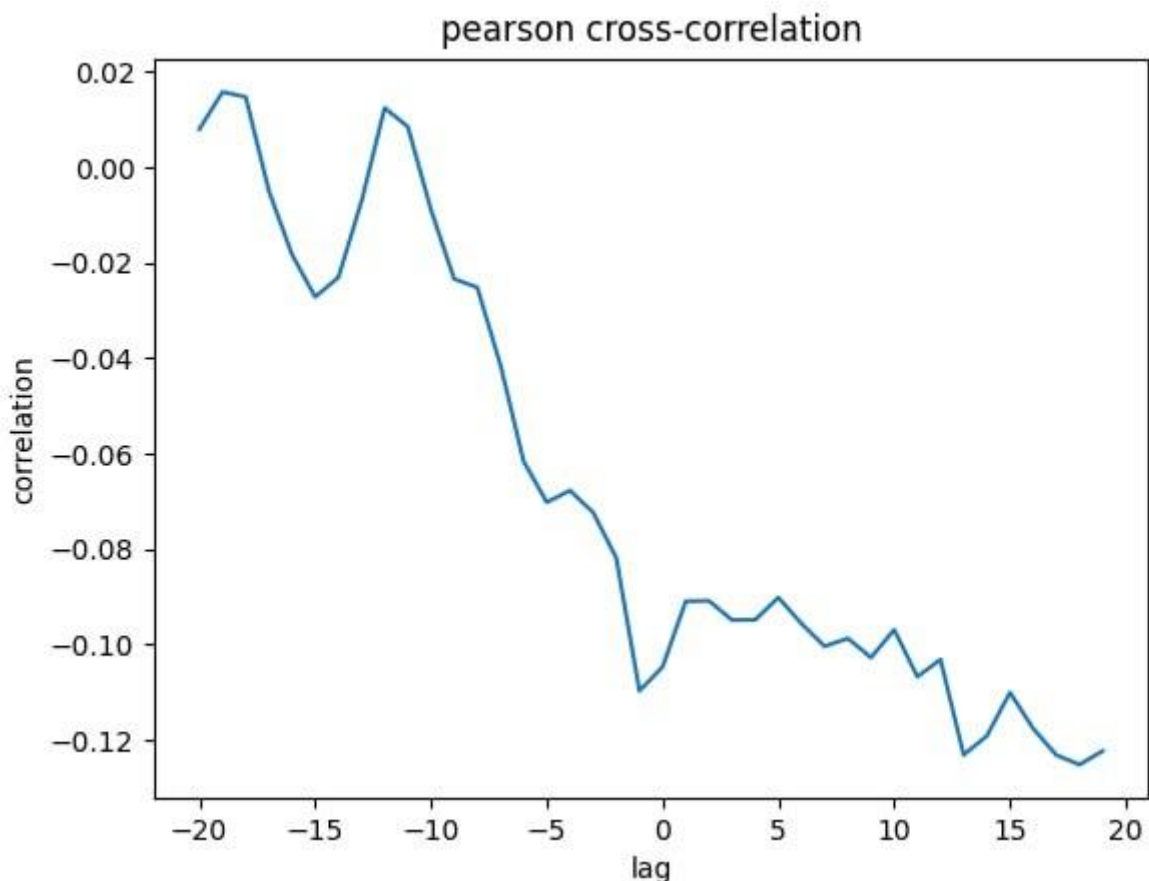


Рисунок 3.3 – Крос-кореляція Пірсона для набору семантичних оцінок

На наступних двох рисунках зображено приклад результатів дослідження репрезентативності вибірки повідомлень на заданому проміжку часу.

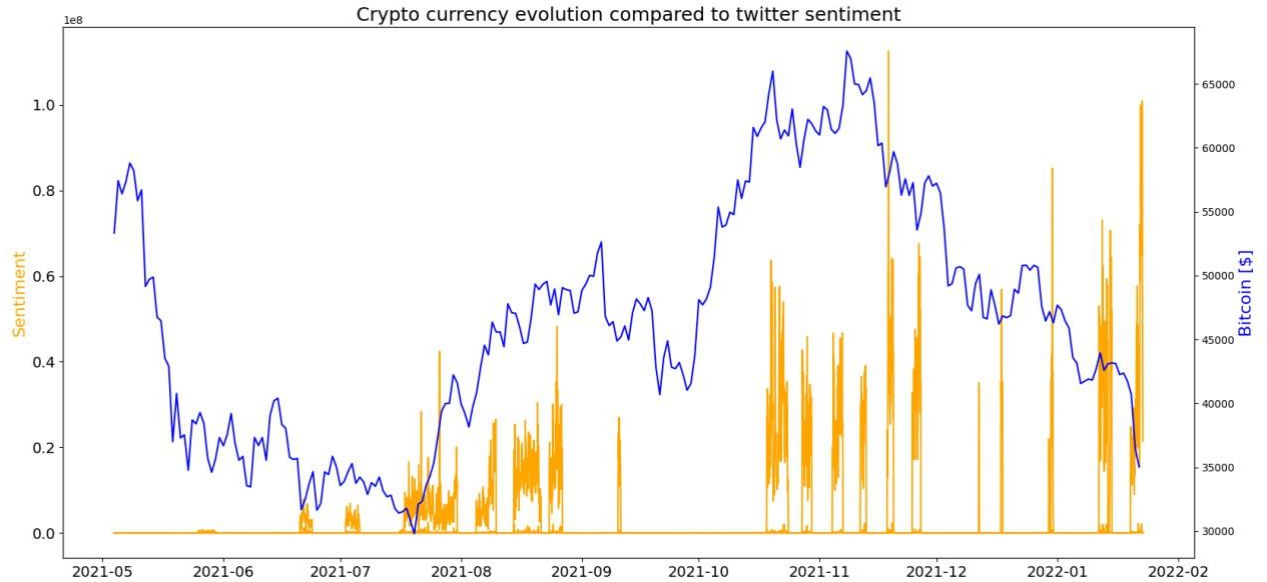


Рисунок 3.4 – Графік ціни криптовалюти у порівнянні з семантичною оцінкою повідомлень

