

Міністерство освіти і науки України
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА АКАДЕМІЯ»
Кафедра Інформатики Факультету Інформатики



Прогнозування курсу криптоактивів

Керівник курсової роботи

Невмержицький Є. І.

“ ____ ” _____ 2022 р.

Виконав студент

Мережко Д. Р.

“ ____ ” _____ 2022 р.

Київ 2022

Тема: Прогнозування курсу криптоактивів

Календарний план виконання роботи:

/п	Назва етапу курсового проекту (роботи)	Термін виконання етапу	Примітка
1.	Отримання завдання на курсову роботу.	20.10.2021	
2.	Огляд технічної літератури за темою роботи.	01.11.2021	
3.	Планування теоретичної та практичної частини роботи. Створення детального плану виконання.	25.11.2021	
4.	Обговорення практичної частини з керівником, визначення напрямку застосування.	20.12.2021	
5.	Вивчення специфіки крипторинку, огляд існуючих інструментів для його аналізу.	05.01.2022	
6.	Закінчення вивчення теоретичної частини, що стосується взаємодії користувачів із біржами, вивчення принципів використання блокчейну у цьому контексті.	15.01.2022	
7.	Вивчення електронних ресурсів щодо лінійної регресії.	05.02.2022	
8.	Визначення засобів реалізації практичної частини.	10.02.2022	
9.	Написання коду для практичної частини пов'язаною із регресією.	08.04.2022	
10.	Написання коду для частини, пов'язаної з часовими рядами.	01.05.2022	
11.	Здача курсової роботи.		

Студент _____

Керівник _____

“ _____ ”

Зміст

Вступ.....	3
РОЗДІЛ 1: Описання ідеї криптоактивів.....	4
1.1 Криптовалюти як явище	4
1.2 Специфіка принципів функціонування крипторинку.....	6
1.3 Ринок та криптобіржі з точки зору комп'ютерних наук.....	8
РОЗДІЛ 2: Використання математичних та статистичних методів для аналізу стану крипторинку	13
2.1 Опис цікавих, з точки зору аналізу, криптоактивів	13
2.2 Математичні та статистичні методи аналізу	16
РОЗДІЛ 3: Практична частина	24
3.1 Аналіз проблеми	24
3.2 Опис запропонованого програмного продукту	24
3.3 Огляд факторів, що підлягають аналізу, та математичних інструментів, що для нього застосовуються.....	25
3.4 Визначення засобів реалізації.....	26
3.5 Опис програмної реалізації продукту та демонстрація роботи	28
Висновки.....	35
Список використаних джерел.....	36
Додатки	37
Додаток А. Software Design Description застосунку для прогнозування курсу	37

Вступ

Феномен виникнення криптовалют у сучасному світі та їх активний вплив як на економіку, так і на політику, став можливим лише завдяки активному розвитку математики, комп'ютерних наук та криптографії. Зі створення генезис-блоку біткоїна у 2009 році, темпи розвитку цієї сфери тільки зростають: впровадження смартконтрактів, можливість цифровізувати та перевести на блокчейн практично будь-який актив із реального світу, вихід на централізовані біржі – це лише деякі з кроків, які зробила криптоіндустрія за останні двадцять років. Із часом авторитет цієї сфери зростає, і тепер вона використовується багатьма міжнародними компаніями, а тренд на використання цифрових активів набуває популярності.

Оскільки класична торгівля на біржі підпорядковується певним математичним законам, дані, отримані з неї, можуть бути проаналізовані, а результати аналізу – вивчені та використані у подальшій економічній діяльності. Один з висновків, що може бути зроблений – той, що стосується прогнозу відносно руху курсу обраного активу на відповідному проміжку часу. Аналогічні прогнози, з певними корегуваннями, можна отримати і відносно криптовалют.

Дана робота присвячена розгляду питання автоматизації аналізу та прогнозування курсу криптоактивів, за допомогою програми, що змогла б, за допомогою машинного навчання, обробити певний об'єм вже наявних даних про торги та на основі них обрахувати прогнозований курс. У роботі описуються технічні засоби, які надають криптовалютні біржі та розробники криптоактивів для взаємодії з даними, що використовуються для дослідження економічних показників. Будуть розглянуті методи аналізу стану ринку з точки зору математики та статистики, а також уже наявні програмні рішення, розроблені для вирішення задачі прогнозування.

РОЗДІЛ 1: Описання ідеї криптоактивів

1.1 Криптовалюти як явище

Поява цифрових грошей була зумовлена недовірою до традиційних фінансових інституцій після економічної кризи 2008 року. Результатом стала необхідність створення валюти, що була б незалежна від банків, регуляторів, та будь-яких інших посередників, що вилилось у розробку біткоїна – перших електронних грошей, на які не діє інфляція.

Криптовалютою називають цифровий актив, функціонування та цінність якого забезпечені за допомогою засобів криптографії, що дозволяє забезпечити цілісність даних, зашифрувати їх відповідно до певних правил, та однозначно ідентифікувати. Наприклад, у логіці біткоїна використовується алгоритм ECDSA (Elliptic Curve Digital Signature Algorithm) – схема криптографії на основі еліптичних кривих. Алгоритм ECDSA використовує еліптичну криву та кінцеве поле для створення підпису даних таким чином, що треті сторони можуть верифікувати автентичність підпису, а сторона, що підписала, зберігає ексклюзивну можливість створювати підпис. У випадку біткоїна, дані, що підписуються - це транзакція передачі права власності. На даній діаграмі наведено схему шифрування та здійснення транзакції біткоїна.



