

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

**Навчально-науковий
«Інститут інформаційних технологій в економіці»**

Кафедра математичного моделювання та статистика

Освітньо-професійна програма	«Економічна кібернетика»
Галузь знань	05 «Соціальні та поведінкові науки»
Спеціальність	051 «Економіка»

Форма навчання: очна (денна)
очна (денна)/дистанційна

КВАЛІФІКАЦІЙНА БАКАЛАВРСЬКА РОБОТА
на тему **«Прогнозування змін тренду котирувань криптовалют на основі
технологій Data Science»**

(назва теми)

здобувача Савченко Дарії Олександрівни
(ПІБ)

(підпис)

Науковий керівник: доктор економічних наук, професор
Матвійчук А.В.

(підпис)

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

В. о. завідувача кафедри: кандидат фізико-математичних наук,
професор Великоіваненко Г.І.

(підпис)

Київ 2024

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

**Навчально-науковий інститут
«Інститут інформаційних технологій в економіці»**

Кафедра математичного моделювання та статистики

Освітньо-професійна програма
Галузь знань
Спеціальність

«Економічна кібернетика»
05 «Соціальні та поведінкові науки»
051 «Економіка»

ПОГОДЖЕНО

Керівник проектної групи (гарант)
освітньо-професійної програми

_____ О.М. Притоманова
(підпис)

_____ 2024 р.

ЗАТВЕРДЖУЮ

В.о. завідувача кафедри математичного
моделювання та статистики

_____ Г.І. Великоіваненко
(підпис)

_____ 2024 р.

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ

здобувача вищої освіти Савченко Дарії Олександрівни

(прізвище, ім'я, по батькові)

_____ очної (денної) _____ **форми навчання**
очної (денної), дистанційної

на підготовку кваліфікаційної бакалаврської роботи

на тему

«Прогнозування змін тренду котирувань криптовалют на основі технологій Data Science»

Тему затверджено наказом ректора Університету від «__» _____ 20__ р. № _____

**Кваліфікаційна бакалаврська робота виконується на матеріалах
сайтів Binance (<https://www.binance.com/uk-UA>), Coinglass (<https://www.coinglass.com/>) та інших
відкритих джерел**

План кваліфікаційної бакалаврської роботи

Розділ 1	Теоретичні засади прогнозування динаміки криптовалют
Розділ 2	Формування бази даних та підготовка предикторів для моделювання
Розділ 3	Моделювання змін тренду котирувань криптовалют за допомогою методів машинного навчання та штучного інтелекту

Об'єкт дослідження:	часові ряди котирувань криптовалют
Предмет дослідження:	методи та моделі машинного навчання прогнозування змін тренду котирувань криптовалют
Мета виконання кваліфікаційної бакалаврської роботи:	розробка та впровадження ефективного підходу до прогнозування змін тренду котирувань криптовалют за допомогою методів машинного навчання та штучного інтелекту

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

У розділі 1: дослідження сучасного стану та тенденцій розвитку ринку криптовалют, аналіз новітніх наукових досліджень та передових методів прогнозування валютних котирувань, розробка концептуального підходу до прогнозування трендів змін криптовалют.

У розділі 2:

збір даних за криптовалютами BTC, ETH, ZRX, BOME, попередня обробка вихідних статистичних даних та формування на їх основі нових предикторів, вибір активів для формування X та Y для навчання моделей, аналіз критеріїв оцінки точності моделювання.

У розділі 3:

застосування логістичної регресії в задачі бінарної класифікації змін тренду котирувань криптовалют, застосування дерев рішень для класифікації трендів, побудова комплексу нейронних мереж, порівняльний аналіз побудованих моделей на тестових вибірках за такими критеріями як точність, коефіцієнт Джині для ROC-кривої, матриця помилок.

Завдання підготував
науковий керівник

(підпис)

А.В. Матвійчук
(ініціали, прізвище)
«__» _____ 2024 р.

Завдання одержав
здобувач

(підпис)

Д.О. Савченко
(ініціали, прізвище)
«__» _____ 2024 р.

Реферат

Кваліфікаційна бакалаврська робота містить 53 сторінки, 9 таблиць, 30 рисунків, перелік використаних джерел з 44 найменувань.

«Прогнозування змін тренду котирувань криптовалют на основі технологій Data Science»

Об'єктом дослідження є часові ряди котирувань криптовалют.

Предметом дослідження є методи та моделі машинного навчання прогнозування змін тренду котирувань криптовалют.

Мета та завдання виконання дослідження. Метою виконання кваліфікаційної бакалаврської роботи є розробка та впровадження ефективного підходу до прогнозування змін тренду котирувань криптовалют за допомогою методів машинного навчання та штучного інтелекту.

Відповідно до поставленої мети визначені такі *завдання*:

- провести аналіз сучасного стану ринку криптовалют;
- вивчити та порівняти існуючі методи прогнозування курсу котирувань криптовалют;
- розробити концептуальний підхід до прогнозування змін тренду криптовалют на основі машинного навчання;
- реалізувати побудовані моделі у програмному середовищі та протестувати їх на основі реальних даних;
- оцінити точність та ефективність моделей.

Теоретична, методична та практична значущість отриманих результатів.

Теоретична значущість роботи полягає в удосконаленні методів прогнозування змін тренду котирувань криптовалют, зокрема в застосуванні машинного навчання для аналізу часових рядів. Важливим елементом теоретико-методичної значущості роботи є розробка алгоритмів для попередньої обробки даних, що дозволяє підвищити точність та надійність прогнозів. У даній роботі розроблена методологія оцінки якості прогностичних моделей, що включає використання таких критеріїв як точність, ROC-AUC, індекс Джині.

Практична значущість полягає у можливості застосування результатів дослідження для прийняття інвестиційних рішень, що дозволить зменшити ризики та підвищити ефективність управління інвестиційним портфелем. Результати дослідження можуть бути використані для розробки ПЗ та аналітичних інструментів для трейдерів та інвесторів, що займаються торгівлею криптовалютами.

Рік виконання кваліфікаційної бакалаврської роботи – 2024. Рік захисту роботи – 2024.

Ключові слова: криптовалюта, прогнозування, тренд, котирування, машинне навчання, логістична регресія, дерево рішень, нейронні мережі.

В І Д Г У К

про кваліфікаційну бакалаврську роботу
здобувача навчально-наукового інституту
«Інститут інформаційних технологій в економіці»
освітньо-професійної програми «Економічна кібернетика»
Савченко Дарії Олександрівни

на тему **«Прогнозування змін тренду котирувань криптовалют на основі технологій Data Science»**
(назва теми)

1. Актуальність теми обумовлюється швидким розвитком ринку криптовалют, який характеризується високою волатильністю та значними ризиками для інвесторів. Використання методів машинного навчання та штучного інтелекту для прогнозування динаміки криптовалют може надати нові можливості для ефективного управління інвестиційними портфелями, зменшення ризиків та підвищення прибутковості. Застосування сучасних технологій аналізу даних дозволяє не лише підвищити точність прогнозів, але й своєчасно реагувати на зміни ринкових умов, що є критично важливим в умовах нестабільного ринку криптовалют. Це вказує на актуальність та своєчасність досліджень у даному напрямку.

2. Позитивні риси кваліфікаційної роботи Савченко Д.О. включають глибокий аналіз сучасного стану ринку криптовалют, основних тенденцій та викликів, методів машинного навчання, які застосовуються для прогнозування фінансових показників. Проведено порівняльний аналіз методів логістичної регресії, дерев прийняття рішень та нейронних мереж, що дозволило обрати ефективний інструментарій для прогнозування динаміки криптовалют. Автор здійснила ретельний збір та підготовку даних, що включало аналіз та відбір предикторів, необхідних для побудови моделей. Студентка чітко визначила мету та поставила задачі, які в кінцевому результаті були виконані.

3. Наявність самостійних розробок автора. Автором бакалаврської роботи самостійно проведено аналіз стану та перспектив розвитку ринку криптовалют, наукових праць з прогнозування фінансових показників, у тому числі із застосуванням методів машинного навчання, сформовано базу даних для проведення моделювання. Розроблено комплекс моделей прогнозування змін тренду котирувань криптовалют на основі методів дерев рішень, моделей

логістичної регресії та нейронних мереж різних архітектур. Практична реалізація моделей була здійснена за допомогою сучасних інструментів, а саме мови програмування Python та спеціалізованих бібліотек для обробки даних та моделювання – Pandas, Sckit-learn, plotly, seaborn, keras. Моделі були протестовані на реальних даних з біржі Binance, що підтвердило їхню ефективність.

4. Розроблені концептуальні аспекти, моделі та методи реалізують оригінальний авторський підхід до прогнозування не точних значень фінансового інструменту, а напрямку його зміни, що дозволило звести задачу прогнозування до бінарної класифікації та застосувати відповідні підходи до моделювання, включаючи специфічні способи обробки даних. Запропоновані методи сприятимуть зниженню ризиків та підвищенню ефективності управління інвестиційними портфелями.

5. Наявність недоліків. Було б корисно розглянути більше методів аналізу динаміки криптовалют для порівняння ефективності різних підходів, зокрема згорткових нейронних мереж та інших типів моделей глибокого навчання, призначених для розпізнавання зображень (у даному випадку графіків цінкових кривих).

6. Загальна оцінка кваліфікаційної магістерської роботи. Кваліфікаційна бакалаврська робота студентки Савченко Дарії Олександрівни виконана у відповідності з завданням із дотриманням всіх вимог. Робота має бути допущена до захисту перед ЕК та заслуговує на високу оцінку.

Науковий керівник: доктор економічних наук, професор, професор кафедри економіко-математичного моделювання Київського національного економічного університету імені Вадима Гетьмана.

Андрій МАТВІЙЧУК

(підпис)

“ ____ ” _____ 20__ р.

ЗМІСТ

ВСТУП.....	3
РОЗДІЛ 1. ТЕОРЕТИЧНІ ЗАСАДИ ПРОГНОЗУВАННЯ ДИНАМІКИ КРИПТОВАЛЮТ	6
1.1 Дослідження сучасного стану та тенденцій розвитку ринку криптовалют.....	6
1.2 Аналіз останніх наукових досліджень та передових методів прогнозування валютних котирувань	12
1.3 Розробка концептуального підходу до прогнозування трендів змін криптовалют.....	17
РОЗДІЛ 2. ФОРМУВАННЯ БАЗИ ДАНИХ ТА ПІДГОТОВКА ПРЕДИКТОРІВ ДЛЯ МОДЕЛЮВАННЯ	20
2.1 Збір даних та їх аналіз	20
2.2 Формування предикторів та індикаторів	23
2.3 Аналіз критеріїв оцінки точності моделювання	29
РОЗДІЛ 3. МОДЕЛЮВАННЯ ЗМІН ТРЕНДУ КОТИРУВАНЬ КРИПТОВАЛЮТ ЗА ДОПОМОГОЮ МЕТОДІВ МАШИННОГО НАВЧАННЯ ТА ШТУЧНОГО ІНТЕЛЕКТУ.....	32
3.1 Застосування логістичної регресії в задачі бінарної класифікації змін тренду котирувань криптовалют	32
3.2 Застосування дерева рішень для класифікації трендів	36
3.3 Моделювання змін тренду криптовалют на основі нейронних мереж різних архітектур.....	41
ВИСНОВКИ	52
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	54
ДОДАТКИ.....	59

ВСТУП

Криптовалюти є одним із важливих технологічних винаходів сучасного світу, що привернув значну увагу як серед фінансових експертів, так і серед звичайних людей. Хоча інвестиції в криптовалюту вважаються досить ризикованими, багато інвесторів позитивно ставляться до їх використання та оптимістично оцінюють перспективи криптовалют. В умовах високої волатильності ринку криптовалют, ефективні методи прогнозування їхніх трендів є надзвичайно актуальними.

Прогнозування трендів криптовалют є актуальною задачею для трейдерів та інвесторів. У багатьох наукових роботах було розглянуто різні підходи до вирішення цього завдання.

Аналізуючи останні наукові дослідження для прогнозування курсу криптовалют з боку прогнозування часових рядів, варто зазначити, що застосовуються різні підходи. Експерименти показують, що найбільш ефективними та перспективними методами для прогнозування курсу криптовалют є методи машинного навчання, зокрема нейронні мережі.

Метою виконання кваліфікаційної роботи є розробка та впровадження ефективного підходу до прогнозування змін тренду котирувань криптовалют за допомогою методів машинного навчання та штучного інтелекту.

Для досягнення цієї мети необхідно виконати наступні *завдання*:

- провести аналіз сучасного стану ринку криптовалют;
- вивчити та порівняти існуючі методи прогнозування курсу котирувань криптовалют;
- розробити концептуальний підхід до прогнозування змін тренду криптовалют на основі машинного навчання;
- реалізувати побудовані моделі у програмному середовищі та протестувати їх на основі реальних даних;

- оцінити точність та ефективність моделей.

Об'єктом дослідження є часові ряди котирувань криптовалют.

Предметом дослідження є методи та моделі машинного навчання прогнозування змін тренду котирувань криптовалют.

Для досягнення поставленої мети в кваліфікаційній роботі використано наступні методи:

- логістична регресія використовується для задач бінарної класифікації, таких як передбачення зростання або падіння курсу криптовалюти;
- дерева рішень, що використовується для класифікації та регресії часових рядів криптовалют;
- штучні нейронні мережі, які використовуються для виявлення складних закономірностей при вирішенні задач класифікації та регресії.

Теоретико-методична значущість. Теоретична значущість роботи полягає в удосконаленні методів прогнозування змін тренду котирувань криптовалют, зокрема в застосуванні машинного навчання для аналізу часових рядів.

Важливим елементом теоретико-методичної значущості роботи є розробка алгоритмів для попередньої обробки даних, що дозволяє підвищити точність та надійність прогнозів. У даній роботі розроблена методологія оцінки якості прогностичних моделей, що включає використання таких критеріїв як точність, ROC-AUC, індекс Джині.

Практична значущість полягає у можливості застосування результатів дослідження для прийняття інвестиційних рішень, що дозволить зменшити ризики та підвищити ефективність управління інвестиційним портфелем. Результати дослідження можуть бути використані для розробки ПЗ та аналітичних інструментів для трейдерів та інвесторів, що займаються торгівлею криптовалютами.

Дані для виконання бакалаврської роботи було взято з біржі Binance. База даних включає різноманітні фактори, які використовуються для прогнозування змін тренду котирувань криптовалют.

Кваліфікаційна робота складається зі вступу, трьох розділів, висновку та списку використаних джерел. Усі моделі створювалися в середовищі Jupyter Notebook із використанням мови програмування Python.

РОЗДІЛ 1

ТЕОРЕТИЧНІ ЗАСАДИ ПРОГНОЗУВАННЯ ДИНАМІКИ КРИПТОВАЛЮТ

1.1 Дослідження сучасного стану та тенденцій розвитку ринку криптовалют

За останні десять років широкого поширення набув новий фінансовий термін - «*криптовалюта*», який виступив як реакція на недоліки традиційної фінансової системи. Держава прагне здійснювати контроль над всіма операціями, включаючи готівкові, безготівкові та депозитні перекази та різні розрахунки з метою їх подальшого оподаткування. Поява цього сучасного фінансового інструменту є цілком логічною відповіддю на дії держави та міжнародних регуляторів, що спрямовані на підвищення прозорості фінансових операцій і боротьбу з корупцією, відмиванням коштів та фінансуванням тероризму, а також усунення затримок у процесі обробки платежів у системі електронних розрахунків фізичних та юридичних осіб [2].

У часи розвитку цифрових технологій криптовалюти привернули увагу як окремих інвесторів, так і організацій по всьому світу завдяки своїм новаторським підходам до фінансових операцій. Хоч криптовалюти не виконують традиційні функції грошей і не є стабільним засобом для інвестування, вони відомі як один із найспекулятивніших та найбільш ризикованих фінансових інструментів, пропонуючи потенційно високі прибутки поряд із високими ризиками.

Криптовалюта – це форма цифрової або віртуальної валюти, що забезпечує безпеку транзакцій за допомогою криптографії та функціонує на базі децентралізованої технології [3]. Особливістю криптовалюти є те, що вона не має

фізичної форми – купюри або монети, тому не може бути обмінною як звичайні гроші. Відмінності цифрових активів від традиційних наступні:

- криптовалюта функціонує без центрального органу управління та емісії (теоретично це робить її стійкою до державного втручання);
- криптовалюта не залежить від будь-якої державної валюти;
- висока волатильність криптовалюти в порівнянні з реальними грошима;
- криптовалюти неможливо підробити на відміну від фізичних коштів.