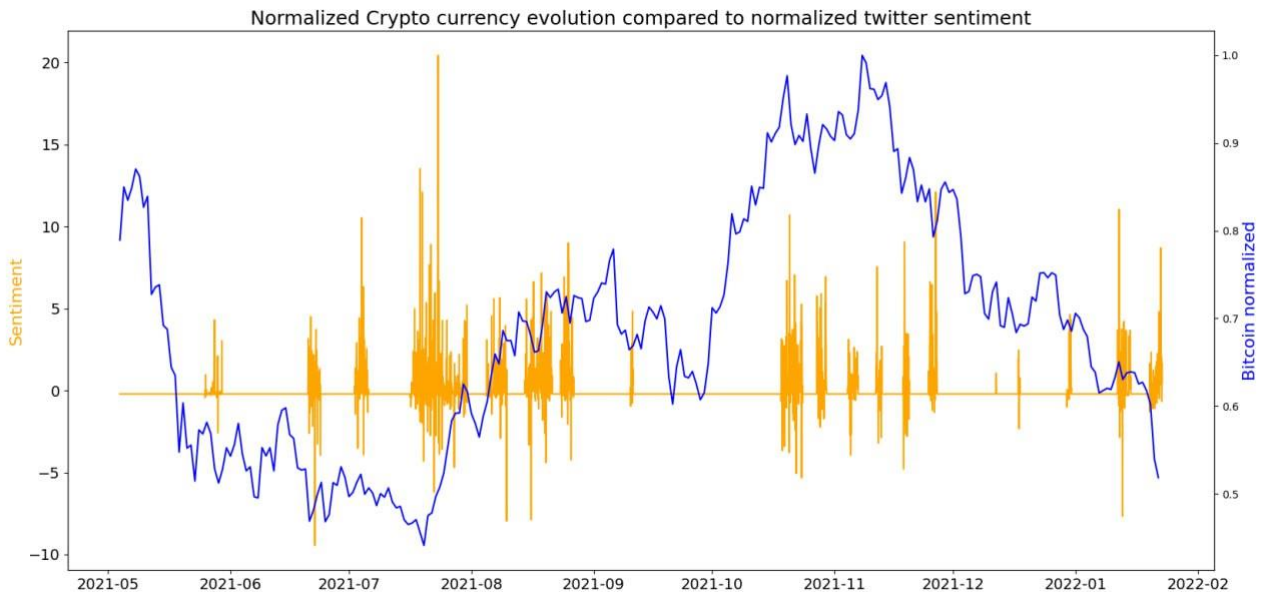


Рисунок 3.5 – Нормалізований графік ціни криптовалюти у порівнянні з семантичною оцінкою повідомлень

Далі наведено порівняльну таблицю результатів прогнозування в залежності від середньої кількості повідомлень за день.

Таблиця 3.2 – Залежність точності від кількості повідомлень за день

	10 повідомлень	20 повідомлень	50 повідомлень
MAPE[42]	7.9223%	4.3108%	3.7966%

За результатами порівняння видно, що точність прямо пропорційна кількості повідомлень, тому доцільно використовувати якнайбільше повідомлень при прогнозуванні.

Також експериментальним шляхом було встановлено, що найефективніше проводити прогнозування на 1 день. Це пояснюється тим, що найбільш значущі повідомлення висвітлюються в засобах масової інформації, спеціалізованих моніторингових системах зранку на наступний день.

Варто зауважити, що при зростанні кількості оброблюваних повідомлень зростають обчислювальні витрати. Порівняння залежності швидкодії від кількості повідомлень наведено в таблиці 3.1.

В даному дослідженні використовувався датасет з 22000 повідомленнями, що давав в середньому 90 повідомлень за день.

### 3.2 Тестування розробленого інтелектуального модулю

Далі наведено результати роботи розробленого інтелектуального модулю для декількох проміжків часу та розмірів інтервалу.

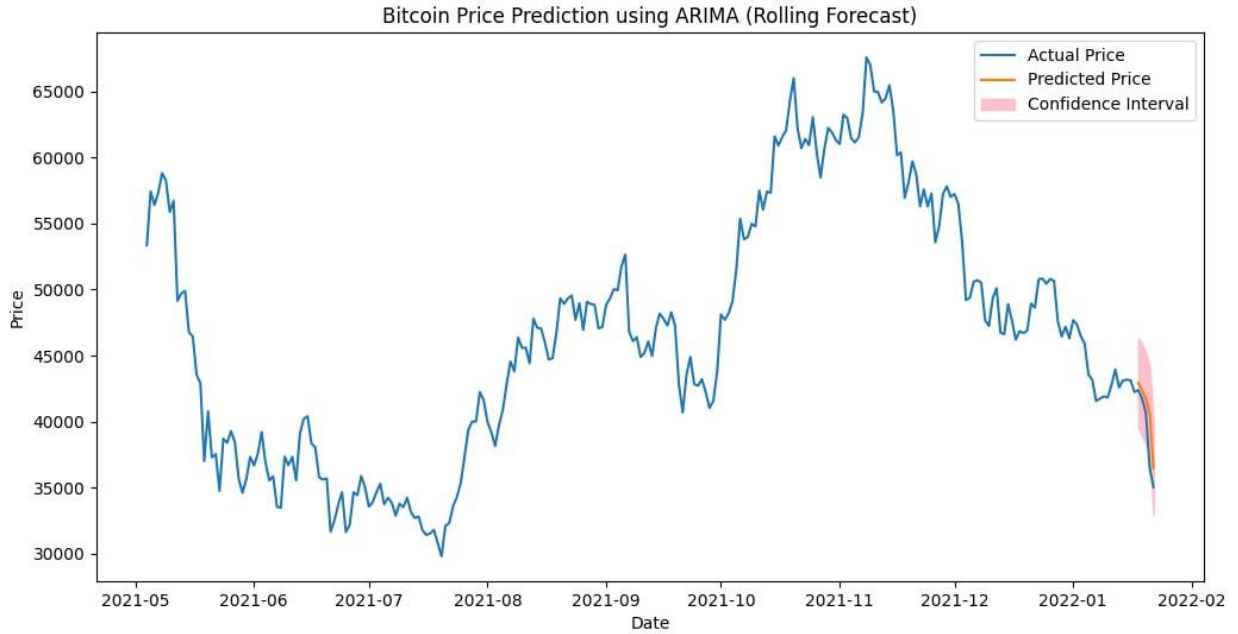


Рисунок 3.6 – Результат роботи програми 1 у вигляді графіку

Cryptocurrency: Bitcoin

Timerange: 2022-01-17 - 2022-01-22

Predicted price values:

42937.8596722088

42375.6328125

41744.328125

40680.41796875

36457.31640625

Mean Absolute Percentage Error (MAPE): 4.222373785716569%

Mean Square Percentage Error (MSPE): 0.32332978545201624%

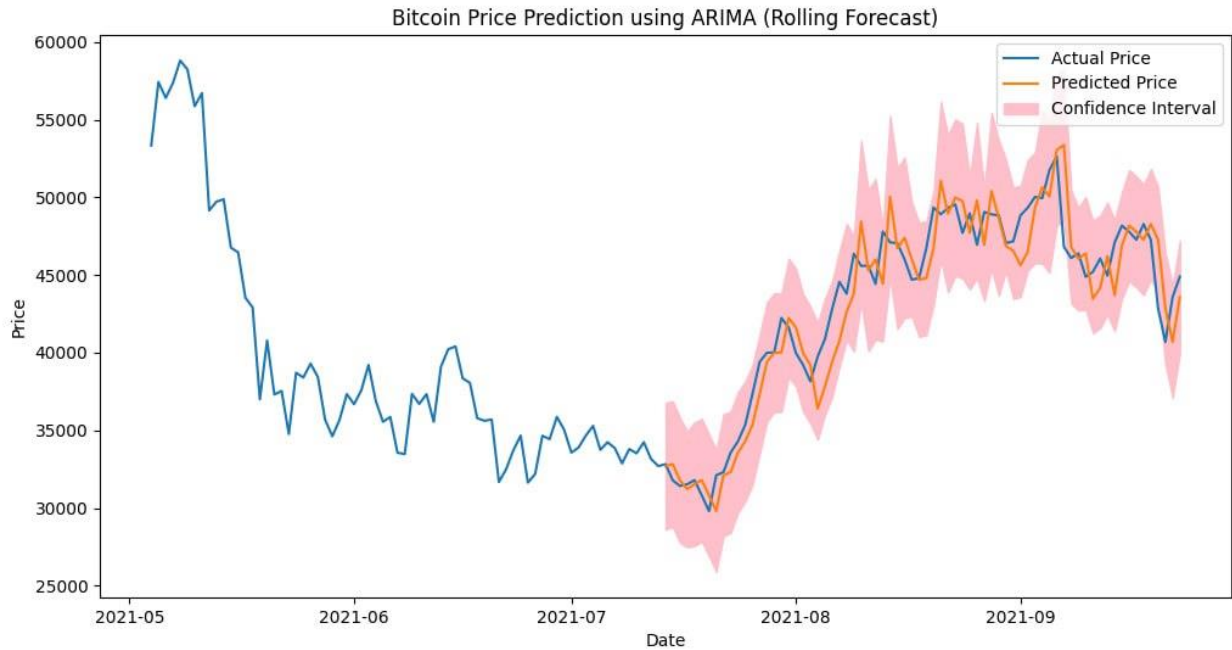


Рисунок 3.7 – Результат роботи програми 2 зі зменшеним набором тренувальних даних у вигляді графіку

Cryptocurrency: Bitcoin

Timerange: 2021-07-14 - 2021-09-23

Predicted price values:

32702.025390625

32822.34765625

31780.73046875

31210.92579530536

31533.068359375

...

47274.14208653528

42843.80078125

40693.67578125

43574.5078125

Mean Absolute Percentage Error (MAPE): 3.5756932019872774%

Mean Square Percentage Error (MSPE): 0.20115292853997505%

В даному прикладі видно, що репрезентативність вибірки повідомлень покращує точність прогнозування нівелюючи зменшення тренувальних даних. (рис.3.10)

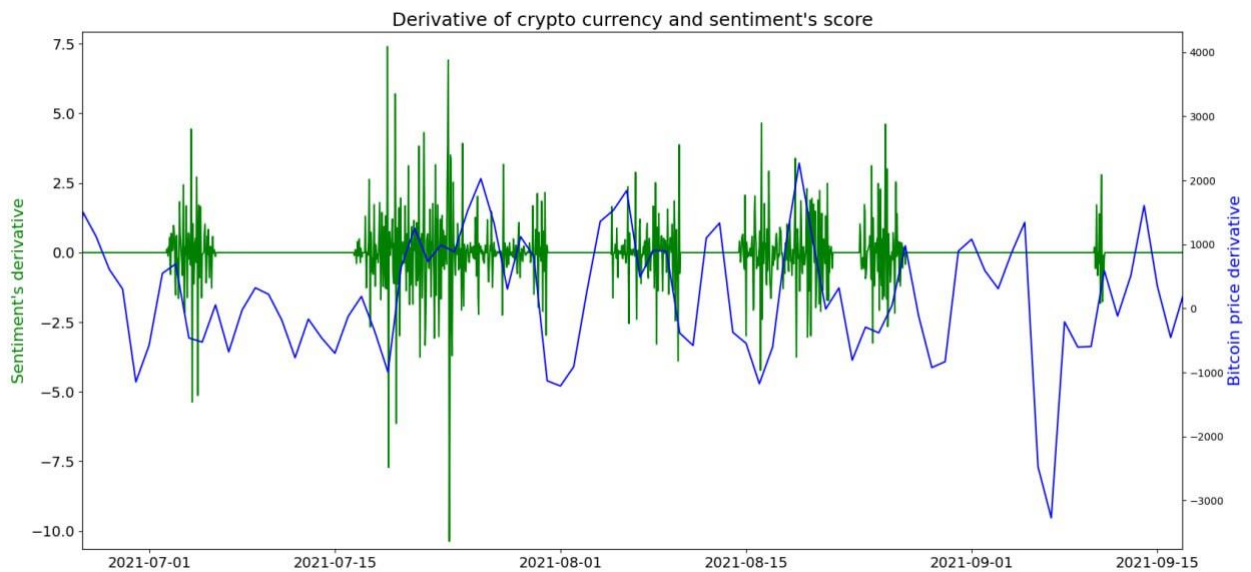


Рисунок 3.8 – Нормалізований графік похідної від ціни криптовалюти у порівнянні з похідною від семантичної оцінки повідомлень, що демонструє репрезентативність вибірки повідомлень на заданому проміжку часу.