Усі дані про транзакції відповідної криптовалюти записуються у блокчейн - розподілену базу даних, у якій для проведення визначеної операції необхідно підтвердження усіх вузлів системи. Це дозволяє практично повністю усунути можливість зламу та викрадення даних. Підтвердження транзакції, у свою чергу, відбувається за певним математичним алгоритмом, визначеним у кожному блокчейні, який називається алгоритмом консенсусу.

Криптовалюта зберігається на публічному ключі користувача – він є унікальним набором символів, що виконує роль цифрового гаманця, з якого криптовалюта не може бути викрадена. У свою чергу, приватний ключ – це унікальний набір символів, що відомий тільки кінцевому власнику криптовалюти та використовується для переведення активів з одного публічного ключа на інший.

Основними способами отримання криптовалюти є торгівля (між кінцевими власниками або за допомогою біржі) та майнінг. Останнє представляє собою обчислення хешів – стрічок, що складаються з літер та чисел, що підтверджують валідність інформації, не розкриваючи її змісту. За знаходження хешів у блоках транзакцій користувачі, які надають свої технічні ресурси, винагороджуються відповідними сумами криптовалюти, майнінг якої відбувається.

Емісія (кількість випущеної криптовалюти) підпорядковується математичним алгоритмам та є контрольованою. Цифрові гроші є дефляційним активом, тобто таким, що не може знецінюватись в результаті збільшення своєї кількості. Це досягнуто завдяки заздалегідь визначеному числу одиниць валюти, а деякі розробники впроваджують механізми знищення певної кількості розроблених токенів щоквартально або щорічно – відправлення певної суми на заблокований публічний ключ, виводячи таким чином валюту з обігу.

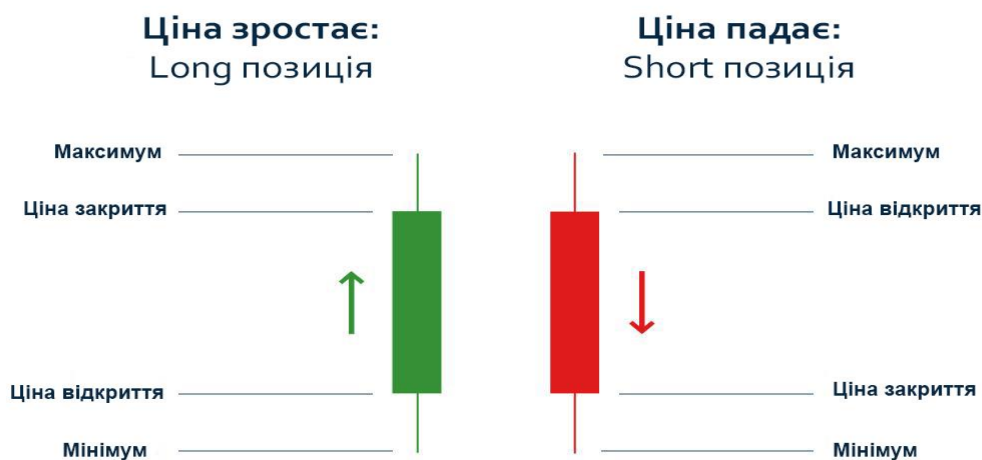
З часом пропозиція криптовалют на ринку буде зменшуватися, що при збереженні попиту надає величезний потенціал росту.

1.2 Специфіка принципів функціонування крипторинку

Крипторинок – це сукупність усіх засобів та інструментів, які стимулюють обмін криптовалютами між зацікавленими особами. Іншими словами, це комплекс умов, в яких формуються попит та пропозиція на купівлю цифрових грошей. Зацікавленість у торгівлі полягає в тому, що вложивши одну суму грошей у певний актив, дочекавшись підняття ціни на нього, трейдер продає його, а різниця в ціні і є прибутком. Крипторинок за принципами свого функціонування дуже схожий на класичний фондовий ринок, але, звісно, зі своєю специфікою. Розглянемо основні засади, на яких формується ринок.

Торгівля відбувається між валютними парами, як і на класичному ринку (наприклад, у парі USD/UAH людина купує долари США, при цьому обмінюючи їх на еквівалентну суму української гривні), лише з тою поправкою, що валютні пари можуть формуватися як між криптоактивами, так і між криптовалютою і фіатними грошима (BTC/ETH - обмін між криптовалютами, BTC/UAH - між фіатом). Отже, купуючи криптовалюту за еквівалент фіатних грошей відповідно до курсу, користувач стає її власником. Тепер він може брати участь у торгах – виставляючи ордери на купівлю або продаж. При цьому угода між продавцем і зацікавленим покупцем відбувається за ціною, яка зараз панує на ринку, користувач не може продати валюту за завищеною ціною. Відкриваючи ордер із прогнозуванням, що ціна на актив буде зростати, трейдер входить у, так звану, “long” позицію, очікуючи, що продасть криптовалюту дорожче, ніж купив. Очікуючи падіння ціни, відкривається ордер на продаж, що називається “short” позицією. Нижче наведено діаграму «японської свічі» -

елемента графіку, що найчастіше використовується для візуалізації руху курсу.