Першою запровадженою криптографічною валютою вважається Bitcoin, яка була випущена в 2009 році анонімною особою або групою людей під псевдонімом «Сатоші Накамото». В 2010 році ціна Bitcoin становила приблизно 0,44 дол. США, а вже в листопаді 2013 року ціна перевищила 1000 дол. США. Ще в 2016 році ціна коливалась до 1000 дол. США, але з 16 грудня 2017 року ціна Bitcoin з 2000 дол. США скочила до 19 188 дол. США [5]. Свій історичний максимум у 73,664 дол. США Bitcoin досягнув 13 березня 2024 року, що показано на рис. 1.1.

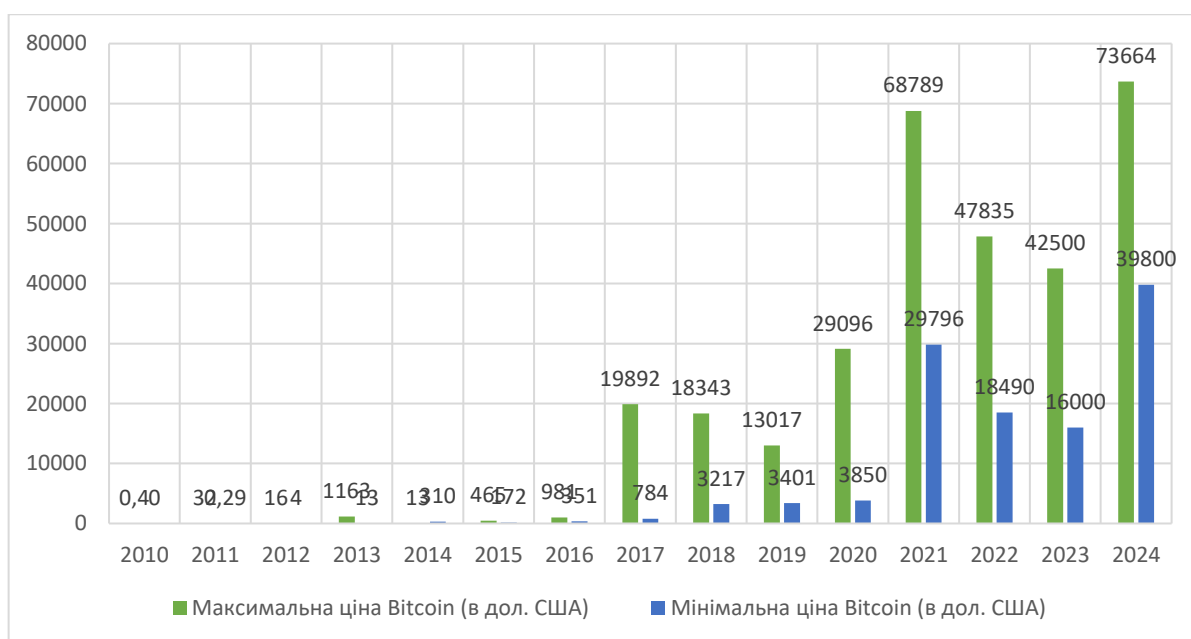


Рисунок 1.1 – Динаміка росту Bitcoin в USD за 2017-2024 роки

Джерело: розроблено автором на основі [6]

Станом на кінець 2021 року на ринку існувало близько 0,44 мільйона криптовалют. Кількість криптовалют продовжувала зростати і в кінці 2023 року досягла 1,98 мільйони, що на 72,3% більше від показника попереднього року. На початку 2024 року кількість криптовалют перевищила 2,52 мільйони, що у 5,7 разів більше порівняно з кінцем 2021 року[10]. Широку популярність криптовалют можна пояснити їх високою ринковою ліквідністю, доступністю купівлі, продажу, передачі та надсилання, а також здатністю забезпечити безпечне зберігання.

На сьогодні Bitcoin залишається лідером за ринковою капіталізацією серед усіх криптовалют, однак інші монети активно намагаються закріпитись на ринку, пропонуючи відмінні послуги, які відрізняються від Bitcoin, всі інші криптовалюти, крім Bitcoin, називають "альткоїнами".

За даними, наведеними в табл. 1.1, видно, що Ethereum посідає друге місце за ринковою капіталізацією, ще однією монетою, яка також відіграє важливу роль у розширенні криптовалютного ринку, є Dogecoin, відомий своєю здатністю обробляти велику кількість транзакцій.

Таблиця 1.1 - Топ-10 криптовалют за ринковою капіталізацією станом на травень 2024 року

№	Криптовалюта	Ринкова капіталізація
1	Bitcoin (BTC)	1,25 трлн. дол. США
2	Ethereum (ETH)	367 млрд. дол. США
3	Tether (USDT)	109,6 млрд. дол. США
4	Binance Coin (BNB)	83 млрд. дол. США
5	Solana (SOL)	64 млрд. дол. США
6	U.S. Dollar Coin (USDC)	33,8 млрд. дол. США
7	XRP (XRP)	27,9 млрд. дол. США
8	Dogecoin (DOGE)	22,5 млрд. дол. США
9	Toncoin (TON)	21 млрд. дол. США
10	Cardano (ADA)	17,1 млрд. дол. США

Джерело: розроблено автором на основі [7]

Більшість криптовалют працюють на основі блокчейн технології, що є видом розподіленого реєстру і функціонує як певний тип бази даних. Блокчейн – це база даних, яка документує кожну транзакцію, що коли-небудь відбувалась в мережі, забезпечуючи захист від несанкціонованого впливу. Дослівно, *blockchain* складається з блоків («block») – інформації про серію транзакцій, здійснених протягом певного часового проміжку, з'єднаних у ланцюг («chain»). Вузол – це комп'ютер або сервер, що підключений до мережі блокчейн. Всі транзакції вважаються «підтвердженими», коли одержувач отримав токен від відправника, інакше транзакція залишається «непідтвердженою». Безпека блокчейна забезпечується тим, що кожен блок у ланцюгу пов'язаний з попереднім. Для зламу всього ланцюга, шахраям потрібно отримати контроль над більшістю вузлів мережі, що є достатньо складним завданням [8]. Завдяки децентралізованій структурі, забезпечується відкритість та прозорість, оскільки кожен учасник має повний запис бази даних, яку підтримує кожен вузол мережі.

З впровадженням криптовалюти, динаміка транзакцій в блокчейні постійно зростає. Ця нова форма валюти здобула велику популярність завдяки своїй автономності та зручності і як наслідок, блокчейн все частіше використовують для різних фінансових операцій (рис. 1.2).



Рисунок 1.2 – Динаміка щоденних транзакцій на блокчейні Bitcoin, січень 2010 - березень 2024

Джерело: розроблено автором на основі [9]

Спираючись на наведені дані, можна проаналізувати динамічне зростання популярності криптовалют, зокрема Bitcoin. Вплив криптовалюти на традиційні валютні системи та методи розрахунку пов'язаний зі скептичним ставленням державних урядів та центральних банків до цифрових грошей. Оскільки держави не мають можливості повного контролю криптовалюти, вони втрачають можливість використовувати криптовалюту для власних цілей, наприклад як експропріація [11], що можливо у випадку з національною валютою. Труднощі з податковою звітністю та негативні реакції на криптовалюту призвели до повної заборони цифрових активів у таких країнах, як Катар, Саудівська Аравія та Китай, остання з яких була світовим лідером з майнінгу біткоїнів.

На рис. 1.3 наведено показники володіння криптовалютою для 10 країн у відсотковому значенні. Лідером цього списку є Об'єднані Арабські Емірати (ОАЕ), які запровадили правила купівлі-продажу криптовалюти та стали першими, хто запровадив нульові податки для її власників. У В'єтнамі велика кількість населення не має доступу до банківських послуг, тому криптовалюти можуть слугувати альтернативним способом здійснення фінансових операцій.



Рисунок 1.3 – Топ-10 країн за володінням криптовалют у 2023 році

Джерело: розроблено автором на основі [12]

За останні роки ринок криптовалют в Україні активно зростає, незважаючи на воєнні умови, і входить до трійки лідерів за обсягами криптовалютних транзакцій у світі. У 2023 році набув чинності Закон України «Про віртуальні активи», що остаточно сприяло легалізації криптовалют. Цей законодавчий акт є фундаментом для розвитку ринку віртуальних криптовалют в Україні, особливо в умовах повномасштабного вторгнення РФ.

Криптовалюти мають свої специфічні характеристики, які впливають на їх використання у світі. Завдяки впровадженню криптографічних методів, транзакції з криптовалютами є безпечними та анонімними, що забезпечує високий рівень приватності користувачів.

Однак ця характеристика відкриває можливості для використання криптовалют з метою фінансування незаконного бізнесу та проведення шахрайських схем, що створює виклики для державних регуляторів. Багато урядів не поспішають інтегрувати криптовалюти у свої фінансові системи, ставлячи під сумнів їхню стабільність та законність.

Одним із важливих аспектів криптовалют є висока волатильність. Різка зміна ціни за короткий проміжок часу створює ризики як для інвесторів, так і для звичайних користувачів. Висока ймовірність втрат знижує привабливість криптовалют як засобу збереження вартості.

Основною метою криптовалюти є модернізація та спрощення здійснення операцій у платіжній системі. Однак, постійне збільшення обсягів криптовалютних операцій вимагає правових рішень з боку фінансових регуляторів. Метою їх впровадження щодо регулювання операцій з криптовалютою є досягнення балансу між захистом фінансових ринків та створенням умов для їх розвитку.

1.2 Аналіз останніх наукових досліджень та передових методів прогнозування валютних котирувань

Прогнозування курсу криптовалют стає все більш важливим завданням на сучасному фінансовому ринку, особливо в контексті динамічного розвитку технологій та збільшення інтересу до криптовалют з боку фінансових аналітиків, інвесторів, трейдерів. Висока волатильність цін на криптовалюти та відсутність юридичного регулювання операцій з ними у багатьох країнах світу призвели до значних інвестиційних ризиків у сфері криптоактивів [15].

Тому розробка та застосування точних та надійних методів прогнозування курсу криптовалют є важливим питанням для забезпечення стабільності фінансової системи та ефективного управління інвестиційним портфелем. Криптовалюти відрізняються високою волатильністю в порівнянні з традиційними валютами, демонструють спекулятивну поведінку, мають обмеження на обсяги операцій та час їх обробки [16].

Ці властивості зумовлені різними причинами. Криптовалюти відрізняються від традиційних валют, оскільки вони не видаються центральними банками та не прив'язані до економіки конкретної країни, тому їхня вартість не базується на будь-якому матеріальному активі. Періоди високої волатильності також пов'язані з хакерськими атаками, кіберзлочинами та регуляторною невизначеністю. Тому для трейдерів та інвесторів важливо розробити ефективний підхід до управління волатильністю криптовалют, що включає в себе дієві методи прогнозування цін на фінансові активи [17].

Традиційний підхід до прогнозування динаміки фінансових інструментів полягає у використанні технічного аналізу, який ґрунтується на вивченні історичних даних динаміки цін. Шляхом аналізу часових рядів за допомогою виявлення паттернів та трендів, технічний аналіз дозволяє передбачити майбутні зміни в цінах. Завдяки цьому підходу інвестори можуть ідентифікувати вдалі

моменти для купівлі або продажу конкретної монети. Технічний аналіз включає в себе використання різних індикаторів, таких як ковзні середні, RSI (Relative Strength Index), MACD (Moving Average Convergence Divergence), стохастичний осцилятор, а також аналіз трендів та визначення рівнів підтримки та опору [19].

Іншим підходом до прогнозування курсу криптовалют є фундаментальний аналіз, що вивчає динаміку цін під впливом макроекономічних чинників [20]. Це включає в себе аналіз економічних, політичних та фінансових факторів, які впливають на ціну активів. Фундаментальний аналіз використовується інвесторами для реалізації довгострокових стратегій, дозволяючи використовувати довготривалі тренди для заробітку. Даний підхід ґрунтується на дослідженні та аналізі трендів, цінних паперів та валют, а також у визначенні чинників і складних взаємозв'язків, що впливають на його розвиток. Основний принцип фундаментального аналізу полягає у побудові причинно-наслідкової казуальної моделі, що ілюструє взаємозв'язки між курсовою вартістю активу та макроекономічними індикаторами [18].

Варто згадати про інший метод прогнозування динаміки фінансових інструментів, що базується на використанні економіко-математичних моделей, та включає аналіз часових рядів. Цей підхід застосовує вивчення історичних даних часових рядів та розробку прогнозів, базуючись на аналізі та обробці минулих спостережень. Серед стандартних моделей часових рядів виділяються моделі ARIMA, Бокса-Дженкінса та їх модифікації, модель GARCH.

Авторегресійна інтегрована ковзна середня (ARIMA), яка є узагальненою моделлю ARMA (Авторегресійна ковзна середня), – це статистичний метод прогнозування часових рядів, що поєднує авторегресію (AR), інтеграцію (I) та ковзне середнє (MA). На відміну від класичної моделі ARMA, яка припускає стаціонарність часових рядів, ARIMA застосовує стаціонарність як додаткову умову та використовує диференціювання для видалення сезонності та трендів, щоб перетворити часові ряди на стаціонарну послідовність даних [22]. Переваги ARIMA полягають в аналізі та прогнозі часових рядів з трендами та сезонністю, а

також у ефективному вмінні справлятися з нестационарними рядами. Однак, такий підхід до прогнозування часових рядів використовує статистичні методи та має певні обмеження. На криптовалютному ринку спостерігаються динамічні залежності та нелінійності в даних, тому такі методи можуть бути недостатньо ефективними для точного прогнозування.

Останнім часом до задач прогнозування фінансових часових рядів активно застосовуються методи машинного навчання (Machine Learning, ML), оскільки вони дозволяють будувати складніші моделі порівняно з класичними статистичними методами. Серед поширених методів машинного навчання можна виділити нейронні мережі (Artificial Neural Network, ANN) та перцепції (Support Vector Regression), які ефективні для прогнозування фінансових активів, зокрема цифрових валют. Це підтверджує дослідження [23], у якому порівнюється LSTM та ARIMA моделі для прогнозування часових рядів криптовалют. Результати показали, що моделі, розроблені з використанням LSTM, мають вищу точність, ніж традиційні моделі, такі як ARIMA. Це робить їх ефективним інструментом для прогнозування курсу криптовалют

Інший метод ML - дерево рішень, яке може бути використане для задач класифікації (дерево класифікації) або регресії (дерево регресії) (Classification and Regression Tree, C&RT). Цей метод ефективно використовує розширений набір даних для побудови моделі та здатний визначати логічні зв'язки та закономірності у часових рядах.

Завдяки здатності обробляти великі обсяги даних та виявляти паттерни, методи машинного навчання стають важливим інструментом для інвесторів та трейдерів, які зацікавлені в аналізі та передбаченні майбутніх трендів та цін цифрових активів. Такі методи забезпечують глибокий аналіз та адаптацію до швидких змін, характерних для ринку криптовалют.

Машинне навчання тісно пов'язане зі штучним інтелектом та зосереджується на розробці алгоритмів, здатних навчатися з даних та робити прогнозування, класифікацію, тощо. Однією з відмінностей від традиційного

моделювання є здатність алгоритмів машинного навчання виявляти та використовувати паттерни у даних, що зменшує необхідність у ручному аналізі та декомпозиції даних [24]. Крім того, методи ML та ШІ є дуже важливими для розвитку технологій на основі блокчейн. Наприклад, фінансові установи використовують алгоритми машинного навчання для підвищення якості обслуговування клієнтів шляхом персоналізації обслуговування та зменшення витрат часу.

Алгоритми машинного навчання в основному поділяються на чотири категорії:

- контрольоване навчання, навчання з учителем (*Supervised Learning*);
- навчання без учителя (*Unsupervised learning*);
- напівкероване навчання (*Semi-supervised Machine Learning*);
- навчання з підкріпленням (*Reinforcement Learning*).

Навчання з учителем (*Supervised Learning*) - це парадигма машинного навчання, в якій модель тренують на основі вхідних даних (наприклад, вектор передбачувальних змінних) та бажаних значень виходу. Це дозволяє створити систему, здатну здійснювати прогнози на основі інформації в умовах невизначеності.

Навчання без учителя орієнтоване на виявлення прихованих паттернів у великих наборах даних [25]. У цьому випадку модель аналізує дані без попередньо визначених міток, самостійно виявляючи структуру або закономірності в них.