### 3.3 Порівняння отриманих результатів з результатами використання інших методів

Для порівняння отриманих результатів роботи інтелектуального модуля прогнозування криптовалют з використанням методів семантичного та технічного аналізів було використано наступні методи:

1. Метод на основі аналізу історичних даних з використанням моделі ARIMA
2. Метод семантичного аналізу з використанням моделі ARIMA

Порівняння з іншими методами не має сенсу, оскільки вони представляють результати у вигляді графіків та не оперують точними числами або виконують іншу задачу.

Результати порівняння наведені в таблиці 3.1

Таблиця 3.3 – Порівняння результатів роботи методів прогнозування  
криптовалют

	MSE 250 днів	MSE 90 днів	MSE 30 днів	MAPE 250 днів	MAPE 90 днів	MAPE 30 днів
Інтелектуальний модуль прогнозування з використанням методів технічного та семантичного аналізів	0.1323%	0.1217%	0.1297%	2.2670%	2.1325%	2.3350%
Метод семантичного аналізу з використанням моделі ARIMA	0.5099%	0.4562%	0.5133%	7.1353%	6.9823%	8.0182%
Метод на основі аналізу історичних даних з використанням моделі ARIMA	0.1320%	0.1032%	0.1534%	2.2926%	2.1325%	2.8596%

### 3.4 Висновки до третього розділу

В даному розділі було описано роботу інтелектуального модулю, проведено тестування розробленого інтелектуального модулю, порівняно розроблений інтелектуальний модуль прогнозування з іншими методами, які виконують подібну задачу.

За результатами порівняльного аналізу було визначено, що поєднання методів технічного та семантичного аналізу можуть покращити точність прогнозування криптовалюти за умови великої кількості семантичних даних а також репрезентативності вибірки повідомлень.

Тестування інтелектуального модулю показало, що він достатньо добре прогнозує курс криптовалют на рівні 98%. Його результати подібні до методу на основі аналізу історичних даних з використанням моделі ARIMA у випадках коли відсутні семантичні дані, оскільки він побудований на його основі, а у випадку наявності репрезентативних даних вони можуть бути трошки кращі.

## ВИСНОВКИ

У випускній кваліфікаційній роботі було проведено аналіз сучасних методів прогнозування криптовалют, досліджено ефективність різних методів прогнозування криптовалют, розроблено інтелектуальний модуль прогнозування криптовалют з використанням методів семантичного та технічного аналізів. Було проведено порівняння розробленого інтелектуального модулю прогнозування криптовалют з іншими методами прогнозування криптовалют на прикладі Bitcoin та соціальної мережі Twitter.

Розроблений інтелектуальний модуль справляється з поставленою задачею та надає корисну інформацію користувачу для прогнозування криптовалют.

Його можна покращити додавши можливість автоматично отримувати дані з різних джерел та обробляти їх.

Він може бути впроваджений в комерційні програми для покращення точності прогнозування.

Тестування інтелектуального модулю показало, що він достатньо добре прогнозує курс криптовалют на рівні 98%. Його результати подібні до методу на основі аналізу історичних даних з використанням моделі ARIMA у випадках коли відсутні семантичні дані, оскільки він побудований на його основі, а у випадку наявності репрезентативних семантичних даних вони можуть бути трошки кращі.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System.  
Режим доступу: <https://bitcoin.org/bitcoin.pdf>
2. Milutinović, Monia (2018). "Cryptocurrency". *Ekonomika*. **64** (1): 105–122. doi:10.5937/ekonomika1801105M. ISSN 0350-137X.
3. Pernice, Ingolf G. A.; Scott, Brett (20 May 2021). "Cryptocurrency". *Internet Policy Review*. **10** (2). doi:10.14763/2021.2.1561. ISSN 2197-6775.
4. Liu, Jinan; Rahman, Sajjadur; Serletis, Apostolos (2020). "Cryptocurrency Shocks". *SSRN Electronic Journal*. doi:10.2139/ssrn.3744260. ISSN 1556-5068. S2CID 233751995.
5. Park, Sehyun; Im, Seongwon; Seol, Youhwan; Paek, Jeongyeup (2019). "Nodes in the Bitcoin Network: Comparative Measurement Study and Survey". *IEEE Access*. **7**: 57009–57022. doi:10.1109/ACCESS.2019.2914098. S2CID 155106629.
6. Krantz M. *Fundamental analysis for dummies* / M.Krantz –Hoboken: Wiley Publishing Inc., 2009. – 387 p.
7. Mizrach, Bruce; Weerts, Susan (27 November 2007). "Highs and Lows: A Behavioral and Technical Analysis". *SSRN* 1118080.
8. Paul V. Azzopardi (2010). *Behavioural Technical Analysis: An introduction to behavioural finance and its role in technical analysis*. Harriman House. ISBN 978-1905641413.
9. Andrew W. Lo; Jasmina Hasanhodzic (2010). *The Evolution of Technical Analysis: Financial Prediction from Babylonian Tablets to Bloomberg Terminals*. Bloomberg Press. p. 150. ISBN 978-1576603499. Retrieved 8 August 2011.
10. Irwin, Scott H.; Park, Cheol-Ho (2007). "What Do We Know About the Profitability of Technical Analysis?". *Journal of Economic Surveys*. **21** (4): 786–826. doi:10.1111/j.1467-6419.2007.00519.x. S2CID 154488391.