Існують такі варіанти торгівлі: спотовий та ф'ючерсний. Різниця між ними полягає у тому, що перший оперує безпосередньо криптоактивами, миттєво проводячи обмін торгової пари між покупцем і продавцем, а другий із контрактами на актив, купуючи які, трейдер зобов'язується продати або купити певну криптовалюту у майбутньому, при цьому забираючи різницю у зміні в ціні, у порівнянні із курсом на момент відкриття позиції. Різниця зростає прямо пропорційно сумі, на яку було куплено ф'ючерсний контракт. Зацікавленість у ф'ючерсній торгівлі полягає у можливості виставити «плече» - множник, який дозволяє торгувати більшою кількістю криптовалюти, ніж є у користувача, відповідно збільшуючи винагороду у разі успішного прогнозу.

Також часто користувачі, для подальшого обміну, переводять фіатні кошти у стейблкоїни. Це цифрові активи, які прив'язані до курсу фіатних валют у відношенні 1:1 (у переважній більшості випадків, але не завжди), тобто є повним їх відповідником на крипторинку. Стейблкоїни можуть бути прив'язані до будь-якого активу із реального світу, на кшталт золота, срібла або нафти. Першою такою валютою став еквівалент долару США, USDT, від компанії Tether, що був випущений у 2015 році. На відміну від інших криптовалют, стейблкоїни не змінюють своєї ціни, а їх введення було

мотивовано надати іншим криптовалютам певний еквівалент цінності з реального світу, але при цьому розташований на блокчейні.

Іншою відмінністю від класичного ринку є волатильність криптовалют, що є частотою і діапазоном зміни ціни криптовалюти. Цей показник набагато більший на крипторинку, оскільки завдяки децентралізації активи не прив'язані до жодної з економік країн світу, що провокує дуже велику залежність від інформаційного фону, найбільше – від новин. Це може провокувати як стрімкі падіння курсу, так і стрімке підняття ціни. Така відсутність будь-яких регуляторів часто і стає причиною відмови фінансових інституцій виходити на ринок цифрових активів.

1.3 Ринок та криптобіржі з точки зору комп'ютерних наук

Торгівля криптовалютами відбувається на криптобіржах. Це платформи, що надають кінцевим користувачам купити цифрові активи, вивівши фіатні кошти зі свого банківського рахунку, використовуючи системи електронних платежів.

На даний момент у п'ятірку найпопулярніших та найбільших за капіталізацією бірж входять:

- Binance із середнім показником об'єму торгів за добу у розмірі 12 млрд. доларів США
- Coinbase Exchange – об'єм торгів за добу: 2 млрд. доларів США
- FTX – об'єм торгів за добу: 1,2 млрд. доларів США
- Huobi Global – об'єм торгів 1,3 млрд. доларів США
- KuCoin – об'єм торгів 1,9 млрд. доларів США

Сама по собі криптобіржа представляє собою веб-сайт або мобільний застосунок, в якому користувач реєструється, прив'язуючи свій акаунт до відповідних публічних ключів. Неможливо зберігати усю криптовалюту, придбану на біржі, на одному публічному ключі через те, що різні криптоактиви використовують різні блокчейни, які можуть мати навіть різну кількість символів у ключі. Тому біржа бере на себе зобов'язання

автоматично реєструвати на користувача певний перелік публічних ключів, в залежності від того, якими криптовалютами він володіє.

Із часом, розвиток цього сегменту спровокував появу повноцінних блокчейн платформ – протоколів, на базі яких можна створювати нові продукти. Це реалізовано за допомогою смарт-контрактів, які є комп'ютерними програмами, які можна завантажити в певний блокчейн і використовувати потужності мережі, валідуючи певні дії на її нодах(вузлах). Більше того, смарт-контракти можуть використовуватися навіть поза крипторинком, практично у будь якій сфері, де необхідно забезпечити чесний результат перевірки. У свою чергу, найпопулярнішими блокчейнами для створення криптовалют є:

- Ethereum
- Binance Smart Chain
- Solana
- Avalanche C-Chain
- Polygon

Кожна з даних мереж має свою специфіку функціонування: вид консенсусу, час обробки транзакцій, час генерації нових блоків, комісії за транзакції, тощо. Також є багато проєктів, що переводять криптовалюти на інший блокчейн, наприклад Wrapped Bitcoin – повний еквівалент біткоїну, реалізований у стандарті BEP20 (Binance Smart Chain). Розглянемо це на прикладі біткоїну та Binance Smart Chain.

Обмін у Binance



Однак, усі описані платформи об'єднує те, що блокчейн надає безпеку, гарантує, що кошти не загубляться при транзакції та забезпечує анонімність.

Однак, використовуючи централізовані біржі, наприклад Binance або Coinbase, користувач погоджується із обробкою персональних даних, оскільки біржа у цьому сценарії виступає регулятором. За своєю ідеєю криптовалюти мають забезпечувати анонімність, тому для цього були створені децентралізовані біржі – DEX (Decentralized Exchange). Це одноранговий ринок, де транзакції відбуваються безпосередньо між крипто-трейдерами. DEX виконує одну з основних можливостей криптовалют: сприяння фінансовим транзакціям, які не здійснюються банками, брокерами, платіжними процесорами чи будь-яким іншим посередником. Найпопулярніші DEX, такі як Uniswap і Sushiswap, використовують блокчейн Ethereum і є частиною зростаючого набору інструментів децентралізованих фінансів (DeFi), які роблять величезний спектр фінансових послуг доступними безпосередньо з сумісного криптовалютного гаманця. DEX активно розвивається — у першому кварталі 2021 року через децентралізовані біржі пройшло 217 мільярдів доларів США. Станом на

квітень 2021 року налічувалося понад два мільйони трейдерів DeFi, що в 10 разів більше, ніж у травні 2020 року.