Таблиця 1.2 - Категорії та алгоритми машинного навчання

Категорія ML	Функція	Завдання	Приклади алгоритмів
<i>Supervised Learning</i>	Набір даних має мітки та результати, використовується для прогнозування на основі даних	1) класифікація 2) регресія	1) Лінійна регресія 2) Логістична регресія 3) Випадковий ліс 4) Нейронні мережі

Продовження таблиці 1.2

<i>Unsupervised learning</i>	Набір даних не має міток, виявляє ключові взаємозв'язки	1) кластеризація 2) асоціація	1) Кластеризація 2) Метод головних компонент 3) Т-розподілене вкладення стохастичної близькості 4) Правило асоціації
<i>Semi-supervised Machine Learning</i>	Поєднання навчання з учителем і без учителя	1) класифікація 2) кластеризація	1) uClassify 2) GATE
<i>Reinforcement Learning</i>	Використовує функцію винагороди для виконання завдань	1) класифікація 2) контроль	1) Q-навчання 2) Дерево пошуку Монте-Карло 3) Метод часових різниць 4) Asynchronous Actor-Critic Agents

Джерело: розроблено автором на основі [25] та [27]

Для прогнозування динаміки цін криптовалют застосовують різні методи машинного навчання, зокрема:

- логістична регресія (*Logistic Regression*);
- лінійна регресія (*Linear regression*);
- метод опорних векторів (*Support Vector Machine*);
- дерево рішень (*Decision Tree*);
- випадковий ліс (*Random Forest*).
- штучні нейронні мережі (ANN).

Найбільш поширений підхід до розробки непараметричної та нелінійної класифікації та регресії ґрунтується на штучних нейронних мережах (*Artificial Neural Networks, ANNs*), особливе місце займають рекурентні нейронні мережі, зокрема архітектура Long Short-Term Memory (LSTM).

Long Short-Term Memory - це архітектура рекурентної нейронної мережі (RNN), яка широко використовується для аналізу та прогнозування задач часових рядів. Мережі LSTM застосовуються у різних галузях, таких як: розпізнавання та обробка природної мови, розпізнавання та генерування описів зображень і т.п. Даний метод продемонстрував ефективність для прогнозування цін на криптовалюту. У дослідженні [23] представлено результати застосування

класичної моделі ARIMA для прогнозування ціни Bitcoin. У роботі Латіфа було показано, що навіть за умови стабільної тенденції ціни криптовалюти, LSTM надає більш точні прогнози. Хоч ARIMA і показує задовільні результати за деякими метриками помилок (див. рис. 1.4), LSTM все ж краще прогнозує ціни, оскільки може передбачати як напрям, так і ціну часових рядів. Натомість ARIMA обмежується відслідковуванням тренду Bitcoin та не має можливості точно прогнозувати актуальні ціни.

TABLE 2

Performance Measurement Metrics				
	MAE	MAPE	RMSE	Accuracy
ARIMA (3,1,3)	837.77	1.79%	940.40	98.21%
RNN-LSTM	126.97	0.27%	151.95	99.73%

Рисунок 1.4 – Результати обчислень метрик ефективності для моделей ARIMA та LSTM

Джерело: [23, с.274]

1.3 Розробка концептуального підходу до прогнозування трендів змін криптовалют

На рис. 1.4 зображено концептуальну схему процесу машинного навчання за допомогою поширеної методології Міжгалузевого стандартного процесу обміну даними (*Cross-Industry Standard Process for Data Mining, CRISP-DM*). Процедура CRISP-DM складається з 6 кроків, які разом з даними утворюють ітеративний процес, що взаємопов'язує всі етапи.

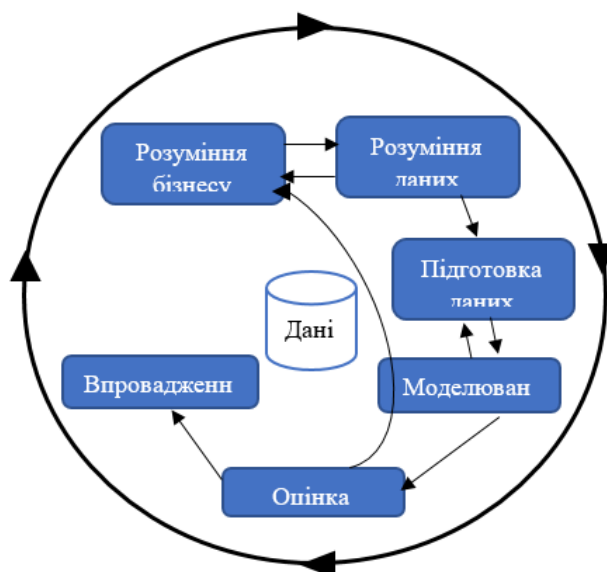


Рисунок 1.4 – Методологія CRISP-DM

Джерело: розроблено автором на основі [29]

1. Розуміння бізнесу (Business understanding). Ця початкова фаза, у якій досліджується мета бізнесу (проєкту), оскільки без чіткого розуміння цілей неможливо здійснити подальші кроки, такі як збір та аналіз даних, побудова моделей тощо. На початковій фазі варто враховувати різні аспекти, такі як волатильність криптовалют, яка значно впливає на результати аналізу. Також слід розглянути типи криптовалют (біткоїн та альткоїни). Зібрана інформація допоможе у постановці задачі та розробці подальшого плану дій для досягнення бізнес-цілей.

2. Розуміння даних (Data understanding). Цей етап починається з виявлення, збору та аналізу даних з різних джерел та за різними характеристиками для подальшого дослідження, які допоможуть досягти цілей проєкту. Для даної роботи джерелом даних є криптовалютна біржа Binance.

3. Підготовка даних (Data preparation). Ця фаза забезпечує придатність даних для подальшого моделювання та коригує відсутні або аномальні значення. На цьому етапі відбувається вивчення структури та якості даних, побудова наборів даних, за потреби їх очищення та редагування. Хороша підготовка даних має вирішальне значення для створення моделей, які мають високу точність і

ефективність. У цій роботі дані включають в себе різні характеристики, такі як ціни відкриття, закриття, максимум, мінімум тощо.

4. Моделювання (Modeling). На цьому етапі необхідно обрати методи відповідно до поставлених завдань дослідження та визначити параметри моделі, що найкраще підходять для обраного типу задач. Вибір моделі також залежить від розміру навчальних даних, лінійності даних, а також кінцевої мети дослідження.

5. Оцінка (Evaluation). Дана фаза характеризується проведенням оцінки ефективності побудованих моделей. Важливо оцінити результати та виявити можливі помилки у роботі моделі, а також обґрунтувати доречність її вибору.

6. Впровадження моделі (Deployment), яка пройшла оцінку та аналіз, і може застосовуватись для вирішення поставлених бізнес-цілей. Головною метою цієї фази є забезпечення стабільної працездатності застосованої моделі.

Вибір оптимальної моделі залежить від конкретних характеристик даних та цілей. У цій роботі буде детально описано реалізацію та оцінку різних методів з акцентом на аналізі точності прогнозування, що дозволить визначити найбільш ефективний підхід до прогнозування динаміки цін на криптовалюту.

РОЗДІЛ 2

ФОРМУВАННЯ БАЗИ ДАНИХ ТА ПІДГОТОВКА ПРЕДИКТОРІВ ДЛЯ МОДЕЛЮВАННЯ

2.1 Збір даних та їх аналіз

Для прогнозування змін тренду котирувань криптовалют було обрано часовий проміжок з 22.02.2023 по 23.02.2024 та інтервалом в 4 години, оскільки на цьому часовому зрізі відбулися нові локальні екстремуми за останній рік для таких монет як BTC та ETH та відбувся лістинг нових монет, таких як BOME, PEPE та інші.

Історичні дані змін котирувань певного активу було взято з сайту Binance, що представлені у вигляді японських свічок (рис. 2.1) та у табличному форматі (рис. 2.2) наступними критеріями:

- *open_time* – дата та час відкриття періоду торгівлі;
- *open* – ціна відкриття, ціна монети на момент початку торговельного періоду;
- *high* – найвища ціна активу протягом періоду торгівлі;
- *low* – найнижча ціна активу протягом періоду торгівлі;
- *close* – ціна закриття (ціна активу на момент закінчення торговельного періоду);
- *volume* – обсяг активу, який був куплений/проданий протягом періоду торгівлі;
- *close_time* – час закриття періоду;
- *q_a_v* (quote asset volume) – обсяг торгів у котирувальній валюті, що був обміняний;

- n_{tr} (number of trades) – кількість торгів, яка відбулась протягом вказаного періоду;
- $t_{b_a_v}$ (taker buy asset volume) – обсяг активів, придбаних учасником («taker»);
- $t_{b_q_a_v}$ (taker buy quote asset volume) – обсяг у котирувальній валюті, придбаний учасником («taker»).

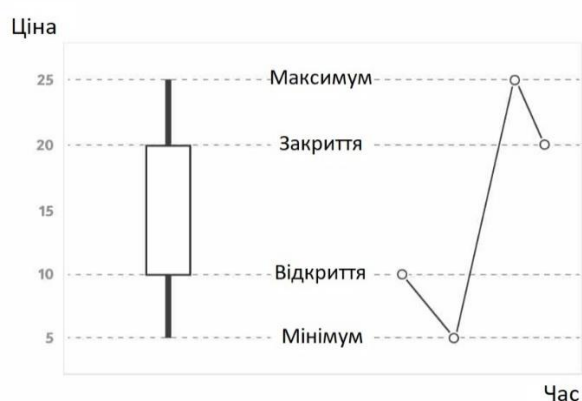


Рисунок 2.1 – Схема японської свічки

Джерело: [30]

BTC.head()												
	open	high	low	close	volume	close_time	q_a_v	n_tr	t_b_a_v	t_b_q_a_v	i	
open_time												
2024-02-22 14:00:00+02:00	51659.1	51766.0	50967.5	51630.0	65259.700	1708617599999	3.351856e+09	797161	32649.831	1.676998e+09	0	
2024-02-22 18:00:00+02:00	51630.0	51785.8	51424.0	51547.0	29668.492	1708631999999	1.531468e+09	409154	14509.290	7.490907e+08	0	
2024-02-22 22:00:00+02:00	51546.9	52130.8	51203.7	51316.9	49747.511	1708646399999	2.571518e+09	614926	24441.999	1.263805e+09	0	
2024-02-23 02:00:00+02:00	51316.9	51576.3	50989.9	51129.7	28843.394	1708660799999	1.478000e+09	397493	13206.624	6.769509e+08	0	
2024-02-23 06:00:00+02:00	51129.8	51352.6	50900.2	50952.3	26692.668	1708675199999	1.365567e+09	367471	12829.396	6.564799e+08	0	

Рисунок 2.2 – Зріз датасету

Джерело: розроблено автором

Кожен рядок датасету представляє інформацію про зміну характеристик (ціна, обсяг та інші) активу в 4-х годинному проміжку. Щоб побачити відсоткові зміни ціни побудуємо графік її зміни для монети BTC у відсотках та дослідимо, на якому інтервалі відбуваються суттєві зміни (більше 1%). На рис. 2.3 по осі X зображено час у форматі (рік-місяць-день година-хвилина-секунда), а по осі Y – відсоткова зміна ціни відносно попереднього часового проміжку, де 0 – 0%, а 1 –

100%. Кожну відсоткову зміну ціни в певному часовому інтервалі виділено окремим кольором:

- синій – 4 години;
- помаранчевий – 8 годин;
- зелений – 12 годин;
- червоний – 16 годин;
- фіолетовий – 20 годин;
- коричневий – 24 години.

Як можна бачити з рис. 2.3, найбільша зміна відбувається у 16-24 годинних інтервалах (локальний максимум не враховується, оскільки такі зміни не є постійними, тоді як зміни $\pm 1\%$ є регулярними), що може вказувати на більшу значущість використання предикторів на зазначених часових проміжках для прогнозування моделі.

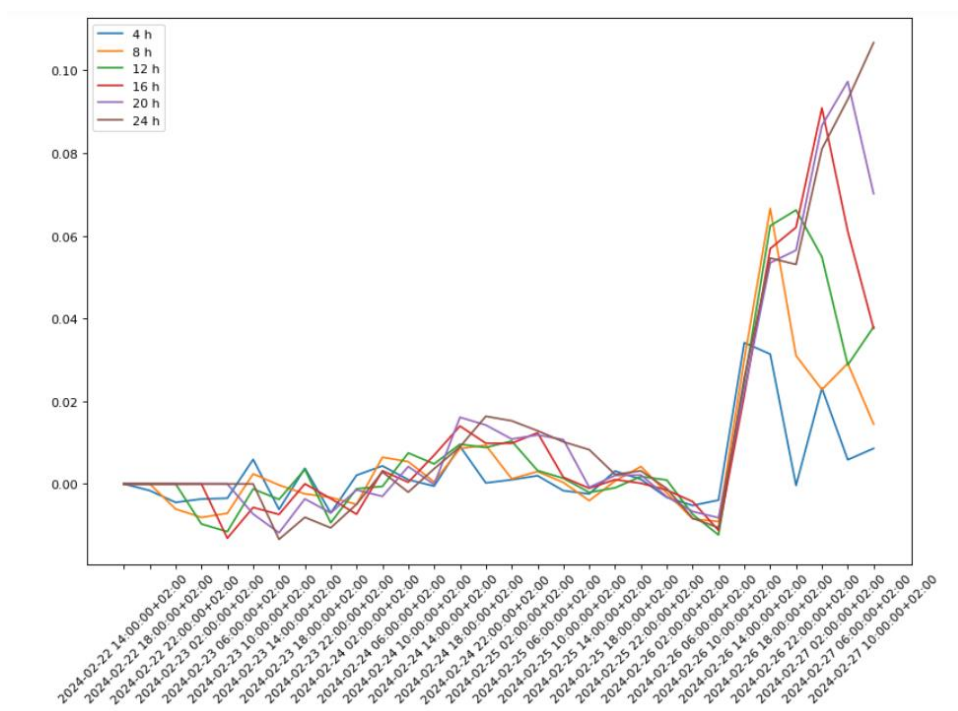


Рисунок 2.3 – Графік кривих відсоткової зміни ціни

Джерело: розроблено автором

Для більш наглядного аналізу 16-24 годинного інтервалу (див. рис. 2.4), де осі X та Y залишаються незмінні. Локальні екстремуми у вигляді 10% зміни ціни є значущими та відбуваються, як правило, внаслідок певних новин (політичних, світових, технічних, фінансових тощо) – такі речі важко передбачити, тому будемо вважати їх за викид і сфокусуємося на зміні тренду, навіть якщо це зміна ціни $\pm 1\%$, як, наприклад, на проміжку між 23 та 25 лютого 2024 року.

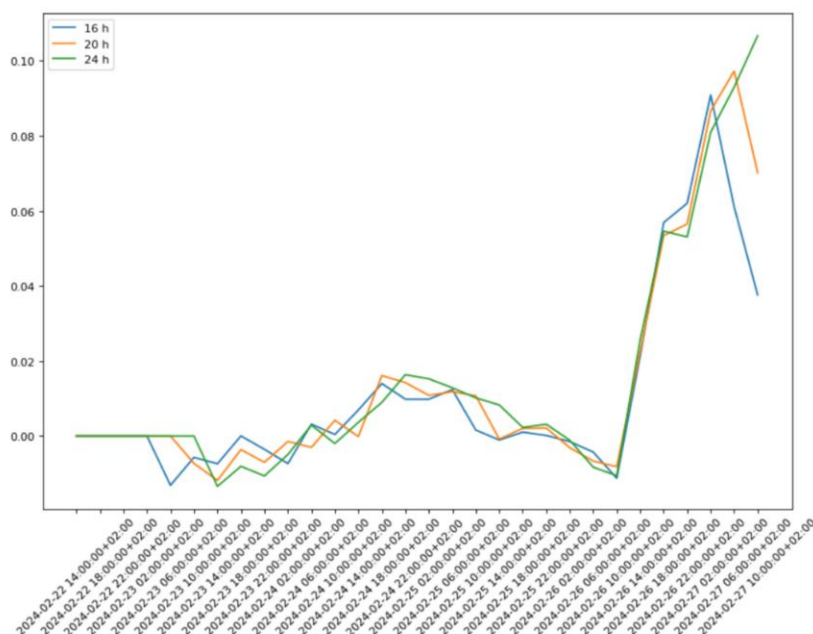


Рисунок 2.4 – Графік відсоткової зміни для 16-, 20- та 24-годинних проміжків

Джерело: розроблено автором

2.2 Формування предикторів та індикаторів

У досліджуваному датасеті представлені показники, які включають абсолютні значення, такі як ціна, обсяг і т.д, це все є значення, котрі є абсолютно різними від монети до монети, наприклад, ціна Bitcoin перевищує ціну інших монет в сотні й більше разів, тому необхідно зробити нормалізацію даних. Процес

нормалізації даних дозволить моделі, незалежно від різних факторів (тип монети, часовий проміжок), більш точно знаходити паттерн та здійснювати прогноз, а викиди стануть менш впливовими на результат, оскільки їх кількість на фоні нормалізованих даних стане меншою ніж до нормалізації, де кожні нові дані інтерпретувалися б як викиди.