11. Osler, Karen (July 2000). "Support for Resistance: Technical Analysis and Intraday Exchange Rates," FRBNY Economic Policy Review (abstract and paper here).
12. Lo, Andrew W.; Mamaysky, Harry; Wang, Jiang (2000). "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation". *Journal of Finance*. **55** (4): 1705–1765. CiteSeerX 10.1.1.134.1546. doi:10.1111/0022-1082.00265.
13. Joseph de la Vega, *Confusión de Confusiones*, 1688
14. Nison, Steve (1991). *Japanese Candlestick Charting Techniques*. pp. 15–18. ISBN 978-0-13-931650-0.
15. Nison, Steve (1994). *Beyond Candlesticks: New Japanese Charting Techniques Revealed*, John Wiley and Sons, p. 14. ISBN 0-471-00720-X
16. Paul V. Azzopardi, "Behavioral Technical Analysis", *ibid*
17. Murphy, John J. *Technical Analysis of the Financial Markets*. New York Institute of Finance, 1999, pp. 1–5, 24–31. ISBN 0-7352-0066-1
18. Archived from the original on 6 January 2015. Retrieved 6 January 2015.
19. Elder (1993), *Part III: Classical Chart Analysis*
20. <https://www.tradingview.com/chart/>
21. <https://santiment.net/>
22. <https://www.tensorcharts.com/>
23. <https://lunarcrush.com/coins/btc/bitcoin>
24. <https://www.coinigy.com/en/cryptocurrencies/btc-bitcoin/>
25. <https://www.cryptocompare.com/coins/btc/analysis/USDT>
26. <https://tradesanta.com/technical-indicators>
27. Turney, Peter (2002). "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews". *Proceedings of the Association for Computational Linguistics*. pp. 417–424.
28. Pang, Bo; Lee, Lillian; Vaithyanathan, Shivakumar (2002). "Thumbs up? Sentiment Classification using Machine Learning Techniques". *Proceedings of*

- the Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 79–86.
29. Shahzad Qaiser, Ramsha Ali - Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents - International Journal of Computer Applications – No.1, July 2018 - [https://www.researchgate.net/publication/326425709\\_Text\\_Mining\\_Use\\_of\\_TF-IDF\\_to\\_Examine\\_the\\_Relevance\\_of\\_Words\\_to\\_Documents](https://www.researchgate.net/publication/326425709_Text_Mining_Use_of_TF-IDF_to_Examine_the_Relevance_of_Words_to_Documents)
30. Daniel Jurafsky & James H. Martin - Speech and Language Processing - <https://web.stanford.edu/~jurafsky/slp3/3.pdf>
31. Xinying Song, Alex Salcianu, Yang Song, Dave Dopson, Denny Zhou - Fast WordPiece Tokenization - Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing - November 7–11, 2021 - <https://aclanthology.org/2021.emnlp-main.160.pdf>
32. The Pros and Cons of Using Natural Language Processing Tools - <https://www.wesuggestsoftware.com/the-pros-and-cons-of-using-natural-language-processing-tools>
33. Diksha Khurana, Aditya Koli, Kiran Khatter and Sukhdev Singh – Natural Language Processing: State of The Art, Current Trends and Challenges - <https://arxiv.org/ftp/arxiv/papers/1708/1708.05148.pdf>
34. Mihalcea, Rada; Banea, Carmen; Wiebe, Janyce (2007). "Learning Multilingual Subjective Language via Cross-Lingual Projections" (PDF). Proceedings of the Association for Computational Linguistics (ACL). pp. 976–983. Archived from the original (PDF) on 2010-07-08.
35. Armstrong J.S. Illusions in Regression Analysis / J.S. Armstrong – Pennsylvania: Penn Press, 2011. – 147 p.
36. Бідюк П. І. Аналіз часових рядів (навчальний посібник) / Бідюк П. І., Романенко В. Д., Тимошук О. Л. – К.: Політехніка, 2010. – 317 с.

37. Adhikari R. An Introductory Study on Time Series Modeling and Forecasting / Adhikari R. – Riga: LAP Lambert Academic Publishing, 2013. – 76 p.
38. Вьюгин В.В. Математические основы теории машинного обучения и прогнозирования / В.В. Вьюгин. – М.: МЦНМО, 2013. – 387 с.
39. Classification and Loss Evaluation - Softmax and Cross Entropy Loss - <https://deepnotes.io/softmax-crossentropy>
40. Kiprono Elijah Koech - Derivative of Sigmoid and Cross-Entropy Functions - <https://towardsdatascience.com/derivative-of-sigmoid-and-cross-entropy-functions-5169525e6705>
41. Грешилов А. А., Стакун В. А., Стакун А. А. Математические методы построения прогнозов. — М.: Радио и связь, 1997. — 112 с. — ISBN 5-256-01352-1.
42. Булашев С. В. Статистика для трейдеров. — М.: Компания Спутник+, 2003. — 245 с.
43. Айвазян С. А. Прикладная статистика. Основы эконометрики. Том 2. — М.: Юнити-Дана, 2001. — 432 с. — ISBN 5-238-00305-6.
44. Деундяк, О. В. Система прогнозування курсу криптовалют на основі фундаментального аналізу : магістерська дис. : 124 Системний аналіз / Деундяк Олександр Володимирович. – Київ, 2018. – 66 с.