На відміну від централізованих бірж, таких як Coinbase, DEX не дозволяють обмінюватися між фіатними та криптовалютами — натомість вони торгують виключно токенами криптовалюти на інші токени. Через централізовану біржу (або CEX) трейдер може обмінювати фіат на криптовалюту (і навпаки) або крипто-крипто пари – частину своїх біткоїнів на ETH. Можна робити більш просунуті кроки, як-от маржинальні операції або встановлення лімітних ордерів. Але всі ці транзакції обробляються самою біржею через «книгу замовлень», яка встановлює ціну на певну криптовалюту на основі поточних замовлень на купівлю та продаж. Нижче наведена діаграма з демонстрацією роботи Uniswap.



З іншого боку, децентралізовані біржі — це просто набір смартконтрактів. Вони встановлюють ціни різних криптовалют щодо кожної з них алгоритмічно і використовують «пули ліквідності», в яких інвестори блокують кошти в обмін на винагороду, подібну до відсотків, для полегшення торгівлі. У той час як транзакції на централізованій біржі реєструються у внутрішній базі даних цієї біржі, транзакції DEX здійснюються безпосередньо в блокчейні.

DEX зазвичай будуються на основі відкритого вихідного коду, а це означає, що кожен зацікавлений може побачити, як саме вони працюють. Це також означає, що розробники можуть адаптувати існуючий код для

створення нових конкуруючих проектів — саме так код Uniswap був адаптований цілим набором DEX кросчейн-проектів із «swap» у своїх назвах, як-от Sushiswap і Pancakeswap.

Однак, така торгівля є незахищеною: деякі зловмисники, випускаючи свої токени, прописують у смарт контрактах такий код, що не дозволяє виводити ці токени на свій гаманець, таким чином забираючи гроші користувачів собі.

Аналіз курсу криптоактиву доречний у тому випадку, коли можна бути впевненим, що гроші не будуть викрадені, тому логічно розглядати торгівлю на централізованих біржах. До того ж, вони можуть надати більшу кількість метрик для статистичного дослідження майбутнього руху.

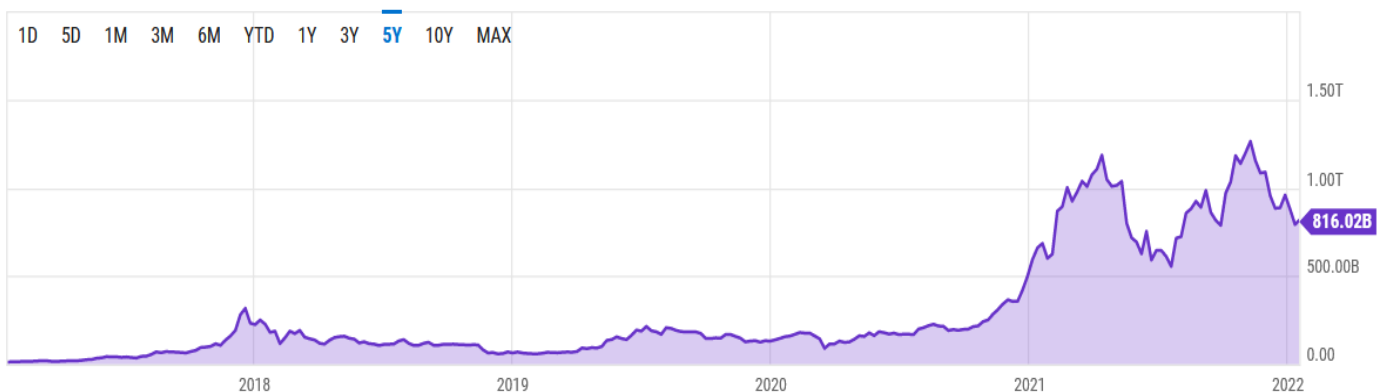
Підбиваючи підсумок, бачимо, що технічні складності торгівлі криптовалютою, як-от невідповідність протоколів, надійність смарт-контракту або недостатня ліквідність токену практично завжди забезпечується централізованими біржами, хоча і DEX, при вмілому використанні, дають таку можливість. Це означає, що сам ринок це стабільна система, що підтримується мільйонами зацікавлених осіб, і яку можна аналізувати за допомогою відповідних метрик, не переймаючись, що досліджувана криптовалюта кудись зникне, чи блокчейн припинить свою роботу. Тому скориставшись необхідними математичними інструментами та індикаторами, можемо прогнозувати рух ціни, що ми і розглянемо далі.

РОЗДІЛ 2: Використання математичних та статистичних методів для аналізу стану крипторинку

2.1 Опис цікавих, з точки зору аналізу, криптоактивів

Розглянемо список активів, аналіз яких був би доречним та дав найбільш ефективні результати дослідження.

На даний момент з усіх криптовалют, яких налічується більше 2,5 тисяч, біткоїн (BTC) має найбільшу долю на ринку та має капіталізацію приблизно 800 мільярдів доларів США. Можна сказати, що біткоїн, власне, і складає авторитет криптовалют, оскільки він з'явився найпершим з усіх та широко згадувався у ЗМІ. Навіть людина, яка не цікавиться економікою або технологіями, обов'язково чула про нього протягом останніх років. На графіку нижче наведені показники капіталізації за останні 5 років.