За основу відносної змінної було обрано відсоткове співвідношення котирувань ціни за певний інтервал, що дає можливість нормалізувати дані в проміжку від 0 до 1, де 0 відповідає 0%, а 1 – 100%. Для більш наглядного аналізу було побудовано графік відсоткової зміни ціни закриття (close) для наступних проміжків: 4 – 24 години з кроком в 4 години. Нові колонки позначені як *pct_1*, *pct_2*, ..., *pct_6*, як зміна за 4, 8, ..., 24 години відповідно.

Для детального аналізу ринку криптовалют важливо використовувати індикатори, які дозволяють проводити більш об'єктивний аналіз динаміки ціни активу, порівняно з простим свічним графіком (рис. 2.5), що може бути підданим маніпуляціям з боку маркет-мейкерів та бірж.



Рисунок 2.5 – Свічний графік ціни Bitcoin

Джерело: розроблено автором

Було обрано індикатори *TSI* та *VPT*, котрі мають пряме відношення до напрямку тренду і визначають його на основі даних про ціну, обсяги, тощо.

Індекс справжньої міцності (*Trend Strength Index*, *TSI*) обчислюється за формулою:

$$TSI_i = \frac{PCDS_i}{APCDS_i} \cdot 100\%, \quad (2.1)$$

де PCDS (Price Close Double Smoothed) - подвійно згладжена ціну закриття;

APCDS (Absolute Price Close Double Smoothed) - абсолютні значення змін подвійно згладженої ціни закриття.

Обчислення зміни ціни (PC) – це різниця між поточною та попередньою ціною.

Згладжена зміна ціни (PCS) - на зміну ціни застосовується техніка згладжування, експоненційна ковзна середня (Exponential Moving Average, EMA), формула якої представлена нижче:

$$EMA_t = \alpha \cdot P_t + (1 - \alpha) \cdot EMA_{t-1}, \quad (2.2)$$

де α - коефіцієнт згладжування (де $0 < \alpha < 1$);

P_t – ціна на момент часу t ;

EMA_{t-1} - експоненційна ковзна середня на попередньому кроці часу $t-1$.

Подвійно згладжена зміна ціни (PCDS) - на PCS застосовується ще одна процедура згладжування за рахунок експоненційної ковзної середньої.

Абсолютна зміна ціни (APC) - абсолютне значення PC.

Згладжена абсолютна зміна ціни (APCS) - на APC застосовується згладжування експоненційної ковзної середньої.

Подвійно згладжена абсолютна зміна ціни (APCDSS) - на APCS застосовується ще одна процедура згладжування шляхом взяття експоненційної ковзної середньої.

Індикатор ефективно визначає зони перекупленості та перепроданості, і його широко використовують як для довгострокового інвестування, так і для короткострокової денної торгівлі з такими зонами інтересу, як $>0,75$ (75%) для перекупленості та $< 0,75$ (-75%) для перепроданості. Ідея цього індикатора полягає у визначенні напрямку руху ціни активу та інтенсивності цього руху. У випадку зростання ціни, значення індикатора збільшується, що свідчить про перевагу попиту над пропозицією, тоді як у випадку зниження ціни активу значення індикатора зменшується, що означає перевагу пропозиції над попитом [31].

Індикатор об'єму ціни та тренду (*Volume Price Trend Indicator, VPTI*) обчислюється за формулою:

$$VPT = \frac{\text{Previous VPT} + \text{Volume} \cdot (\text{Today's Closing Price} - \text{Previous Closing Price})}{\text{Previous Closing Price}}, \quad (2.3)$$

de Previous VPT – попереднє значення VPT;

Today's Closing Price - ціна закриття сьогоднішнього торгового часового проміжку (година, день, тиждень і тд);

Previous Closing Price - ціна закриття попереднього торгового часового проміжку.

VPT використовується для визначення балансу між попитом та пропозицією на ринку криптовалют.

Індикатор VPT є важливим інструментом, що дозволяє передбачати ринкові тенденції, прогнозувати розвороти тренду, підтверджувати тренди, ідентифікувати розриви (breakouts), а також прогнозувати можливі зміни на ринку [32]

Розбіжність конвергенції ковзного середнього (*Moving Average Convergence Divergence, MACD*) - технічний індикатор, що допомагає визначати цінові тенденції та потенційні торгові можливості [6]. Загальна стратегія полягає в пошуку сигналів на покупку, коли лінія MACD перетинає сигнальну лінію вгору, і сигналів на продаж, коли, відповідно, вниз.

Хоч даний показник широко використовуваний, але вже застарілий і не підходить для ринку криптовалют. MACD розроблений для довгострокових таймфреймів, таких як тиждень чи місяць, і тому він не є ефективним для ринків з високою волатильністю. На денних та тижневих графіках Bitcoin спостерігається висока волатильність та хвилеподібний характер, тому такий індикатор як MACD часто видає запізнілі сигнали, що створює ризики до втрат потенційно вигідних моментів для входу та виходу з торгівлі.

У цьому дослідженні Y включає напрямок тренду - зростаючий або спадаючий, що формується зі значення відсоткової зміни ціни на наступний часовий проміжок, котрий ми позначимо як зростаючий (1) або спадаючий (0).

Основним завданням є передбачення напрямку тренду, результативна змінна позначає напрямок та виражена через формулу (2.4).

$$Y = f(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (2.4)$$

де x – відсоткова зміни ціни відносно попереднього часового проміжку.

У рамках запропонованого методологічного підходу спочатку обчислимо значення TSI та VPTI для 4 годинних інтервалів для абсолютної ціни закриття інтервалу, потім побудуємо відсоткову зміну значення кожного з індикаторів для наступних проміжків: 4 – 24 години з кроком в 4 години та позначимо нові колонки як $tsi_1, tsi_2, \dots, tsi_6$ та $vpti_1, vpti_2, \dots, vpti_6$ як зміна за 4, 8, ..., 24 години відповідно. Відсоткові зміни індикаторів, наприклад, для 4 та 8 годин, видно на рис. 2.6. - зміна відбувається в діапазоні від -2000% до 500%, що могло вказувати на зміну тренду.

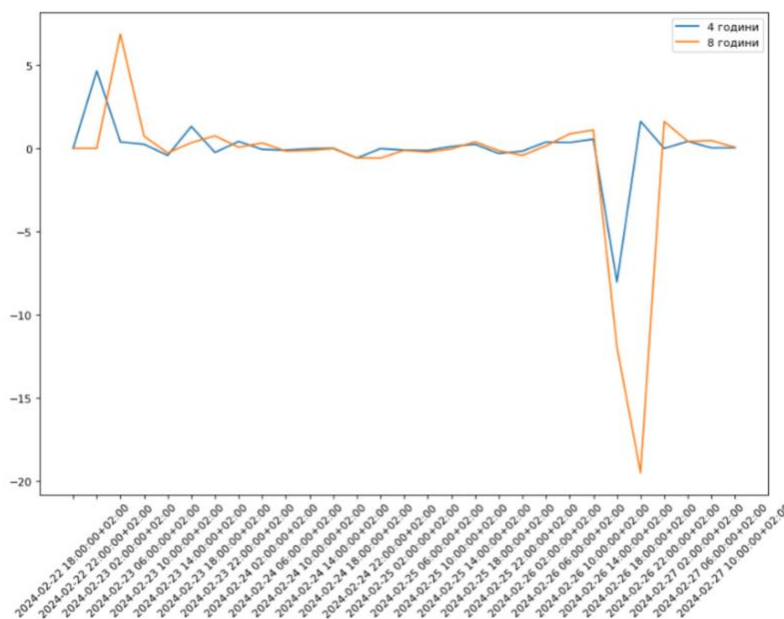


Рисунок 2.6 – Графік відсоткової зміни VPTI для 4 та 8 годин

Джерело: розроблено автором

Відобразимо в тому ж форматі зміни для індикатора TSI (див. рис. 2.7), де спостерігається різкий спад до рівня -200, що означає -20000% зміна порівняно з значенням на попередньому часовому проміжку. Як бачимо, індикатори можуть

показувати суттєві коливання, навіть при мінімальних цінових змінах, тому правильне їх використання може допомогти у прогнозуванні напрямку тренду.

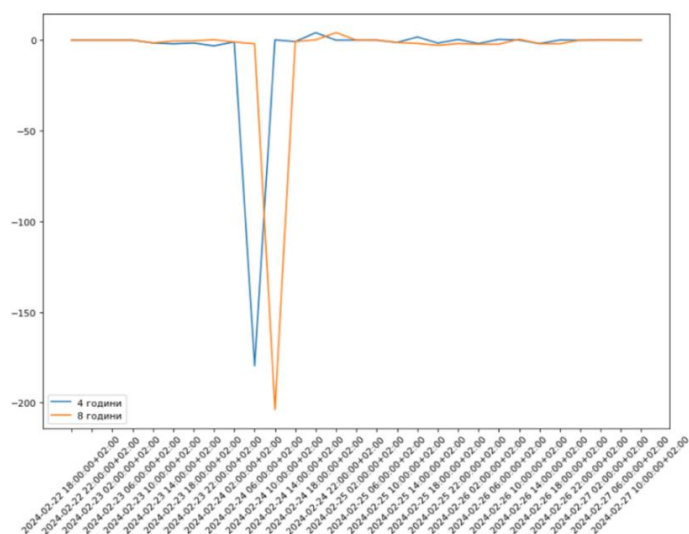


Рисунок 2.7 – Графік відсоткової зміни TSI для 4 та 8 годин

Джерело: розроблено автором

Дані для тренування моделей обираються в залежності від часового проміжку, тобто для 4-годинного часового інтервалу всі значення індикаторів та відсоткові зміни ціни мають бути відносно 4 годин. На прикладі графіків на рис. 2.6 та 2.7 видно, що для більшого часового проміжку зміни є значними, що цілком логічно, оскільки ціна має тенденцію до більших змін за більший часовий інтервал, що наводить на думку, що моделі мають показати себе краще для прогнозування тренду на більшому часовому проміжку.

Для перевірки моделей даного дослідження було вибрано декілька криптовалют, які відрізняються рівнем впливу на ринок.

Bitcoin – основна криптовалюта з найбільшою капіталізацією та найбільшим впливом на ринок криптовалюти. Ethereum – друга за капіталізацією та впливом монета, ZRX – монета на базі інфраструктурного протоколу 0x який дозволяє користувачам легко торгувати токенами ERC20 та іншими активами у блокчейні Ethereum. BOME (Book of Meme) – монета, яка нещодавно увійшла в лістинг та показала значний початковий зріст. Гіпотеза дослідження полягає в тому, що зміна ціни Bitcoin впливає абсолютно на всі монети, Ethereum впливає на

пов'язані з ним криптовалюти, наприклад, ZRX, а BOME – має вплив від загального ринку і дуже легко піддається зросту або падінню.

Всі дані активів було поділено на попередні тренувальні та тестові вибірки – 120 часових проміжків з інтервалом в 4 години для тренувальної та 60 для тестової.

Після нормалізації таких даних як відсоткова зміна ціни та значення індикаторів, тренувальний датасет стає готовим для навчання моделі.

В дослідженні аналізується декілька часових інтервалів для перевірки того, як моделі показують себе для різних монет на різних часових проміжках, тому для кожного з них буде здійснено прогноз на основі зібраних даних для подальшого порівняння ефективності моделей.

2.3 Аналіз критеріїв оцінки точності моделювання

Оскільки дане дослідження передбачає бінарну класифікацію - клас задач класифікації елементів набору даних, розбитих на дві групи на підставі правила класифікації (2.4), то для аналізу ефективності моделей було використано наступні критерії оцінки та допоміжні інструменти аналізу точності моделі:

- точність;
- ROC-AUC;
- матриця помилок;
- коефіцієнт Джині.

Ці інструменти дозволять оцінити не тільки загальний прогноз моделі, але й проаналізувати ефективність передбачення по кожному класу, визначаючи, яка з моделей буде краще передбачати зростаючий або спадаючий тренд.

Точність (precision) - це частка правильно класифікованих прикладів серед усіх прикладів. Обчислюється за формулою:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, \quad (2.5)$$

де TP – істинно позитивні результати, модель правильно передбачає 1;

TN – істинно негативні результати, модель правильно передбачає 0;

FP – хибні позитивні результати, модель передбачає 1, коли насправді має бути 0;

FN – хибні негативні результати, модель передбачає 0, коли насправді має бути 1.

ROC-AUC (Area Under the Receiver Operating Characteristic Curve), або площа під кривою, вимірює здатність моделі бінарної класифікації розрізняти між двома класами. Цей показник оцінюється шляхом обчислення площі під кривою, що створюється під час графічного представлення співвідношення між чутливістю (Recall) і специфічністю ($1 - \text{Specificity}$) при різних порогових значеннях [33].

Формально ROC-AUC можна розрахувати наступним чином:

1. ROC-крива визначається як графік, де по осі Y відкладається значення True Positive Rate (TPR, інакше - Recall або Sensitivity), а по осі X - False Positive Rate (FPR, $1 - \text{Specificity}$). Для кожного можливого порогу:

- $TPR(\text{Sensitivity}) = \frac{TP}{TP + FN}$.
- $FPR(\text{Sensitivity}) = \frac{FP}{FP + TN}$.

2. AUC (Area Under the Curve) вимірює площу під цією кривою. Чим ближче значення AUC до 1, тим краще класифікатор відрізняє між класами. Якщо $AUC = 0,5$, класифікатор не має жодної дискримінаційної здатності (еквівалентно випадковому вгадуванню).

Щоб обчислити AUC математично, часто використовують метод чисельного інтегрування або наприклад метод трапецій для апроксимації площі під кривою ROC.

Матриця помилок (Confusion Matrix) - це таблиця, яка візуалізує кількість TP , TN , FP , і FN при конкретному значенні точки відсікання. Це дає змогу детально оцінити, як модель впоралась з класифікацією для кожного класу.

Коефіцієнт Джині — вимірює здатність класифікатора розрізняти між двома класами. Значення «0» означає, що модель не має жодної дискримінаційної

здатності, значення близьке до «1» означає ідеальну дискримінацію, значення «-1» вказує на антидискримінаційну здатність, тобто модель систематично прогнозує неправильні результати.

Коефіцієнт Джині розраховується як подвоєна площа між кривою ROC і діагоналлю або як

$$Gini = 2 \cdot AUC - 1, \quad (2.6)$$

Коефіцієнт Джині завжди знаходиться в межах від 0 до 1, причому більше число означає кращий класифікатор.

Всі ці критерії дадуть ширший аналіз та розуміння вдалості моделі для прогнозування тренду котирувань криптовалют.

РОЗДІЛ 3

МОДЕЛЮВАННЯ ЗМІН ТРЕНДУ КОТИРУВАНЬ КРИПТОВАЛЮТ ЗА ДОПОМОГОЮ МЕТОДІВ МАШИННОГО НАВЧАННЯ ТА ШТУЧНОГО ІНТЕЛЕКТУ

3.1 Застосування логістичної регресії в задачі бінарної класифікації змін тренду котирувань криптовалют

Логістична регресія (*Logistic regression*) – це статистична модель, яка використовується для прогнозування ймовірності бінарного результату на основі незалежних змінних. Цей метод зазвичай використовується у машинному навчанні та аналізі даних для задач класифікацій. Використовуючи набір вхідних змінних, логістична регресія спрямована на моделювання ймовірності конкретного результату. Вихідна змінна в логістичній регресії є двійковою — вона може приймати лише одне з двох можливих значень (наприклад, «0» або «1», «так» або «ні»).

Логістична регресія використовує логістичну функцію, яка називається сигмоїдною. На рис. 3.1 зображена S-подібна крива, яка наближається до значень від 0 до 1, *threshold value* – це порогове значення. Якщо результат сигмоїдної функції перевищує значення 0,5, то результат визначається як 1, і навпаки, якщо дана функція нижча за значення 0,5, результат класифікується як 0. У даному дослідженні точка відтинання для визначення класу всюди буде 0,5 для проведення вибору найкращої моделі в однакових умовах.

Формула сигмоїдної функції записується таким чином:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3.1)$$

де e – основа натурального логарифму;

x – стандартне рівняння регресії.