## ДОДАТОК А Код модулю обробки вхідної інформації від користувача

```
import matplotlib

matplotlib.use('Qt5Agg') # or 'TkAgg', or 'WXAgg'

import matplotlib.pyplot as plt

from time import sleep

import json

import pandas as pd

import io

import re

import numpy as np

from tqdm import tqdm

import pandas as pd

from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer

from tqdm import trange, tqdm_notebook, tqdm

from sklearn import preprocessing

df_clean = pd.read_csv('sorted_file.csv')

analyzer = SentimentIntensityAnalyzer()

compound = []

for i,s in enumerate(tqdm(df_clean['text'],position=0, leave=True)):

    # print(i,s)

    vs = analyzer.polarity_scores(str(s))

    compound.append(vs["compound"])

df_clean["compound"] = compound

df_clean.head(2)

scores = []
```

```
for i, s in tqdm(df_clean.iterrows(), total=df_clean.shape[0], position=0, leave=True):

    try:

        scores.append(s["compound"] * ((int(s["user_followers"]))) *
        ((int(s["user_favourites"])+1)/int(s["user_followers"]+1)) * ((int(s["is_retweet"])+1)))

    except:

        scores.append(np.nan)

df_clean["score"] = scores

print(scores)

df_clean.head(2)

# df_clean.to_csv("scored_file.csv")

df_price = pd.read_csv("bitcoin_historical_data.csv")

df_price.Date = pd.to_datetime(df_price.Date)

# df_price.Timestamp = pd.to_datetime(df_price.Timestamp,unit='s')

df_price.head(2)
```

## ДОДАТОК Б Код модулю для семантичної оцінки повідомлень

```

import matplotlib
matplotlib.use('Qt5Agg') # or 'TkAgg', or 'WXAgg'
import matplotlib.pyplot as plt
from time import sleep
import json
import pandas as pd
import io
import re
import numpy as np
from tqdm import tqdm
import pandas as pd
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
from tqdm import trange, tqdm_notebook, tqdm

from sklearn import preprocessing
# sentiment analysis
df_clean = df_clean.drop_duplicates()
tweets = df_clean.copy()
tweets['date'] = pd.to_datetime(tweets['date'],utc=True)
tweets.date = tweets.date.dt.tz_localize(None)
tweets.index = tweets['date']

# tweets_grouped = tweets.groupby(pd.TimeGrouper('1h'))['score'].sum()
tweets_grouped = tweets.resample('1h').sum()

crypto_usd = df_price.copy()
crypto_usd['Date'] = pd.to_datetime(crypto_usd['Date'], unit='s')
crypto_usd.index = crypto_usd['Date']
# crypto_usd['Timestamp'] = pd.to_datetime(crypto_usd['Timestamp'], unit='s')
# crypto_usd.index = crypto_usd['Timestamp']

# crypto_usd_grouped = crypto_usd.groupby(pd.TimeGrouper('1h'))['Weighted_Price'].mean()

```

```
crypto_usd_grouped = crypto_usd.resample('D')['Close'].mean()
```

## ДОДАТОК В Код модулю для скорінгової оцінки повідомлень

```

import matplotlib

matplotlib.use('Qt5Agg') # or 'TkAgg', or 'WXAgg'

import matplotlib.pyplot as plt

from time import sleep

import json

import pandas as pd

import io

import re

import numpy as np

from tqdm import tqdm

import pandas as pd

from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer

from tqdm import trange, tqdm_notebook, tqdm

from sklearn import preprocessing

# Normalize time series data
min_max_scaler = preprocessing.StandardScaler()
score_scaled = min_max_scaler.fit_transform(tweets_grouped['score'].values.reshape(-1,1))
# df_clean['score'] = score_scaled
# df_clean.to_csv('scored_file_new.csv')
tweets_grouped['normalized_score'] = score_scaled
print(tweets_grouped)
# tweets_grouped.to_csv('tweets_g.csv')
# crypto_used_grouped_scaled = min_max_scaler.fit_transform(crypto_usd_grouped.values.reshape(-1,1))
crypto_used_grouped_scaled = crypto_usd_grouped / max(crypto_usd_grouped.max(), abs(crypto_usd_grouped.min()))
# crypto_usd_grouped['normalized_price'] = crypto_used_grouped_scaled

```

## ДОДАТОК Д Код модулю для перевірки репрезентативності скорінгових оцінок

```
import matplotlib

matplotlib.use('Qt5Agg') # or 'TkAgg', or 'WXAgg'

import matplotlib.pyplot as plt

from time import sleep

import json

import pandas as pd

import io

import re

import numpy as np

from tqdm import tqdm

import pandas as pd

from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer

from tqdm import trange, tqdm_notebook, tqdm

from sklearn import preprocessing

df_clean = pd.read_csv('sorted_file.csv')

analyzer = SentimentIntensityAnalyzer()

compound = []

for i,s in enumerate(tqdm(df_clean['text'],position=0, leave=True)):

    # print(i,s)

    vs = analyzer.polarity_scores(str(s))

    compound.append(vs["compound"])

df_clean["compound"] = compound

df_clean.head(2)
```

```

scores = []

for i, s in tqdm(df_clean.iterrows(), total=df_clean.shape[0],position=0, leave=True):

    try:

        scores.append(s["compound"] * ((int(s["user_followers"]))) *
((int(s["user_favourites"])+1)/int(s["user_followers"]+1)) *((int(s["is_retweet"])+1)))

    except:

        scores.append(np.nan)

df_clean["score"] = scores

print(scores)

df_clean.head(2)

# df_clean.to_csv("scored_file.csv")

df_price = pd.read_csv("bitcoin_historical_data.csv")

df_price.Date = pd.to_datetime(df_price.Date)

# df_price.Timestamp = pd.to_datetime(df_price.Timestamp,unit='s')

df_price.head(2)

# sentiment analysis

df_clean = df_clean.drop_duplicates()

tweets = df_clean.copy()

tweets['date'] = pd.to_datetime(tweets['date'],utc=True)

tweets.date = tweets.date.dt.tz_localize(None)

tweets.index = tweets['date']

# tweets_grouped = tweets.groupby(pd.TimeGrouper('1h'))['score'].sum()

tweets_grouped = tweets.resample('1h').sum()

crypto_usd = df_price.copy()