Через це біткоїн, хоч і не був створений із розрахунку на це, регулює майже весь крипторинок. Волатильність головної криптовалюти суттєво впливає майже на всі альткоїни, хоча напряду вони ніяк не пов'язані. Таку залежність пов'язують із тим, що через незалежність від економік найпотужніших держав світу, біткоїн став складати так званий «net worth» усієї криптоіндустрії. На жаль, це провокує сильну залежність від новин, що пов'язані із ринком. Наприклад, суттєві падіння відбуваються на фоні

засідань фінансових регуляторів США або Китаю, біткоїн також реагує майже на будь яку політичну нестабільність. Також треба сказати, що біткоїн не має блокчейн-платформи, що використовувалась поза фінансовою сферою. Це робить його далеко не найбільш технологічним, особливо з огляду на інші проекти, що з'явилися останнім часом.

Однак, проводячи аналіз, брати до уваги головну криптовалюту просто необхідно, оскільки саме вона визначає глобальний тренд усього ринку та диктує зміну курсу іншим криптоактивам.

Іншою криптовалютою, що має велику вагу на крипторинку має ефір (ETH), капіталізація якого складає приблизно 370 мільярдів доларів США. Ефір відрізняється тим, що став першою криптовалютою, у якій почали використовуватись смарт контракти, що надало можливість використовувати можливості платформи ефіра для узгодження та валідації інформації зі сфер, що не пов'язані напряду із фінансовим сектором. Вони можуть використовуватись для безпечних купівель товарів, децентралізованого зберігання даних, у медицині та навіть у механізмі виборів, щоб забезпечити однозначну аутентифікацію та унікальність голосу.

Ефір розроблений за допомогою стандарту ERC20, що надає широкі можливості для створення на базі цього блокчейну інших проектів та валют. Зокрема, набуває розвитку сегмент децентралізованих фінансових застосунків (DeFi), які допомагають користувачам у фінансовій сфері, забезпечуючи у розрахунках усі переваги, які надає блокчейн. Єдиний мінус, який наявний при використанні ETH mainnet (головна мережа ефіру, через яку відбуваються транзакції) є відносно велика сума комісії за проведення транзакції. Через це багато застосунків використовують для аналогічних операцій, як-от розміщення NFT у мережі, блокчейн Polygon. Нижче наведена діаграма екосистеми DeFi, базована на ефірі.

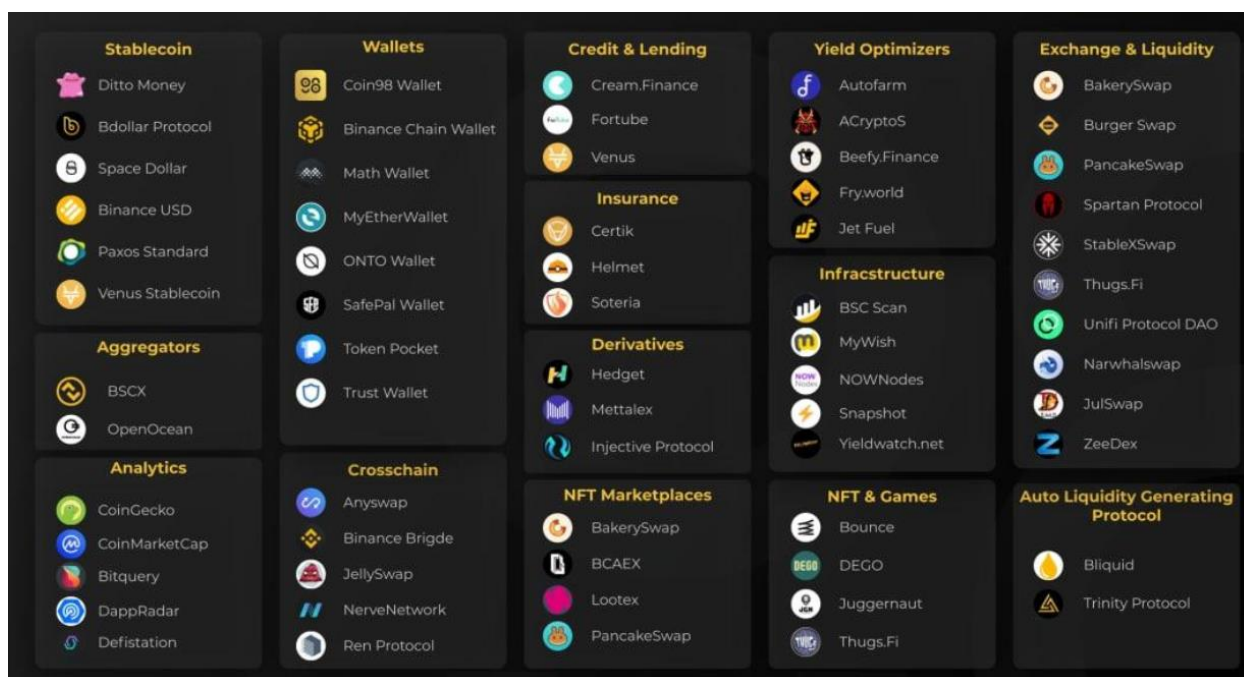


Також ERC20 породив таке явище як крипто-ігри (стандарт ERC721 зокрема), тобто мобільні або веб-застосунки, проводячи час у яких гравець отримує винагороду у виді криптовалюти, прив'язаної до гри (наприклад, Axie Infinite). Іншим популярним використанням блокчейну ефіра стали NFT, які дозволили цифровізувати та уніфікувати об'єкти мистецтва.