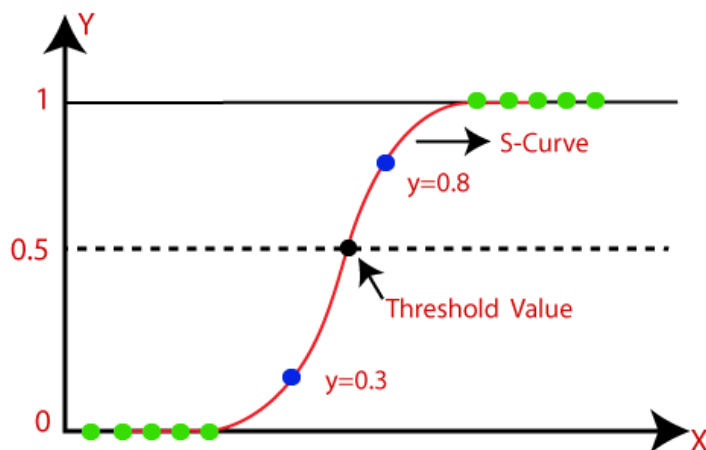


Рисунок 3.1 – Графік сигмоїдної функції

Джерело: [34]

У ході даного дослідження було побудовано логістичну регресію для вирішення задачі класифікації напрямку зміни тренду криптовалюти. Введемо наступні позначення (де «_train» – тренувальна вибірка, «_test» - тестова вибірка):

X_{train} , X_{test} – набір ознак, що складається з показників, які потенційно можуть впливати на ціну криптовалюти. В даному випадку – це технічні індикатори TSI та VPT ;

y_{train} , Y_{test} – вектор цільових значень, де приймаються значення «0» або «1».

Отримані результати тренування логістичної регресії показують, що точність моделі для Bitcoin коливається в діапазоні від 33,9% до 47,5% (див. табл. 3.1). Ті самі дії було проведено для інших досліджуваних криптовалют: ETH, ZRX та BOME, що показано у табл. 3.1.

Таблиця 3.1 – Порівняльна таблиця точності моделей логістичної регресії

Криптовалюта	Точність моделі для різних часових інтервалів					
	4h	8h	12h	16h	20h	24h
BTC	0,475	0,44	0,339	0,407	0,389	0,355

Продовження таблиці 3.1

ETH	0,566	0,591	0,641	0,633	0,66	0,725
ZRX	0,558	0,65	0,65	0,658	0,691	0,716
BOME	0,452	0,452	0,452	0,333	0,404	0,333

Джерело: розроблено автором

Модель найкраще передбачила дані з 4 годинним часовим проміжком, проте слід зазначити, що точність може бути вищою за моделі з більшим значенням інших критеріїв, тому необхідно орієнтуватися також на AUC та Джині, що демонструє табл. 3.2. Значення AUC варіюється від 0,465 до 0,513, коефіцієнт Джині для моделі з часовим проміжком у 16 годин є найвищим і становить 0,513, що вказує на найкращу дискримінативну здатність серед усіх 6 моделей (табл. 3.2).

Таблиця 3.2 – Порівняльна таблиця показників AUC та Джині моделей логістичної регресії

Показники	Криптовалюта - Bitcoin					
	4h	8h	12h	16h	20h	24h
AUC	0,465	0,5	0,5	0,513	0,5	0,5
Джині	-0,068	0	0	0,027	0,0	0

Джерело: розроблено автором

На основі наведених показників, модель для часового проміжку у 16 годин, показала найкращі результати за такими критеріями:

- AUC – 0,513;
- Джині – 0,027;
- Точність – 0,407

Найгірше модель справляється для 12 годинного проміжку з результатом точності 0,339.

Фундаментальним інструментом для оцінки моделей бінарної класифікації є ROC-крива, також відома як крива похибок, яка відображає зв'язок між істиною

позитивними результатами (TPR) з помилково позитивними результатами (FPR) [33]. На рис. 3.3 зображено побудовану ROC-криву для найкращої моделі логістичної регресії для Bitcoin.

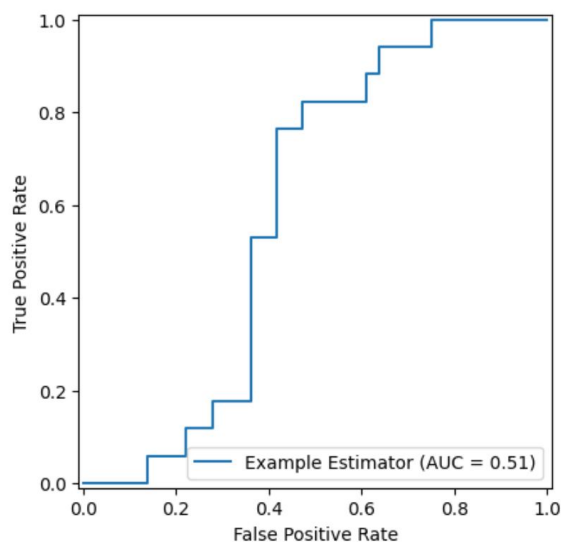


Рисунок 3.3 – ROC-крива для моделі з найвищим значенням точності для BTC

Джерело: розроблено автором

Ще одним інструментом для оцінки ефективності моделей класифікації є матриця помилок, результати якої зображено на рис. 3.4 для криптовалюти BTC.



Рисунок 3.4 – Матриця помилок найкращої моделі логістичної регресії для BTC

Джерело: розроблено автором

Матриця помилок показує, що модель змогла правильно класифікувати 2 випадки класу 0 і здійснила 35 неправильних передбачень. Для класу 1 було передбачено 0 правильних передбачень і 22 неправильних. Аналіз матриць помилок показав, що жодна модель логістичної регресії не змогла досягти успіху у класифікації класу 0 та продемонструвала високий рівень помилкових передбачень, що вказує на необхідність корекції порогу 0,5 у більшу сторону.

Для прогнозування цінових трендів криптовалют, логістична регресія може бути недостатньо точною та ефективною. Щоб підвищити точність прогнозів криптовалютних трендів, варто розглянути можливість використання більш складних моделей, які можуть краще враховувати складні взаємозв'язки у даних.

3.2 Застосування дерева рішень для класифікації трендів

Дерево рішень (*Decision tree*) — це ієрархічна модель підтримки прийняття рішень, яка використовує деревоподібну модель рішень та їхніх можливих наслідків, включаючи результати випадкових подій, витрати на ресурси та корисність [35]. Дерева рішень є потужним інструментом машинного навчання з багатьма перевагами, які роблять їх привабливими для використання в задачах класифікації та регресії. Цей метод особливо корисний в ситуаціях, коли дані є складними та нелінійними, та коли зв'язки між вхідними змінними та залежною (цільовою) змінною нелегко помітити.

Дерева рішень є потужним інструментом машинного навчання з багатьма перевагами, які роблять їх привабливими для використання в задачах класифікації та регресії.

Однією з головних переваг дерев рішень є їх інтуїтивність і простота розуміння. Завдяки своїй структурі дерева рішень дозволяють користувачам легко інтерпретувати результати та пояснювати моделі. Ще однією перевагою дерев

рішень є їх здатність працювати з мінімальною підготовкою даних. Цей метод не потребує нормалізації або стандартизації даних, обробки відсутніх значень чи усунення викидів, що значно скорочує витрати часу та зусиль.

Дерева рішень можуть обробляти як категоріальні, так і числові дані, а також працювати з нелінійними зв'язками між характеристиками і цільовими змінними, що робить їх універсальними інструментами для різних типів аналізу.

Дерева рішень — це моделі з високою дисперсією, і деякі зміни в даних можуть кардинально змінити прогнози, створені моделлю, ця проблема ускладнює використання даних, які є динамічними та можуть змінюватися протягом певного періоду часу [36].

На рис. 3.5 наведено схему побудови дерева рішень. Для того, щоб сформулювати уявлення щодо даного алгоритму машинного навчання, наведемо визначення певних термінів.

Кореневий вузол – найвищий вузол, місце, де починається дерево рішень.

Листовий вузол – кінцевий вихідний вузол, після якого неможливе розділення.

Розділення – процес поділу вузла на два або більше підвузлів.

Батьківський та дочірній вузли – вузол, від якого йде розділення на підвузли, називається батьківським, які в свою чергу називаються дочірніми.

Вузол прийняття рішень – вузол, що розбивається на додаткові вузли.

Гілка або піддерево – дерево, утворене розділенням дерева.

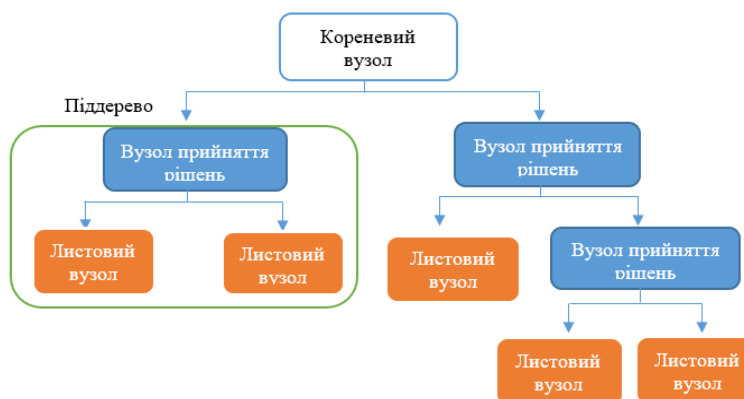


Рисунок 3.5 – Схема побудови дерева рішень

Джерело: розроблено автором

Процес створення дерева рішень є спадним. Алгоритм має віднайти критерій розщеплення, який зможе розбити дерево на гілки. Дуже важливо зрозуміти, який атрибут розмістити на різних рівнях дерева, адже випадковий підхід може дати неточні результати. Найвищий атрибут приросту інформації вибирається як критерій розщеплення.

Існує декілька критеріїв розщеплення. Найбільш відомими є ентропія та коефіцієнт Джині.

Критерій ентропії розраховується наступним чином:

$$E = - \sum_i^c p_i \log_2 p_i, \quad (3.2)$$

де p_i – ймовірність випадкового вибору елемента класу i , де елементом може бути будь-який об'єкт, що належить до цього класу.

Ентропія використовується для оцінки ступеня розподілу даних в наборі та визначає випадковість даних.

Для оцінки точності моделей дерева рішень на різних часових проміжках, було проведено порівняльний аналіз криптовалют BTC, ETH, ZRX і BOME. Результати наведено в таблиці 3.3.

Таблиця 3.3 – Порівняльна таблиця точності моделей дерева рішень

Криптовалюта	Точність моделі для різних часових інтервалів					
	4h	8h	12h	16h	20h	24h
BTC	0,448	0,614	0,767	0,727	0,796	0,754
ETH	0,457	0,338	0,338	0,305	0,305	0,322
ZRX	0,457	0,457	0,508	0,474	0,440	0,389
BOME	0,463	0,463	0,463	0,365	0,414	0,341

Джерело: розроблено автором

Найвища точність моделі для BTC спостерігається на 20-годинному інтервалі – 0,796 (див. табл. 3.3). Для ETH найвища точність буде на 4-годинному, а для ZRX та BOME – на 4-годинному та 8 годинному інтервалах.

Порівняльний аналіз також було здійснено для показників AUC та Gini для криптовалюти Bitcoin на різних часових інтервалах. Результати представлені в таблиці 3.4.

Таблиця 3.4 – Порівняльна таблиця показників AUC та Gini моделей дерева рішень

Показники	Криптовалюта - Bitcoin					
	4h	8h	12h	16h	20h	24h
AUC	0,448	0,609	0,77	0,754	0,819	0,795
Gini	-0,103	0,219	0,54	0,509	0,638	0,591

Джерело: розроблено автором

Значення AUC варіюється від 0,448 до 0,796, а коефіцієнт Джині для моделі з 20 годинним проміжком є найвищим і становить 0,638 з точністю 0,796, що вказує на найкращу дискримінативну здатність серед усіх 6 моделей (табл. 3.4).

Аналізуючи ROC-криву (див. рис. 3.6) для моделі $k=5$, можна спостерігати, що крива підіймається майже вертикально, що означає швидкість досягання моделі високого рівня чутливості з мінімальною кількістю хибнопозитивних результатів.

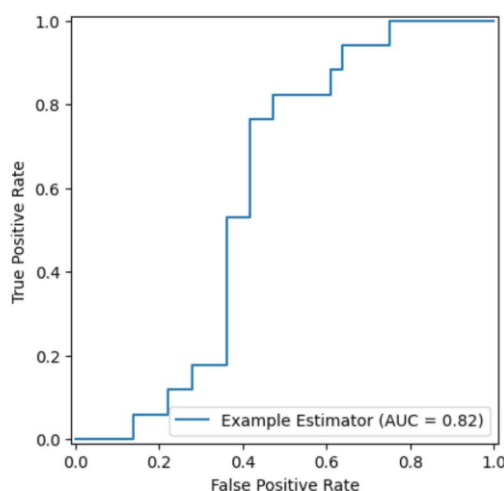


Рисунок 3.6 – ROC-крива для моделі дерева рішень з найвищим значенням точності для BTC

Джерело: розроблено автором

З урахуванням результатів матриць помилок, модель для 20-годинного проміжку (див. рис. 3.7) є найбільш ефективною для задачі передбачення змін тренду котирувань криптовалют. Вона демонструє найкращу здатність розрізняти класи та найвищий коефіцієнт Джині. Матриця помилок показує, що модель робить найменшу кількість похибок, особливо важливих для позитивного класу. В даному випадку мінімізація помилок є критичною, оскільки неправильна класифікація позитивного класу може призвести до втрати потенційних можливостей для інвестування.



Рисунок 3.7 – Матриця помилок найкращої моделі дерев рішень для BTC

Джерело: розроблено автором

Як видно з проведеного аналізу, модель дерева рішень, як і логістична регресія, не дає точних результатів для передбачення цін на криптовалюту на основі Bitcoin. Таким чином, найкраща модель дерева рішень може бути корисною для початкового етапу аналізу та прогнозування курсу криптовалют. Однак, для більш точних і надійних прогнозів, особливо в умовах високої волатильності та невизначеності, доцільніше використовувати більш складні моделі та методи машинного навчання.

3.3 Моделювання змін тренду криптовалют на основі нейронних мереж різних архітектур

Штучні нейронні мережі (ШНМ, Artificial Neural Network, ANN), які зазвичай просто називають нейронними мережами (НМ, Neural Networks, NN) або нейромережами (Neural Nets), це математичні, програмні та апаратні моделі, побудовані за принципом організації й функціонування біологічних нейронних мереж клітин живого організму [37]. Спрощене пояснення того, як це працює:

- Вхідний шар отримує початкові дані, які оброблятиме нейронна мережа, кожен нейрон у вхідному шарі відповідає за обробку певних вхідних даних.
- Приховані шари нейронів виконують обчислення вхідних даних, де кожен нейрон прихованого шару отримує дані від нейронів попереднього шару, застосовує математичну функцію активації і передає результат нейронам наступного шару.
- Вихідний рівень нейронів - останній шар нейронної мережі, який надає результат розрахунку нейронної мережі. Кількість нейронів у цьому шарі залежить від типу проблеми, яку вирішує нейронна мережа, наприклад, у задачі двійкової класифікації (де результатом є «так» або «ні») у вихідному шарі буде один нейрон.
- Зв'язки - кожен нейрон у шарі з'єднаний з кожним нейроном у сусідніх шарах, ці зв'язки пов'язані з вагами, які нейронна мережа коригує під час навчання, щоб оптимізувати свою ефективність.
- Функція активації - як згадувалося раніше, кожен нейрон застосовує функцію активації до зваженої суми своїх вхідних даних, ця функція вносить нелінійність у мережу, дозволяючи їй виявляти складні закономірності в даних.
- Тренування нейронної мережі - нейронні мережі навчаються на даних за допомогою процесу, який називається тренуванням нейромережі. Під час навчання мережі подається з вхідними даними правильні результати (мітки),

відбувається порівняння результатів нейромережі та міток, після чого за допомогою градієнтного спуску мінімізується функція втрат.

- Прогнозування - після навчання нейронна мережа може робити прогнози щодо нових даних, пропускаючи їх через мережу та отримуючи вихідні дані з вихідного рівня нейронів.

По суті, нейронна мережа навчається розпізнавати закономірності в даних, регулюючи свої внутрішні параметри (ваги) на основі прикладів, наданих під час навчання. Це дозволяє їй узагальнювати та робити прогнози щодо нових даних.

Існує багато видів нейромереж, але в даній роботі розглянуто лише основні. Персептрон — математична або комп'ютерна модель сприйняття інформації мозком (кібернетична модель мозку). Модель персептрона також розглядається як один з найкращих і найпростіших типів штучних нейронних мереж.