```

```

crypto_usd['Date'] = pd.to_datetime(crypto_usd['Date'], unit='s')

crypto_usd.index = crypto_usd['Date']

# crypto_usd['Timestamp'] = pd.to_datetime(crypto_usd['Timestamp'], unit='s')

# crypto_usd.index = crypto_usd['Timestamp']

# crypto_usd_grouped = crypto_usd.groupby(pd.TimeGrouper('1h'))['Weighted_Price'].mean()

crypto_usd_grouped = crypto_usd.resample('D')['Close'].mean()

def crosscorr(datax, datay, lag=0, method="pearson"):
    """ Lag-N cross correlation.

    Parameters
    -----

    lag : int, default 0

    datax, datay : pandas.Series objects of equal length

    Returns
    -----

    crosscorr : float

    """
    return datax.corrwith(datay.shift(lag), method=method)['score']

# xcov = [crosscorr(tweets_grouped, crypto_usd_grouped, lag=i, m='pearson') for i in range(-20,20)]

# tweets_grouped.corrwith(crypto_usd_grouped,method='pearson')

beginning = max(tweets_grouped.index.min().replace(tzinfo=None), crypto_usd_grouped.index.min())

end = min(tweets_grouped.index.max().replace(tzinfo=None), crypto_usd_grouped.index.max())

tweets_grouped = tweets_grouped[beginning:end]

crypto_usd_grouped = crypto_usd_grouped[beginning:end]

```

```
fig, ax1 = plt.subplots(figsize=(20,10))

ax1.set_title("Crypto currency evolution compared to twitter sentiment", fontsize=18)

ax1.tick_params(labelsize=14)

ax2 = ax1.twinx()

ax1.plot_date(tweets_grouped.index, tweets_grouped, 'orange')

ax2.plot_date(crypto_usd_grouped.index, crypto_usd_grouped, 'blue')

ax1.set_ylabel("Sentiment", color='orange', fontsize=16)

ax2.set_ylabel("Bitcoin [$]", color='blue', fontsize=16)

plt.show()

xcov = [crosscorr(tweets_grouped, crypto_usd_grouped, lag=i, method="pearson") for i in range(-20,20)]

plt.plot(range(-20,20), xcov)

plt.title("pearson cross-correlation")

plt.xlabel("lag")

plt.ylabel("correlation")

plt.show()

xcov = [crosscorr(tweets_grouped, crypto_usd_grouped, lag=i, method="kendall") for i in range(-20,20)]

plt.plot(range(-20,20), xcov)

plt.title("kendall cross-correlation")

plt.xlabel("lag")

plt.ylabel("correlation")

plt.show()

xcov = [crosscorr(tweets_grouped, crypto_usd_grouped, lag=i, method="spearman") for i in range(-20,20)]

plt.plot(range(-20,20), xcov)

plt.title("spearman cross-correlation")
```

```
plt.xlabel("lag")
```

```
plt.ylabel("correlation")
```

```
plt.show()
```

## ДОДАТОК Е Код модулю для побудови прогнозів ціни криптовалюти на історичних даних

```

import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
import matplotlib
matplotlib.use('Qt5Agg') # or 'TkAgg', or 'WXAgg'
import matplotlib.pyplot as plt
import pmdarima as pm
import numpy as np
# Read the CSV file
data = pd.read_csv('merged_data.csv')

# Assuming the date is in a column named 'Date' and the price in 'Price'
data['Date'] = pd.to_datetime(data['Date'], format='%Y-%m-%d')
data.set_index('Date', inplace=True)
column = 'normalized_score'

# Calculate the minimum and maximum values of the 'score' column
min_value = data[column].min()
max_value = data[column].max()

# Set the new minimum and maximum values
new_min = 0.95
new_max = 1.05

# Normalize the 'score' column using the formula
for i in range(len(data['normalized_score'])):
    if data['normalized_score'][i] != -0.2048931397104815:
        data['normalized_score'][i] = (data[column][i] - min_value) * (new_max - new_min) / (max_value - min_value) +
        new_min
    else:
        data['normalized_score'][i] = 1

# Split the data into a training set and a testing set

```

```

split_point = len(data) -15
train_data = data['Close'][:split_point]
test_data = data['Close'][split_point:]
# test_dates = data['Date'][split_point:]

# Function to fit the ARIMA model and make a one-step prediction
def predict_one_step(train_data):
    model = pm.auto_arima(
        train_data,
        seasonal=False,
        stepwise=True,
        suppress_warnings=True,
        trace=False
    )
    forecast, conf_int = model.predict(n_periods=1, return_conf_int=True)
    return forecast[0], conf_int[0]

# Rolling forecast
predictions = []
conf_intervals = []
for i in range(len(test_data)):
    forecast, conf_int = predict_one_step(train_data)
    predictions.append(forecast)
    conf_intervals.append(conf_int)

# Add the actual value to the training set and continue
train_data = train_data.append(test_data.iloc[i:i + 1])
for i in range(len(predictions)):
    # print(f'date predictions {predictions[i]} score{data['normalized_score'][i]}')
    predictions[i] = predictions[i]

# Calculate the error using the mean absolute percentage error (MAPE)
# print(test_data.values)
print(f"
Cryptocurrency: Bitcoin
Timerange: - 2022-01-22

```

Predicted price values:

```
""")
for i in range(len(predictions)):
    print(f"
{predictions[i]}
")
mape = np.mean(np.abs((test_data.values - predictions) / test_data.values)) * 100
mse = np.mean(np.abs(((test_data.values - predictions) / test_data.values)** 2)) * 100
# Print the results
print(f"Mean Absolute Percentage Error (MAPE): {mape}%")
print(f"Mean Square Percentage Error (MSPE): {mse}%")

# Plot the results
plt.figure(figsize=(12, 6))
plt.plot(data['Close'], label='Actual Price')
plt.plot(test_data.index, predictions, label='Predicted Price')
plt.fill_between(
    test_data.index, np.array(conf_intervals[:, 0]), np.array(conf_intervals[:, 1]),
    color='pink', label='Confidence Interval'
)
plt.legend()
plt.title('Bitcoin Price Prediction using ARIMA (Rolling Forecast)')
plt.xlabel('Date')
plt.ylabel('Price')
plt.show()
```