Загалом, ефір є фундаментальною криптовалютою, яка використовує технологію, що підкріплена десятками застосунків. Відповідно, прогножуючи курс криптоактивів, що базуються на ERC20, необхідно розуміти динаміку руху ефіра.

Наступною фундаментальною криптовалютою, яку було б доречно аналізувати, є BNB, власний токен біржі Binance, що створений за стандартом BEP20. На ньому випущений стейблкоїн BUSD (Binance USD), а також CAKE (нативний токен децентралізованої біржі PancakeSwap). Зараз смарт контракти багатьох криптовалют мають адреси у Binance Smart Chain, що робить важливим для аналізу як сам блокчейн, так і саму біанс коін.

Нижче наведена схема, на якій зображено проекти, що працюють на BSC.



Звісно, розглянуті криптовалюти можуть бути не найбільш технологічними з точки зору коду та логіки. Наразі є проекти, що надають дійсно нові можливості у сфері криптоактивів, наприклад TORNADO.Cash, що надає можливість майнінгу анонімності або Ensurg, що надає можливість застрахувати активи та отримувати з цього прибуток. Однак, три розглянуті криптовалюти є найбільш масштабними інфраструктурними проектами, що визначають напрямок руху ціни і які необхідно брати до уваги при детальному дослідженні ситуації на ринку.

2.2 Математичні та статистичні методи аналізу

Перш за все, необхідно визначити, якими засобами та інструментами можна ефективно аналізувати ринок загалом та курс конкретних цифрових активів зокрема. Необхідно обрати такий спосіб, що дозволяє проводити обробку великих об'ємів даних, зібраних за певний період, та обробляти їх, використовуючи певну модель, що аналізує дані на основі визначеного алгоритму.

Краще за все для таких функцій підходить машинне навчання – сукупність математичних методів, за допомогою яких програма «навчається»

за допомогою певного, наданого розробником, набору вхідних даних та може використовувати результати цієї обробки задля подальшого використання у прогнозуванні можливих варіантів поведінки системи.

Відповідно – чим більше модель «тренується», тобто оброблює тестові дані, тим більш точним буде подальше прогнозування, що може принести розробнику користь, оскільки він знатиме, як приблизно може розвиватися ситуація, що у сфері крипторинку є особливо важливим, тому, що це веде до збільшення можливого заробітку.

Машинне навчання у сфері торгівлі можна застосувати для того, щоб ефективніше бачити тенденції ринку та виключити із торгівлі людський фактор, який провокує паніку або жадібність у трейдера під час значних коливань ціни. На крипторинку є певні послідовності у перебігу торгівлі, знаючи які, можна зробити торгівлю на біржі більш успішною.

Розглянемо три алгоритми машинного навчання, що найчастіше використовуються у контексті прогнозування курсу криптовалют: лінійна регресія, нейронні мережі LSTM (long short-term memory) та ...