Він буває двох типів: одношаровий персептрон (див. рис. 3.8), який не містить прихованих шарів, та багатошаровий персептрон (див. рис. 3.9), який містить один або кілька прихованих шарів.

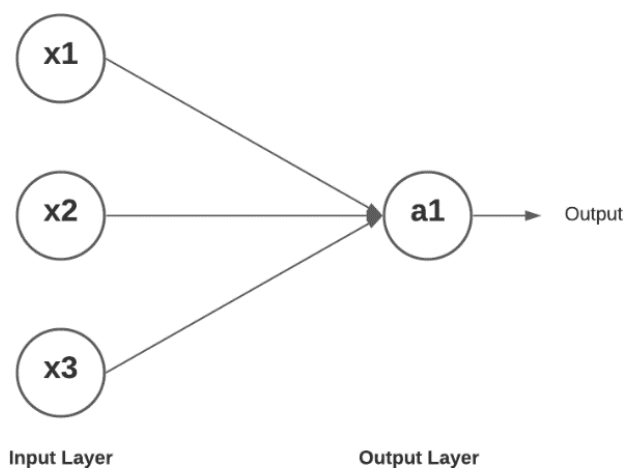


Рисунок 3.8 – Зображення одношарової нейронної мережі

Джерело: [38]

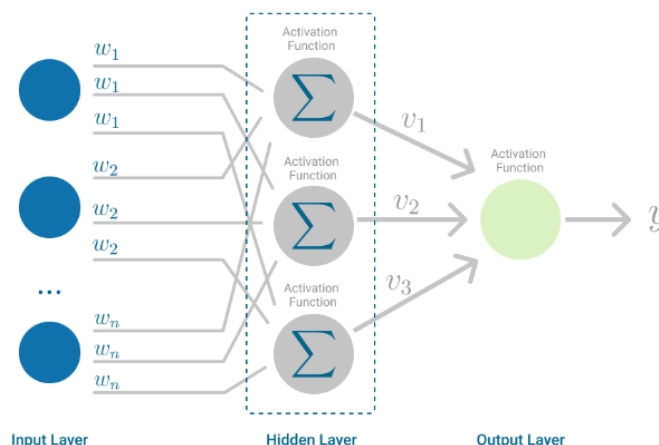


Рисунок 3.9 – Зображення багатошарової нейронної мережі

Джерело: [39]

Одним із головних недоліків використання одношарового персептрона є те, що він не може обробляти нелінійно розділені дані. Це означає, що при відсутності лінійної роздільності, персептрон не підходить для більш складних задач.

Мережі довгострокової короткочасної пам'яті (Long Short-Term Memory, LSTM) є типом рекурентної нейронної мережі RNN (див. рис. 3.10), призначеної для виявлення довгострокових залежностей у послідовних даних. На відміну від традиційних мереж прямого зв'язку, мережі LSTM мають комірки пам'яті та шлюзи, які дозволяють зберігати або забувати інформацію з часом вибірково. Це робить LSTM ефективними для розпізнавання мовлення, обробки природної мови, перекладу та аналізу часових рядів.

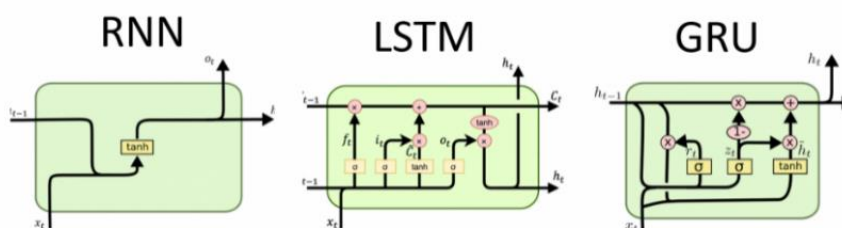


Рисунок 3.10 – Архітектура рекурентних нейронних мереж

Джерело: [40]

Мережі довгострокової короткочасної пам'яті мають широке застосування в обробці послідовних даних завдяки своїй здатності моделювати довгострокові

залежності. Завдяки можливості фіксувати довгострокові залежності в даних часових рядів, мережі довгострокової короткочасної пам'яті є придатними для таких завдань, як прогноз фондового ринку.

Згорткові нейронні мережі (CNN або ConvNets) — це спеціалізовані нейронні архітектури, які переважно використовуються для завдань комп'ютерного зору, таких як класифікація зображень і розпізнавання об'єктів (див. рис. 3.11). Ці нейронні мережі використовують потужність лінійної алгебри, зокрема через операції згортки, для ідентифікації шаблонів в зображеннях.

Convolutional Neural Network

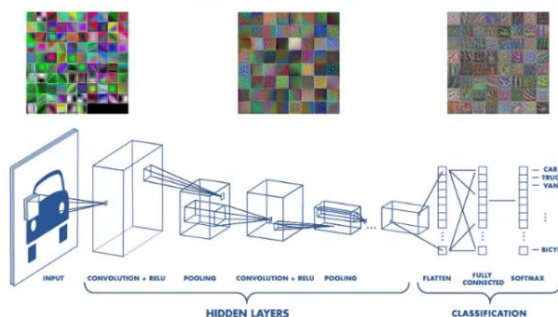


Рисунок 3.11 – Архітектура згорткової нейронної мережі для задач класифікації зображень

Джерело: [41]

Останні дослідження вказують, що використання CNN у задачах прогнозування часових рядів показує кращі результати в порівнянні з LSTM [42], тому ми побудуємо ряд нейромереж з функцією втрат *binary_crossentropy*, а саме: LSTM, CNN, CNN+LSTM та порівняємо їх. Функція *binary_crossentropy*:

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (3.4)$$

де y — фактичні мітки класів;

$\hat{y}$ - передбачені ймовірності приналежності до класу;

N — кількість спостережень у наборі даних.

Для побудови нейромереж був взятий фреймворк Keras. В даному випадку на рис. 3.12, 3.13 та 3.14 представлена інформація про:

- шари та їх тип (Layer (type));
- вихідну форму кожного шару (Output Shape);
- кількість параметрів у кожному шарі (Param#).

Архітектура кожної з нейромереж представлена нижче:

- LSTM – з 32 нейронами пам'яті, оптимізатором Adam, функцією втрат *binary_crossentropy* та активаційною функцією вихідного шару *sigmoid* (рис. 3.12).

На рис. 3.12 зображені такі дані як:

- Layer (type) - lstm_83 – це шар LSTM, dense_82 – це шар Dense.
- Output Shape - (None, 32) - шар 32 вихідні нейрони.
- Param # - шар lstm_83 має 4,352 параметри, шар dense_82 - 33 параметри.

Варто зауважити, що None вказує на змінну кількість зразків у партії (batch size).

Layer (type)	Output Shape	Param #
lstm_83 (LSTM)	(None, 32)	4,352
dense_82 (Dense)	(None, 1)	33

Рисунок 3.12 – Архітектура LSTM нейронної мережі

Джерело: розроблено автором

За аналогією, на рис. 3.13 та 3.14 також можна побачити ці дані для моделей CNN та CNN+LSTM відповідно.

- CNN – з кількістю фіч 32, оптимізатором Adam, функцією втрат *binary_crossentropy* та активаційною функцією вихідного шару *sigmoid* (рис. 3.13).

Layer (type)	Output Shape	Param #
conv1d_1 (Conv1D)	(None, 3, 32)	64
flatten_82 (Flatten)	(None, 96)	0
dense_88 (Dense)	(None, 1)	97

Total params: 161 (644.00 B)

Trainable params: 161 (644.00 B)

Рисунок 3.13 – Архітектура CNN нейронної мережі

Джерело: розроблено автором

- CNN+LSTM - з кількістю фіч 32 та з 32 нейронами пам'яті, оптимізатором Adam, функцією втрат *binary_crossentropy* та активаційною функцією вихідного шару *sigmoid* (рис. 3.14).

Layer (type)	Output Shape	Param #
conv1d_25 (Conv1D)	(None, 3, 32)	64
lstm_107 (LSTM)	(None, 32)	8,320
dense_112 (Dense)	(None, 1)	33

Total params: 8,417 (32.88 KB)

Trainable params: 8,417 (32.88 KB)

Рисунок 3.14 – Архітектура комбінованої CNN+LSTM нейронної мережі

Джерело: розроблено автором

Тренування відбувалося на тих самих вибірках, що і для інших моделей машинного навчання, протягом 1000 епох та розміром батчу 16 (розмір батчу (batch size) визначає кількість зразків, які використовуються за одну епоху для тренування нейронної мережі [43]).

Результати перевірки моделей на тестових даних зображені у табл. 3.5 та 3.6.

Таблиця 3.5 – Порівняльна таблиця показників AUC та Джині нейронних мереж

Показники	Криптовалюта - Bitcoin					
	4h	8h	12h	16h	20h	24h
Модель	LSTM					
AUC	0,517	0,456	0,435	0,233	0,305	0,527
Gini	0,034	-0,087	-0,12	-0,53	-0,38	0,055
Модель	CNN					
AUC	0,465	0,63	0,83	0,7	0,569	0,611
Gini	-0,068	0,261	0,659	0,4	0,138	0,222
Модель	CNN+LSTM					
AUC	0,458	0,516	0,698	0,675	0,614	0,591
Gini	-0,08	0,032	0,397	0,351	0,228	0,18

Джерело: розроблено автором

Таблиця 3.6 – Порівняльна таблиця точності нейронних мереж

Монета	Часовий проміжок					
	LSTM					
	4h	8h	12h	16h	20h	24h
BTC	0,517	0,385	0,428	0,428	0,240	0,471
ETH	0,457	0,322	0,474	0,237	0,237	0,389
ZRX	0,4576	0,542	0,711	0,610	0,593	0,576
BOME	0,463	0,536	0,317	0,414	0,487	0,512
CNN						
BTC	0,465	0,438	0,859	0,690	0,444	0,320
ETH	0,440	0,338	0,762	0,711	0,508	0,322
ZRX	0,491	0,525	0,711	0,711	0,627	0,508
BOME	0,487	0,560	0,585	0,634	0,658	0,609
CNN+LSTM						
BTC	0,5172	0,508	0,571	0,581	0,592	0,584
ETH	0,491	0,525	0,627	0,661	0,610	0,627
ZRX	0,474	0,593	0,677	0,677	0,711	0,677
BOME	0,487	0,512	0,634	0,682	0,780	0,707

Джерело: розроблено автором

Найкращий результат виявився для моделі CNN з інтервалом в 12 годин та $AUC = 0,83$. Графік ROC-кривої зображений на рис. 3.15. З матриці помилок на рис. 3.16 можна визначити, що модель гарно передбачає висхідний тренд. Ці результати вказують на те, що модель добре справляється з виявленням позитивних випадків, але все ж має проблеми з хибнопозитивними результатами.

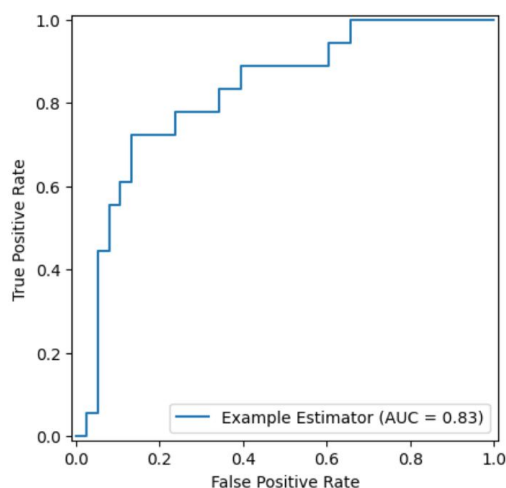


Рисунок 3.15 – ROC-крива для моделі нейронних мереж з найвищим значенням точності для BTC

Джерело: розроблено автором



Рисунок 3.16 – Матриця помилок для моделі нейронних мереж з найвищим значенням точності для BTC

Джерело: розроблено автором

Для аналізу процесу тренування нейронних мереж було побудовано графіки зміни значення функції втрат та точності з масштабом 1:100. На цих графіках видно, що функція втрат зменшується досить рівно та лінійно, що свідчить про те, що нейромережа може бути ще краще донавчена. Точність поступово збільшується, що вказує на відсутність перенавчання (overfitting) нейромережі.

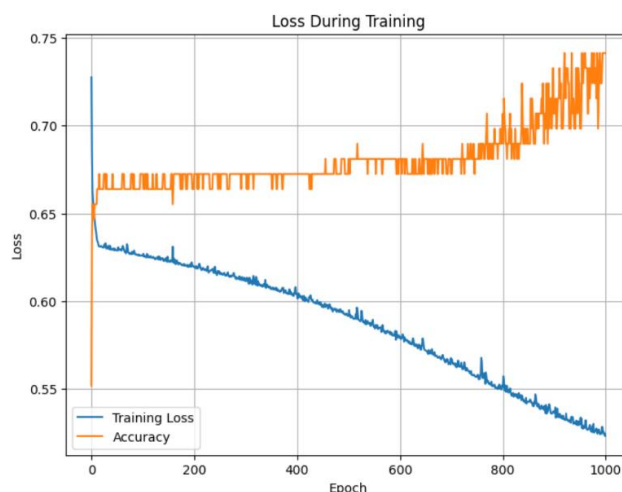


Рисунок 3.17 – Графік зміни функції втрат та точності під час тренування моделі CNN

Джерело: розроблено автором

Побудуємо графіки тренування для LSTM та CNN+LSTM та проаналізуємо їх. Нижче (див. рис. 3.18) графік тренування LSTM з найкращими результатами для 4h. Ми можемо бачити, що значення функції втрат після приблизно 500 епохи не спадає, а знаходиться в певному діапазоні, що може свідчити про те, що нейромережа досягла своїх можливостей через особливості архітектури та, відповідно, архітектуру треба змінювати.

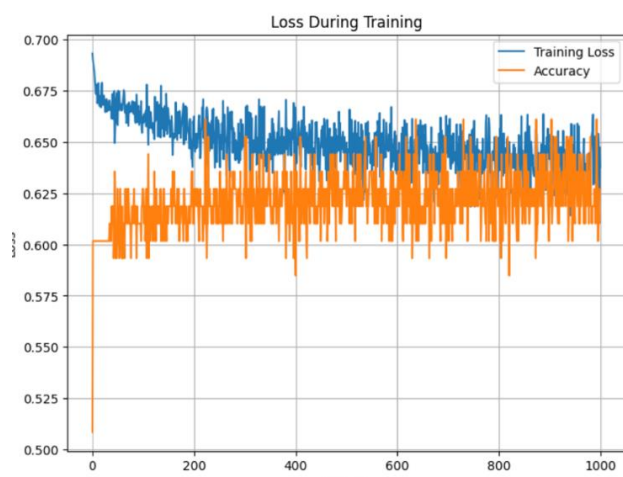


Рисунок 3.18 – Графік моделі LSTM

Джерело: розроблено автором

Наступний графік (див. рис. 3.19) - це графік тренування CNN+LSTM з найкращими результатами для 20h. Ми можемо бачити, що значення функції втрат

поступово спадає, але при цьому кожного разу коливається в певному діапазоні і чим більше епох, тим більший діапазон стає. Тож можна зробити висновок, що архітектура вибрана досить вдало, але має певні недоліки, котрі заважають покращувати результат.



Рисунок 3.19 – Графік моделі CNN+LSTM

Джерело: розроблено автором

Для того, щоб ефективно оцінити результативність побудованих моделей та обрати найкращий метод машинного навчання, нижче представлено табл. 3.7 з найкращими результатами для кожного з методів.

Таблиця 3.7 - Порівняння найкращих результатів моделей для прогнозування котирувань Bitcoin (BTC)

Модель	Інтервал (год)	Точність	AUC	Gini
Bitcoin (BTC)				
Логістична регресія	16	0,407	0,513	0,027
Дерево рішень	20	0,796	0,819	0,638
Нейронна мережа (CNN)	12	0,859	0,83	0,659

Джерело: розроблено автором

Найкращі результати серед досліджуваних моделей для прогнозування котирувань Bitcoin показали моделі дерева рішень та нейронної мережі (CNN). Дерево рішень має високу точність (0,796), що свідчить про його здатність

правильно передбачати більшість випадків. Нейронна мережа (CNN) демонструє найвищий результат точності (0,859) та найкращі показники AUC та Джині, що вказує на її високу здатність розрізняти класи. Це є важливим для точного прогнозування зростання і падіння котирувань, а також допомагає знизити фінансові ризики для інвесторів та трейдерів.

З огляду на проведений аналіз, можна зробити висновок, що найкращою моделлю прогнозування котирувань криптовалют є нейронна мережа типу CNN. Висока точність моделі CNN та її високі показники AUC та Джині роблять цю модель ефективною для задач передбачень змін тренду криптовалют, забезпечуючи точність та надійність результатів.