## ДОДАТОК Є Код модулю для побудови прогнозів ціни криптовалюти на історичних даних та скорінгових оцінках повідомлень

```

import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
import matplotlib
matplotlib.use('Qt5Agg') # or 'TkAgg', or 'WXAgg'
import matplotlib.pyplot as plt
import pmdarima as pm
import numpy as np

# Read the CSV file
data = pd.read_csv('merged_data.csv')

# Assuming the date is in a column named 'Date' and the price in 'Price'
data['Date'] = pd.to_datetime(data['Date'], format='%Y-%m-%d')
data.set_index('Date', inplace=True)
column = 'normalized_score'

# Calculate the minimum and maximum values of the 'score' column
min_value = data[column].min()
max_value = data[column].max()

# Set the new minimum and maximum values
new_min = 0.95
new_max = 1.05

# Normalize the 'score' column using the formula
for i in range(len(data['normalized_score'])):
    if data['normalized_score'][i] != -0.2048931397104815:
        data['normalized_score'][i] = (data[column][i] - min_value) * (new_max - new_min) / (max_value - min_value) +
new_min
    else:
        data['normalized_score'][i] = 1

# Split the data into a training set and a testing set
split_point = len(data) - 15

```

```

train_data = data['Close'][:split_point]
test_data = data['Close'][split_point:]
# test_dates = data['Date'][split_point:]

# Function to fit the ARIMA model and make a one-step prediction
def predict_one_step(train_data):
    model = pm.auto_arma(
        train_data,
        seasonal=False,
        stepwise=True,
        suppress_warnings=True,
        trace=False
    )
    forecast, conf_int = model.predict(n_periods=1, return_conf_int=True)
    return forecast[0], conf_int[0]

# Rolling forecast
predictions = []
conf_intervals = []
for i in range(len(test_data)):
    forecast, conf_int = predict_one_step(train_data)
    predictions.append(forecast)
    conf_intervals.append(conf_int)

# Add the actual value to the training set and continue
train_data = train_data.append(test_data.iloc[i:i + 1])
for i in range(len(predictions)):
    # print(f'date predictions {predictions[i]} score{data['normalized_score'][i]}')
    predictions[i] = predictions[i] * data['normalized_score'][i]
# Calculate the error using the mean absolute percentage error (MAPE)
# print(test_data.values)
print(f"
Cryptocurrency: Bitcoin
Timerange: - 2022-01-22
Predicted price values:
")
for i in range(len(predictions)):
    print(f"
{predictions[i]}
")
mape = np.mean(np.abs((test_data.values - predictions) / test_data.values)) * 100

```

```
mse = np.mean(np.abs(((test_data.values - predictions) / test_data.values)** 2)) * 100
# Print the results
print(f"Mean Absolute Percentage Error (MAPE): {mape}%")
print(f"Mean Square Percentage Error (MSPE): {mse}%")

# Plot the results
plt.figure(figsize=(12, 6))
plt.plot(data['Close'], label='Actual Price')
plt.plot(test_data.index, predictions, label='Predicted Price')
plt.fill_between(
    test_data.index, np.array(conf_intervals)[:, 0], np.array(conf_intervals)[:, 1],
    color='pink', label='Confidence Interval'
)
plt.legend()
plt.title('Bitcoin Price Prediction using ARIMA (Rolling Forecast)')
plt.xlabel('Date')
plt.ylabel('Price')
plt.show()
```

## ДОДАТОК Ж Зразок вхідних даних про повідомлення

1765847,David Jeffries,,16 Year Old Investor • Entrepreneur • Be One Step Ahead • @meta\_dragons,2020-03-17 13:53:18,1116.0,346.0,7883.0,False,2021-11-19 15:28:10,"My answer when someone ask me what's that online currency you were telling me about 4 years ago!

Bitcoin /FjXmdTOZmm",['Bitcoin'],Twitter for iPhone,False

1980447,Guy Swann □,,"#Bitcoin ""The Guy who's read more about Bitcoin than anyone else you know."" □ Host of @BitcoinAudible and @ShitcoinInsider",2014-03-22 03:09:06,40419.0,2323.0,203442.0,False,2021-12-30 15:59:30,"I'm betting everything because it's our best chance, hands down. And I'm willing to risk it all to know that I did all I could to see a better future.

That said, I hold no illusions, nothing is guaranteed. But win or lose, Bitcoin is the only path worth fighting for.",['Bitcoin'],Twitter Web App,False

527605,SONDOUSSE,,https://t.co/Frh4gphMwc,2021-02-08 14:06:18,30.0,485.0,503.0,False,2021-07-19 12:32:28,"joined

Hswap DeFi HECO Airdrop Bitcoin",,['Hswap', 'DeFi', 'HECO', 'Airdrop', 'Bitcoin'],Twitter Web App,False

354046,Truth Shots,"McKinney, TX", "I'm Back,

Still don't give a damn about left or right shit",2021-03-27 18:35:58,58.0,191.0,6139.0,False,2021-07-18 14:55:15,"If you take vaccine God not gonna punish you □ □ □ CovidVaccine,['CovidVaccine'],Twitter for Android,False

27934,CTrading.io,Madrid - España,"Ctrading es una plataforma tecnológica de análisis, predicción y operativa avanzada en criptomercados para profesionales

#Trading #Criptomercados #BTC #ETH",2018-12-24 15:56:10,363.0,954,2215,False,2021-02-14 15:02:04,"□ DeFI Markets □

Plataform:Maker

Category:Lending

USD:\$6,587,069,240

Bitcoin BTC ETH Cryptocurrency",,['Bitcoin', 'BTC', 'ETH', 'Cryptocurrency'],Data Ctrading News,False

1029181,CryptoHorse,,"\$BTC \$ETH \$ALTS #Crypto #Diversify #KillEmotion #UseYourHead #DYOR #HaveNoIdeaWhatIdo

My tweets are not a financial advise.",2017-11-13 10:37:44,3119.0,1701.0,2103.0,False,2021-08-20 10:07:36,DXY time to fall and send BTC to 100k /2iptWsfmJc,['BTC'],Twitter Web App,False