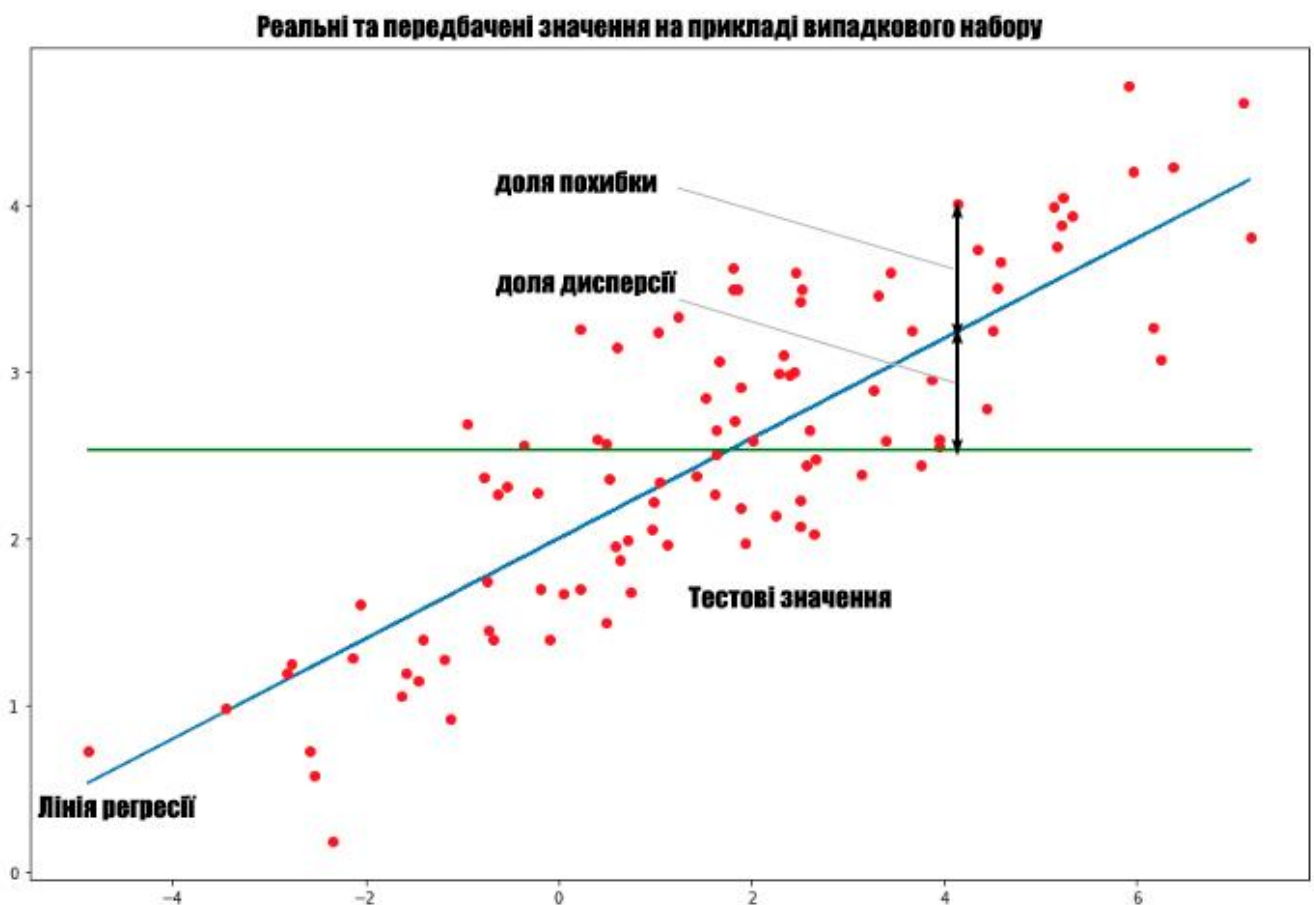
Лінійна регресія

Лінійна регресія є основним і широко використовуваним типом прогнозного аналізу. Регресія - спосіб вибрати із сімейства функцій ту, яка мінімізує функцію втрат. Вона характеризує, наскільки сильно тестова функція відхиляється від значень у заданих точках. Якщо точки отримані експериментально, вони неминуче містять помилку вимірювань, шум, тому розумніше вимагати, щоб функція передавала загальну тенденцію, а не точно проходила всі точки.

Ми хочемо провести криву якомога ближче до точок і при цьому зберегти її максимально простою, щоб вловити загальну тенденцію. За баланс між цими суперечливими бажаннями відповідає функція втрат (в англійській літературі «loss function» або «cost function»). Мета регресії — знайти коефіцієнти лінійної комбінації базисних функцій, і цим визначити

регресійну функцію f (яку також називають моделлю). Лінійну регресію називають лінійною саме через лінійну комбінацію базисних функцій — це не пов'язано з самими функціями (вони можуть бути як лінійними, так і нелінійними).

Найпростіша форма рівняння регресії з однією залежною та однією незалежною змінною визначається за формулою $y = c + b \cdot x$, де y = оцінка залежної змінної, c = константа, b = коефіцієнт регресії і x = оцінка для незалежна змінна. Проста лінійна регресія може бути представлена як показано нижче. Червоні крапки є тестовими значеннями, що використовуються для навчання моделі, а синя лінія демонструє передбачені значення.



По-перше, регресія може бути використана для визначення сили впливу незалежної змінної (змінних) на залежну змінну. Типовими питаннями є те, яка сила зв'язку між дозою та ефектом, витратами на продажі та маркетинг чи віком і доходом.

По-друге, його можна використовувати для прогнозування наслідків або впливу змін. Тобто регресійний аналіз допомагає нам зрозуміти, наскільки залежна змінна змінюється при зміні однієї чи кількох незалежних змінних.

По-третє, регресійний аналіз передбачає тенденції та майбутні значення. Для отримання точкових оцінок можна використовувати регресійний аналіз. Саме для нашої задачі це підходить ідеально, дозволяючи спрогнозувати, якою буде ціна досліджуваного криптоактиву через певний проміжок часу.

Лінійна регресія може бути простою, коли існує залежність лише від однієї змінної, однак у контексті прогнозування курсів криптовалют доречно використовувати залежність від багатьох змінних, оскільки ціна завжди залежить від великої кількості факторів, як от:

- Ринкова капіталізація
- Обсяг торгів
- Кількість транзакцій
- Середній розмір блоку
- Комісія за транзакцію
- Унікальні адреси
- Хешрейт
- Індекс NASDAQ
- Індекс Dow Jones

Більше того, до чинників, що впливають на ціну, належать загальна економічна ситуація у світі, політичні кризи, заяви Федеральної Резервної Служби США, тощо. Однак, ми будемо розглядати ситуацію із математичної точки зору, беручи до уваги лише статистичні показники, виключаючи можливість появи гучних інфоповодів.

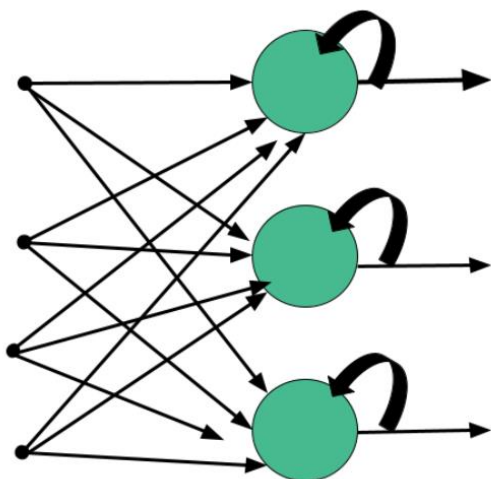
При виборі моделі для аналізу важливим фактором є її відповідність. Додавання незалежних змінних до моделі лінійної регресії завжди збільшує дисперсію, що негативно впливає на результати. Прості моделі зазвичай краще, ніж складні, оскільки статистично, якщо модель включає велику

кількість змінних, деякі з них будуть статистично значущими лише через випадковість.