ВИСНОВКИ

В даній кваліфікаційній роботі були розроблені моделі машинного навчання для прогнозування трендів котирувань криптовалют. Вони базуються на сучасних алгоритмах та демонструють високу точність у передбаченні змін тренду криптовалют. Проведене дослідження дозволило зробити наступні висновки:

1. Проведений аналіз сучасного стану ринку криптовалют показав значне зростання цього цифрового активу. Це, в свою чергу, підкреслює необхідність розробки надійних та ефективних методів прогнозування курсу криптовалют, що дозволить трейдерам та інвесторам приймати швидкі та обґрунтовані рішення, а також передбачати майбутню цінову динаміку.

2. Традиційні методи, такі як технічний та фундаментальний аналіз, надають важливі інструменти для дослідження ринку, що багато років використовуються трейдерами та аналітиками. Для технічного аналізу характерне застосування таких інструментів як ковзні середні, індикатори тренду, осцилятори, технічні індикатори TSI (True Strength Index) та VPT (Volume Price Trend), а також свічкові графіки. В свою чергу, фундаментальний аналіз передбачає оцінку внутрішньої вартості активу на основі макроекономічних показників, новин, світових подій та інших факторів, що впливають на ринок. Однак, такі традиційні підходи мають обмежену здатність враховувати складні взаємозв'язки та патерни в даних.

3. Поруч з такими методами машинного навчання як логістична регресія та дерева рішень, свою ефективність у прогнозуванні часових рядів криптовалют демонструють нейронні мережі, які набувають все більшої популярності, забезпечуючи більш точні прогнози. Розробка методів прогнозування, які забезпечують високу точність може позитивно вплинути на ефективність роботи трейдерів та аналітиків, зменшуючи час на обробку даних та прийняття рішень, що особливо важливо в швидкоплинних умовах ринку криптоактивів.

4. Розробка концептуального підходу до прогнозування змін тренду криптовалют на основі машинного навчання включала визначення основних характеристик та показників, які використовуються для прогнозування. Було розроблено алгоритм підготовки даних, що включає нормалізацію. Крім того, була розроблена модель для бінарної класифікації напрямків змін криптовалюти.

5. Розроблені моделі на основі логістичної регресії, дерев рішень, LSTM, CNN та LSTM+CNN продемонстрували здатність прогнозувати зміни трендів криптовалют на різних часових проміжках та для різних криптовалют по-різному, зокрема нейронні мережі та дерева рішень показали високі значення точності та показники Джині та AUC. Найкращий результат продемонструвала модель CNN в порівнянні з іншими підходами. Це вказує на практичну цінність цих моделей для інвесторів та трейдерів ринку криптовалют.

Отже, дане дослідження підтвердило доцільність використання сучасних методів машинного навчання, таких як нейронні мережі, для прогнозування трендів криптовалют. Побудовані моделі можуть бути успішно застосовані для покращення точності прогнозів, зменшення ризиків та оптимізації прийняття інвестиційних рішень на ринку криптовалют.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Гонак І. М., Горин В. П. Криптовалюта як об'єкт інвестування. Наукові записки Національного університету «Острозька академія». Серія «Економіка»: науковий журнал. Острог: Видавництво НаУОА, вересень 2022. № 26(54). С. 71-84.
2. Чумаченко О.Г., Баластрик О. Г. Передумови виникнення криптовалюти. *Наукові праці НДФІ*, 2017. № 4. С. 96-99.
3. What Is Cryptocurrency? *Crypto.com*. URL: <https://crypto.com/university/what-is-cryptocurrency>. (дата звернення: 16.04.2024).
4. Панахов Ф. Сутність і функції криптовалют. *Економічний вісник Донбасу*, (4 (70), 119–120. URL: [https://doi.org/10.12958/1817-3772-2022-4\(70\)-119-120](https://doi.org/10.12958/1817-3772-2022-4(70)-119-120).
5. Edwards J. Bitcoin's Price History. *Investopedia*. URL: <https://www.investopedia.com/articles/forex/121815/bitcoins-price-history.asp>.
6. Bitcoin's price history: 2009-2024. *Forbes India*. URL: <https://www.forbesindia.com/article/explainers/bitcoin-price-history/92523/1> (дата звернення: 20.04.2024).
7. Tretina K. Top 10 cryptocurrencies of may 2024. *Forbes*. URL: <https://www.forbes.com/advisor/au/investing/cryptocurrency/top-10-cryptocurrencies/> (дата звернення: 25.04.2024).
8. Блокчейн технології. *Avitar*. URL: <https://www.avitar.legal/industries/blokchein-tekhnologii> (дата звернення: 26.04.2024).
9. Number of daily network transactions on the blockchain of Bitcoin from January 2009 to April 11, 2024. *Statista*. URL: <https://www.statista.com/statistics/730806/daily-number-of-bitcoin-transactions/> (дата звернення: 26.04.2024).

10. 5,300 new tokens launched daily in 2024 so far. *CoinGecko*. URL: <https://www.coingecko.com/research/publications/how-many-cryptocurrencies-are-there> (дата звернення: 30.04.2024).
11. Іванов М. П., Гострик О. М. Аналіз ринку криптовалют і його особливості в сучасних умовах. *Інформаційні технології в економіці і управлінні*: зб. наук. студ. праць. Одеса : ОНЕУ, 2020. Вип. 2. С. 16.
12. Lu M. Countries with the highest rates of crypto ownership. *Visual Capitalist*. URL: https://www.visualcapitalist.com/countries-with-the-highest-rates-of-crypto-ownership/#google_vignette (дата звернення: 11.05.2024).
13. Лук'янчук Р. В. Сучасні тенденції розвитку вітчизняного сектору в умовах правового режиму воєнного стану. *Інформація і право*. 2023. № 2. С. 113–124. URL: [https://doi.org/10.37750/2616-6798.2023.2\(45\).282329](https://doi.org/10.37750/2616-6798.2023.2(45).282329).
14. Qaroush Z., Zakarneh S., Dawabsheh A. Cryptocurrencies advantages and disadvantages: a review. *International journal of applied sciences and smart technologies*. 2022. Vol. 4, no. 1. P. 1–20. DOI:10.24071/ijasst.v4i1.4610
15. Дербенцев В., Великоіваненко Г., Даценко Н. Застосування методів машинного навчання до прогнозування часових рядів криптовалют. *Нейро-нечіткі технології моделювання в економіці*, 2019. № 8. С. 65-93. <http://doi.org/10.33111/nfmte.2019.065> (дата звернення: 11.05.2024).
16. Середняк Т. К. Сучасні гроші, банківські послуги та фінансові інновації в умовах інтеграції України в ЄС: матеріали V всеук. наук.-практ. інтерн. конф. студ. аспір., молод. вчених і пров. фах-ів. Київ, КНЕУ — Дніпро: 2023, 15 с.
17. G. Dudek et al. Forecasting cryptocurrencies volatility using statistical and machine learning methods: a comparative study. *Applied soft computing*. 2024. Vol. 151. URL: <https://doi.org/10.1016/j.asoc.2023.111132>.
18. Даценко Н. В. Застосування дерев класифікації та регресії до прогнозування часових рядів фінансових інструментів. *Вчені записки*: зб. наук. пр. М-во освіти і науки України, ДВНЗ «Київ. нац. екон. ун-т ім. Вадима

Гетьмана». Київ : КНЕУ, 2018. Вип. 19. С. 195–206. DOI.10.33111/vz_kneu.19.18.01.18.124.130.

19. Kanat E. The validity of technical analysis in the cryptocurrency market: evidence from machine learning methods. *Journal of business, economics and finance (JBEF)*. 2023. Vol. 12, no. 3. P. 102–109. DOI: 10.17261/Pressacademia.2023.1821.

20. Сохацька О.М., Панасюк В.М., Роговська-Іщук І.В., Вінницький С.І. Фундаментальний та технічний аналізи міжнародних ринків. Тернопіль: ЗУНУ, 2022. 309 с.

21. Вечерковська А., Поперешняк С. Огляд алгоритмів машинного навчання та їх застосування для прогнозування цін купівлі криптовалюти. *Вісник Херсонського національного технічного університету*. 2023. № 4(87). С. 223–229. URL: <https://doi.org/10.35546/kntu2078-4481.2023.4.26>.

22. Kontopoulou V. I. A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. *Future internet*. 2023. Vol. 15, 8. P. 255. URL: <https://doi.org/10.3390/fi15080255> (дата звернення: 17.05.2024).

23. Latif N. Comparative performance of LSTM and ARIMA for the short-term prediction of bitcoin prices. *Australasian accounting, business and finance journal*. 2023. Vol. 17, no. 1. P. 256–276. URL: <https://doi.org/10.14453/aabfj.v17i1.15> (дата звернення: 17.05.2024).

24. Yusifov M., Bakirova L. Machine learning concepts and applications. Azerbaijan State Oil and Industry University. 2021.

25. Pakeeza H. Concepts of Machine Learning. URL: https://www.researchgate.net/publication/379477181_Concepts_of_Machine_Learning#pf2 (дата звернення: 18.05.2024).

26. Ranjan R. Effectiveness of machine learning algorithms in cryptocurrency price forecasting - phemex. *Phemex*. URL: <https://phemex.com/student-foundation/papers/machine-learning-crypto> (дата звернення: 18.05.2024).

27. Taye M. Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. *Academic journal*. 2023. Vol. 12, no. 5. P. 19. DOI: 10.3390/computers12050091.

28. Sarker I. H. Machine learning: algorithms, real-world applications and research directions. *SN computer science*. 2021. Vol. 2, no. 3. URL: <https://doi.org/10.1007/s42979-021-00592-x> (дата звернення: 18.05.2024).
29. Barreto C. Entendendo mais sobre o ciclo de vida de projetos de machine learning, 2017. *Medium*. URL: <https://medium.com/@carleto/crisp-dm-entendendo-mais-sobre-o-ciclo-de-vida-de-projetos-de-machine-learning-b2ffc44cd3b7> (дата звернення: 18.05.2024).
30. Японські свічки: вчимося читати і розуміти рух ринку / *ITstatti.in.ua*. URL: https://itstatti.in.ua/15-treidinh/906-yaponski-svichky.html#google_vignette (дата звернення: 20.05.2024).
31. What is true strength index (tsi): crypto divergence & trend trading. *Phemex*. URL: <https://phemex.com/academy/what-is-true-strength-index> (дата звернення: 17.05.2024).
32. Volume price trend indicator (vpt). Corporate finance institute. URL: <https://corporatefinanceinstitute.com/resources/career-map/sell-side/capital-markets/volume-price-trend-indicator-vpt/> (дата звернення: 20.05.2024).
33. Dash S. Understanding the roc and auc intuitively. *Medium*. URL: <https://medium.com/@shaileydash/understanding-the-roc-and-auc-intuitively-31ca96445c02> (дата звернення: 20.05.2024).
34. Ansari Y. Understanding logistic regression: a beginner's guide, 2023. *Medium*. URL: https://medium.com/@novus_afk/understanding-logistic-regression-a-beginners-guide-73f148866910 (дата звернення: 20.05.2024).
35. Logistic regression: definition, use cases, implementation. *V7labs*. URL: <https://www.v7labs.com/blog/logistic-regression> (дата звернення: 21.05.2024).
36. Даценко Н. В. Система моделей оцінювання та прогнозування інноваційних фінансових інструментів (на прикладі криптовалют): дис. ... на здобуття канд. екон. наук: 08.00.11. Київ, 2019. 173 с. (дата звернення: 21.05.2024).
37. Карпа Д. М., Тсмотс І. Н., Оротіак Ю. В. Нейромережеві засоби прогнозування споживання енергоресурсів. *Scientific Bulletin of UNFU*. 2018. Т.

28, № 5. С. 140–146. URL: <https://doi.org/10.15421/40280529> (дата звернення: 27.05.2024).

38. Manav. Decision trees: explained in simple steps. *Medium*. URL: <https://medium.com/analytics-vidhya/decision-trees-explained-in-simple-steps-39ee1a6b00a2> (дата звернення: 21.05.2024).

39. Ravi S. Introduction to neural networks: build a single layer perceptron in pytorch. *Medium*. URL: <https://medium.com/@shashankshankar10/introduction-to-neural-networks-build-a-single-layer-perceptron-in-pytorch-c22d9b412ccf> (дата звернення: 21.05.2024).

40. Ali R., Hussain J., Lee S. W. Multilayer perceptron-based self-care early prediction of children with disabilities. *Digital health*. 2023. Т. 9. URL: <https://doi.org/10.1177/20552076231184054> (дата звернення: 21.05.2024).

41. Saidi R., Jarray F., Alsuhaiban M. Comparative analysis of recurrent neural network architectures for arabic word sense disambiguation. *Conference: 18th International Conference on Web Information Systems and Technologies*. 2022. DOI:10.5220/0011527600003318

42. Vaid N. K. Convolution neural networks. *Medium*. URL: <https://neeraj-kumar-vaid.medium.com/convolution-neural-networks-9516833e118d> (дата звернення: 20.05.2024).

43. Wang X. Automated recognition of epilepsy from EEG signals using a combining space–time algorithm of CNN-LSTM. *Scientific reports*. 2023. Т. 13, № 1. URL: <https://doi.org/10.1038/s41598-023-41537-z.4>

44. Relation between learning rate and batch size. *Baeldung on computer science*. URL: <https://www.baeldung.com/cs/learning-rate-batch-size> (дата звернення: 20.05.2024).

ДОДАТКИ

Додаток А

Код програми для підготовки та аналізу тестових даних

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.pyplot import figure
BTC = pd.read_csv('./btc.csv', sep='\t')
ETH = pd.read_csv('./eth.csv', sep='\t')
ZRX = pd.read_csv('./zrx.csv', sep='\t')
BOME = pd.read_csv('./BOME.csv', sep='\t')
BTC.set_index('open_time', inplace=True)
ETH.set_index('open_time', inplace=True)
ZRX.set_index('open_time', inplace=True)
BOME.set_index('open_time', inplace=True)
#показуємо дані
BTC.tail()
ETH.head()
print(f'Len BTC data: {len(BTC)} splitted by 4hour, total days: {len(BTC)*4/24}')
print(f'Len ETH data: {len(ETH)} splitted by 4hour, total days: {len(ETH)*4/24}')
#Візуалізація змін ціни BTC за перші 5 днів
figure(figsize=(12, 8), dpi=80)
plt.xticks(rotation=45)
plt.plot(BTC['close'][:30])
# формування нових предикторів
for k in range(1,28//4,1):
    BTC[f'close_pct_{k}'] = BTC['close'].pct_change(periods=k)
    ETH[f'close_pct_{k}'] = ETH['close'].pct_change(periods=k)
    ZRX[f'close_pct_{k}'] = ZRX['close'].pct_change(periods=k)
    BOME[f'close_pct_{k}'] = BOME['close'].pct_change(periods=k)
BTC.fillna(0, inplace = True)
ETH.fillna(0, inplace = True)
ZRX.fillna(0, inplace = True)
BOME.fillna(0, inplace = True)
```

```

for k in range(1,28//4,1):#1 - 4h, 2 - 8h, 3 - 12h
    BTC[f'close_pct_{k}_logistic'] = BTC[f'close_pct_{k}']>0
    ETH[f'close_pct_{k}_logistic'] = ETH[f'close_pct_{k}']>0
    BTC[f'close_pct_{k}_logistic'] = BTC[f'close_pct_{k}_logistic'].astype(int)
    ETH[f'close_pct_{k}_logistic'] = ETH[f'close_pct_{k}_logistic'].astype(int)
    ZRX[f'close_pct_{k}_logistic'] = ZRX[f'close_pct_{k}']>0
    BOME[f'close_pct_{k}_logistic'] = BOME[f'close_pct_{k}']>0
    ZRX[f'close_pct_{k}_logistic'] = ZRX[f'close_pct_{k}_logistic'].astype(int)
    BOME[f'close_pct_{k}_logistic'] = BOME[f'close_pct_{k}_logistic'].astype(int)
# відобразимо графік зміни ціни в проміжках часу для аналізу
figure(figsize=(12, 8), dpi=80)
for k in range(1,28//4, 1):
    plt.plot(BTC[f'close_pct_{k}'][:30], )
plt.legend([f'{k*4} h' for k in range(1,28//4, 1)])
plt.xticks(rotation=45)
plt.show()
figure(figsize=(12, 8), dpi=80)
for k in range(4,28//4, 1):
    plt.plot(BTC[f'close_pct_{k}'][:30], )
plt.legend([f'{k*4} h' for k in range(4,28//4, 1)])
plt.xticks(rotation=45)
plt.show()
from ta.volume import VolumePriceTrendIndicator
from ta.momentum import tsi, TSIIndicator
BTC_train, BTC_test = BTC[:120].copy(), BTC[120:].copy()
ETH_train, ETH_test = ETH[:120].copy(), ETH[120:].copy()
ZRX_train, ZRX_test = ZRX[:120].copy(), ZRX[120:].copy()
BOME_train, BOME_test = BOME[:120].copy(), BOME[120:].copy()
BTC_train['vpti'] = VolumePriceTrendIndicator(BTC_train['close'], BTC_train['volume'],
fillna=True).volume_price_trend()
BTC_test['vpti'] = VolumePriceTrendIndicator(BTC_test['close'], BTC_test['volume'],
fillna=True).volume_price_trend()
ETH_train['vpti'] = VolumePriceTrendIndicator(ETH_train['close'], ETH_train['volume'],
fillna=True).volume_price_trend()

```

```

ETH_test['vpti'] = VolumePriceTrendIndicator(ETH_test['close'], ETH_test['volume'],
fillna=True).volume_price_trend()
ZRX_train['vpti'] = VolumePriceTrendIndicator(ZRX_train['close'], ZRX_train['volume'],
fillna=True).volume_price_trend()
ZRX_test['vpti'] = VolumePriceTrendIndicator(ZRX_test['close'], ZRX_test['volume'],
fillna=True).volume_price_trend()
BOME_train['vpti'] = VolumePriceTrendIndicator(BOME_train['close'], BOME_train['volume'],
fillna=True).volume_price_trend()
BOME_test['vpti'] = VolumePriceTrendIndicator(BOME_test['close'], BOME_test['volume'],
fillna=True).volume_price_trend()
for k in range(1,28//4, 1):
    BTC_train[f'tsi_{k}'] = TSIIndicator(BTC_train['close'], window_slow = 2,
window_fast=1).tsi().pct_change( periods=k).fillna(0)
    ETH_train[f'tsi_{k}'] = TSIIndicator(ETH_train['close'], window_slow = 2,
window_fast=1).tsi().pct_change( periods=k).fillna(0)
    ZRX_train[f'tsi_{k}'] = TSIIndicator(ZRX_train['close'], window_slow = 2,
window_fast=1).tsi().pct_change( periods=k).fillna(0)
    BOME_train[f'tsi_{k}'] = TSIIndicator(BOME_train['close'], window_slow = 2,
window_fast=1).tsi().pct_change( periods=k).fillna(0)
    BTC_train[f'vpti_{k}'] = VolumePriceTrendIndicator(BTC_train['close'],
BTC_train['volume']).volume_price_trend().pct_change( periods=k).fillna(0)
    ETH_train[f'vpti_{k}'] = VolumePriceTrendIndicator(ETH_train['close'],
ETH_train['volume']).volume_price_trend().pct_change( periods=k).fillna(0)
    ZRX_train[f'vpti_{k}'] = VolumePriceTrendIndicator(ZRX_train['close'],
ZRX_train['volume']).volume_price_trend().pct_change( periods=k).fillna(0)
    BOME_train[f'vpti_{k}'] = VolumePriceTrendIndicator(BOME_train['close'],
BOME_train['volume']).volume_price_trend().pct_change( periods=k).fillna(0)
    BTC_test[f'tsi_{k}'] = TSIIndicator(BTC_test['close'], window_slow = 2,
window_fast=1).tsi().pct_change( periods=k).fillna(0)
    ETH_test[f'tsi_{k}'] = TSIIndicator(ETH_test['close'], window_slow = 2,
window_fast=1).tsi().pct_change( periods=k).fillna(0)
    ZRX_test[f'tsi_{k}'] = TSIIndicator(ZRX_test['close'], window_slow = 2,
window_fast=1).tsi().pct_change( periods=k).fillna(0)
    BOME_test[f'tsi_{k}'] = TSIIndicator(BOME_test['close'], window_slow = 2,
window_fast=1).tsi().pct_change( periods=k).fillna(0)

```

```

BTC_test[f'vpti_{k}'] = VolumePriceTrendIndicator(BTC_test['close'],
BTC_test['volume']).volume_price_trend().pct_change( periods=k).fillna(0)
ETH_test[f'vpti_{k}'] = VolumePriceTrendIndicator(ETH_test['close'],
ETH_test['volume']).volume_price_trend().pct_change( periods=k).fillna(0)
ZRX_test[f'vpti_{k}'] = VolumePriceTrendIndicator(ZRX_test['close'],
ZRX_test['volume']).volume_price_trend().pct_change( periods=k).fillna(0)
BOME_test[f'vpti_{k}'] = VolumePriceTrendIndicator(BOME_test['close'],
BOME_test['volume']).volume_price_trend().pct_change( periods=k).fillna(0)
# Візуалізуємо індикатор TSI для проміжку 4год*3 = 12 год, тобто підрахунок вівся для
кожних 12 год
figure(figsize=(12, 8), dpi=80)
plt.xticks(rotation=45)
# plt.plot(ZRX_test['tsi_3'][:30])
# plt.plot(ZRX_test['tsi_4'][:30])
plt.plot(BTC_train['tsi_1'][1:30], label='4 години')
plt.plot(BTC_train['tsi_2'][1:30], label='8 години')
# plt.plot(BTC_train['tsi_1'][20:50], label='4 години')
# plt.plot(BTC_train['close'].pct_change()[20:50], label='ціна')
# plt.plot(BTC_train['tsi_2'][20:50], label='8 годин')
# plt.plot(ZRX_test['tsi_3'][20:50],label='24 години')
# Візуалізуємо індикатор VPTI
figure(figsize=(12, 8), dpi=80)
plt.xticks(rotation=45)
plt.plot(BOME_train['vpti'][0:30])
plt.legend()
#Побудова логістичної регресії
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, roc_curve, RocCurveDisplay, auc, confusion_matrix
import seaborn as sns
results = {}
for k in range(1,28//4, 1):
    X_train = BTC_train[[f'tsi_{k}', f'vpti_{k}',f'close_pct_{k}']].values[k:-1]
    y_train = BTC_train[f'close_pct_{k}_logistic'].values[k+1:]

```

```

X_test = BTC_test[['tsi_{k}', 'vpti_{k}', 'close_pct_{k}']].values[k:-1]
y_test = BTC_test['close_pct_{k}_logistic'].values[k+1:]
# Створимо екземпляр моделі логістичної регресії і навчимо його
model = LogisticRegression()
model.fit(X_train, y_train)
# Виконуємо прогнозування на основі тестових даних
y_pred = model.predict(X_test)
y_pred_proba = model.predict_proba(X_test)[:, 1]
# Оцінимо точність моделі(в проміжку від 0 до 1, де 1- 100%, а 0 - 0%)
accuracy = accuracy_score(y_test, y_pred)
print(f'Точність моделі: {accuracy} for k: {k}')
results[k] = [accuracy, model]
# Розрахунок AUC ROC
fpr, tpr, thresholds = roc_curve(y_test, y_pred_proba)
roc_auc = auc(fpr, tpr)
print(f'ROC AUC: {roc_auc} for k: {k}')
# Відображення ROC кривої
display = RocCurveDisplay(fpr=fpr, tpr=tpr, roc_auc=roc_auc, estimator_name='Example
Estimator')
display.plot()
plt.show()
# Розрахунок коефіцієнта Джині
gini_coefficient = 2 * roc_auc - 1
print(f'Коефіцієнт Джинні: {gini_coefficient} for k: {k}')
# Вивід матриці помилок
cm = confusion_matrix(y_test, y_pred)
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=['Клас 0', 'Клас 1'],
yticklabels=['Клас 0', 'Клас 1'])
plt.xlabel('Передбачені мітки')
plt.ylabel('Актуальні мітки')
plt.title('Матриця помилок')
plt.show()
#Перевірка моделі для ЕТН
def check_model_accuracy(df, results):
    for k in range(1,28//4, 1):

```

```

# Згенеруємо деякі вигадані дані для класифікації
# X - це набір ознак, y - вектор цільових значень (0 або 1)
X_test = df[['f'tsi_{k}', 'f'vpti_{k}', 'f'close_pct_{k}']].values[:-1]
y_test = df[f'close_pct_{k}_logistic'].values[1:]
model = results[k][1]

# Виконуємо прогнозування на основі тестових даних
y_pred = model.predict(X_test)

# Оцінимо точність моделі
accuracy = accuracy_score(y_test, y_pred)

print(f'Точність моделі: {accuracy} for k: {k}, on BTC was: {results[k][0]}")

print('Results for ETH')
check_model_accuracy(ETH_train, results)
print('Results for ZRX')
check_model_accuracy(ZRX_train, results)
print('Results for BOME')
check_model_accuracy(BOME_train, results)

#Побудова моделі дерева рішень
results = {}

for k in range(1,28//4, 1):
    X_train = BTC_train[['f'tsi_{k}', 'f'vpti_{k}', 'f'close_pct_{k}']].values[k:-1]
    y_train = BTC_train[f'close_pct_{k}_logistic'].values[k+1:]
    X_test = BTC_test[['f'tsi_{k}', 'f'vpti_{k}', 'f'close_pct_{k}']].values[k:-1]
    y_test = BTC_test[f'close_pct_{k}_logistic'].values[k+1:]

    # Створимо екземпляр моделі логістичної регресії і навчимо його
    classifier = DecisionTreeClassifier(max_depth=3) # Можна варіювати параметр max_depth
    classifier.fit(X_train, y_train)

    # Робимо прогнози
    y_pred = classifier.predict(X_test)
    y_pred_proba = classifier.predict_proba(X_test)[:, 1]

    # Оцінимо точність моделі(в проміжку від 0 до 1, де 1- 100%, а 0 - 0%)
    accuracy = accuracy_score(y_test, y_pred)

    print(f'Точність моделі: {accuracy} for k: {k}")
    results[k] = [accuracy, model]

# Розрахунок AUC ROC
fpr, tpr, thresholds = roc_curve(y_test, y_pred_proba)

```

```

roc_auc = auc(fpr, tpr)
print(f"ROC AUC: {roc_auc} for k: {k}")
# Відображення ROC кривої
display = RocCurveDisplay(fpr=fpr, tpr=tpr, roc_auc=roc_auc, estimator_name='Example
Estimator')
display.plot()
plt.show()
# Розрахунок коефіцієнта Джині
gini_coefficient = 2 * roc_auc - 1
print(f"Коефіцієнт Джинні: {gini_coefficient} for k: {k}")
# Вивід матриці помилок
cm = confusion_matrix(y_test, y_pred)
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=['Клас 0', 'Клас 1'],
yticklabels=['Клас 0', 'Клас 1'])
plt.xlabel('Передбачені мітки')
plt.ylabel('Актуальні мітки')
plt.title('Матриця Помилок')
plt.show()
print('Results for ETH')
check_model_accuracy(ETH_test, results)
print('Results for ZRX')
check_model_accuracy(ZRX_test, results)
print('Results for BOME')
check_model_accuracy(BOME_train, results)

```



метадані

Заголовок

Савченко_Матвійчук_плагіат4

Автор

Науковий керівник / Експерт

Савченко**Матвійчук**

підрозділ

кафедра математичного моделювання та статистики

Тривога

У цьому розділі ви знайдете інформацію щодо текстових спотворень. Ці спотворення в тексті можуть говорити про МОЖЛИВІ маніпуляції в тексті. Спотворення в тексті можуть мати навмисний характер, але частіше характер технічних помилок при конвертації документа та його збереженні, тому ми рекомендуємо вам підходити до аналізу цього модуля відповідально. У разі виникнення запитань, просимо звертатися до нашої служби підтримки.

Заміна букв		3
Інтервали		0
Мікропробіли		3
Білі знаки		0
Парафрази (SmartMarks)		15

Обсяг знайдених подібностей

Коефіцієнт подібності визначає, який відсоток тексту по відношенню до загального обсягу тексту було знайдено в різних джерелах. Зверніть увагу, що високі значення коефіцієнта не автоматично означають плагіат. Звіт має аналізувати компетентна / уповноважена особа.



КП 1

25

Довжина фрази для коефіцієнта подібності 2



КП 2

9059

Кількість слів



КЦ

66940

Кількість символів

Подібності за списком джерел

Нижче наведений список джерел. В цьому списку є джерела із різних баз даних. Колір тексту означає в якому джерелі він був знайдений. Ці джерела і значення Коефіцієнту Подібності не відображають прямого плагіату. Необхідно відкрити кожне джерело і проаналізувати зміст і правильність оформлення джерела.

10 найдовших фраз

Копір тексту

ПОРЯДКОВИЙ НОМЕР	НАЗВА ТА АДРЕСА ДЖЕРЕЛА URL (НАЗВА БАЗИ)	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)	
1	https://ir.kneu.edu.ua/bitstreams/151df6cf-1099-4401-925c-5a0fe3423590/download	27	0.30 %
2	https://ir.kneu.edu.ua/bitstreams/151df6cf-1099-4401-925c-5a0fe3423590/download	25	0.28 %
3	http://ippi.org.ua/sites/default/files/maket.pdf	17	0.19 %
4	Plurality the types of art in contemporary art: specifics of media art Тетяна Міронова;	14	0.15 %
5	https://ir.kneu.edu.ua/server/api/core/bitstreams/01411c28-e954-480c-9741-533e5822b9ac/content	13	0.14 %

6	https://tradersunion.com/ua/interesting-articles/best-cryptocurrency-to-invest/private-cryptocurrency/	12	0.13 %
7	Вечерковська_Поперешняк.docx 12/4/2023 Publishing House "Helvetica" (Видавничий дім "Гельветика")	11	0.12 %
8	https://stud.wiki/programming/3c0a65625b3ad78a4d53a89521216c36_0.html	11	0.12 %
9	Прогнозування обсягів збуту хлібопекарського підприємства 6/3/2020 Kyiv National Economic University named after Vadym Hetman KNEU (кафедра економіко-математичного моделювання)	10	0.11 %
10	http://cit-journal.com.ua/index.php/cit/article/view/170/243	8	0.09 %

з бази даних RefBooks (0.15 %)

ПОРЯДКОВИЙ НОМЕР	ЗАГОЛОВОК	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)	
джерело: Paperity			
1	Plurality the types of art in contemporary art: specifics of media art Тетяна Міронова;	14 (1)	0.15 %

з домашньої бази даних (0.20 %)

ПОРЯДКОВИЙ НОМЕР	ЗАГОЛОВОК	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)	
1	Прогнозування обсягів збуту хлібопекарського підприємства 6/3/2020 Kyiv National Economic University named after Vadym Hetman KNEU (кафедра економіко-математичного моделювання)	10 (1)	0.11 %
2	Прогнозування споживчого попиту в Україні 6/8/2020 Kyiv National Economic University named after Vadym Hetman KNEU (кафедра економіко-математичного моделювання)	8 (1)	0.09 %

з програми обміну базами даних (0.12 %)

ПОРЯДКОВИЙ НОМЕР	ЗАГОЛОВОК	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)	
1	Вечерковська_Поперешняк.docx 12/4/2023 Publishing House "Helvetica" (Видавничий дім "Гельветика")	11 (1)	0.12 %

з Інтернету (1.51 %)

ПОРЯДКОВИЙ НОМЕР	ДЖЕРЕЛО URL	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)	
1	https://ir.kneu.edu.ua/bitstreams/151df6cf-1099-4401-925c-5a0fe3423590/download	52 (2)	0.57 %
2	https://otherreferats.allbest.ru/programming/01283830_0.html	18 (3)	0.20 %
3	http://ippi.org.ua/sites/default/files/maket.pdf	17 (1)	0.19 %
4	https://ir.kneu.edu.ua/server/api/core/bitstreams/01411c28-e954-480c-9741-533e5822b9ac/content	13 (1)	0.14 %

5	https://tradersunion.com/ua/interesting-articles/best-cryptocurrency-to-invest/private-cryptocurrency/	12 (1)	0.13 %
6	https://stud.wiki/programming/3c0a65625b3ad78a4d53a89521216c36_0.html	11 (1)	0.12 %
7	http://cit-journal.com.ua/index.php/cit/article/view/170/243	8 (1)	0.09 %
8	https://ua-referat.com/uploaded/metodichni-rekomendaciyi-shodo-pidgotovki-ta-zahistu-kvalifika/index4.html	6 (1)	0.07 %

Список прийнятих фрагментів (немає прийнятих фрагментів)

ПОРЯДКОВИЙ НОМЕР	ЗМІСТ	КІЛЬКІСТЬ ОДНАКОВИХ СЛІВ (ФРАГМЕНТІВ)
------------------	-------	---------------------------------------