Нейронні мережі RNN (recurrent neural network)

Рекурентна нейронна мережа (RNN) — це тип штучної нейронної мережі, яка використовує послідовні дані або дані часових рядів. Ці алгоритми глибокого навчання зазвичай використовуються для таких задач як переклад, обробка природної мови, розпізнавання мовлення та підписання зображень; вони включені в популярні програми, такі як Siri, голосовий пошук і Google Translate. RNN відрізняються своєю «пам'яттю», оскільки вони беруть інформацію з попередніх входів, щоб впливати на поточний вхід і вихід даних. У той час як традиційні глибокі нейронні мережі припускають, що входи та виходи незалежні один від одного, вихід рекурентних нейронних мереж залежить від попередніх елементів у послідовності.

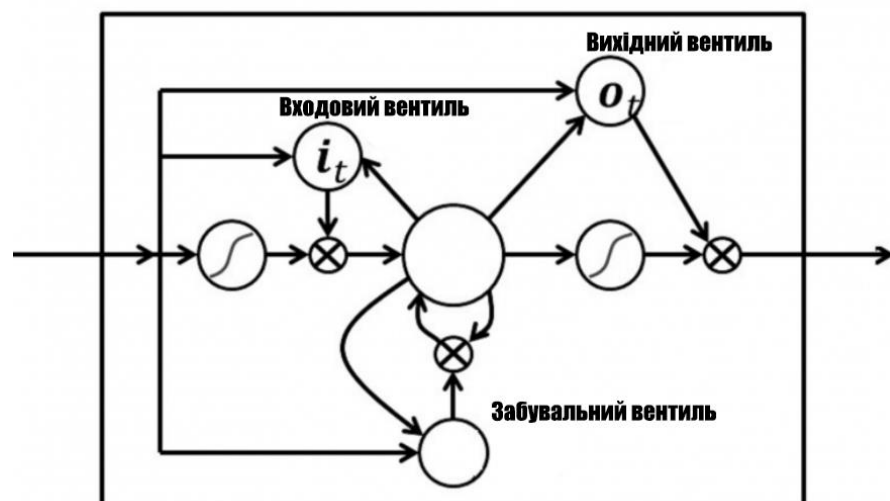
Алгоритм дуже добре працює для послідовних даних, таких як часові ряди, мова, текст, фінансові дані, аудіо, відео, погода тощо. RNN здатні сформувати набагато глибше розуміння послідовності та її контексту порівняно з іншими алгоритмами. Саме такі властивості роблять алгоритм надзвичайно ефективним для вирішення задачі прогнозування курсу валют. У RNN інформація проходить цикл. Приймаючи рішення, алгоритм враховує поточний вхід, а також те, що він навчився з вхідних даних, отриманих раніше.



Схематична демонстрація
роботи RNN

LSTM (long-short term memory) є RNN, які в основному розширюють гнучкість щодо пам'яті. Цей алгоритм добре підходить, щоб вчитися на важливих рішеннях, які мають дуже тривалі часові проміжки між собою. LSTM дозволяють RNN запам'ятовувати вхідні дані протягом тривалого періоду часу тому, що LSTM містять інформацію в пам'яті, так само, як пам'ять комп'ютера. LSTM може читати, записувати та видаляти інформацію зі своєї пам'яті.

У LSTM у вас є три вентиля: вхідний, забутий і вихідний. Ці вентиля визначають, чи дозволяти новий вхід (вхідний клапан), видаляти інформацію, оскільки вона не важлива (забувальний клапан), або дозволяти їй впливати на вихід на поточному часовому етапі (вихідний клапан).



У контексті прогнозування ціни криптоактивів послідовність дій для застосування LSTM може бути такою:

- Початкова обробка даних з біржі (нормалізація даних ціни у вигляді діапазону (0;1))
- Розділення даних на тренувальні та тестові (найбільш часте розділення відбувається у пропорції 80/20)
- Трансформація цих даних у формат часового ряду, в тому числі часовий крок

- Для LSTM часовий крок представляє кількість або довжину попередніх точок даних, включаючи поточну точку даних для передбачення
- Далі ми застосуємо функцію до нашого набору даних, включаючи визначений часовий крок
- Визначаємо кількість шарів LSTM
- Коли модель, шари, рівні активації, функції втрат визначені, ми можемо починати тренування моделі
- Визначивши кількість циклів тренування із всім датасетом та пропорцію тренувальних даних що беруться з датасету до обробки, можемо оцінити ефективність моделі, порівнявши функції втрат тренувальних та тестових даних
- Після знаходження оптимальних для моделі параметрів та подальшого тренування, в кінцевому результаті ми отримаємо графік ціни криптоактиву за вказаний період, що ми отримали з тренувальних даних; це і є той результат і механізм, який можна застосовувати для прогнозування на довшу перспективу

Баєсівська регресія

Насправді регресія по Баєсу є різновидом лінійної регресії. З точки зору Байєса, ми формулюємо лінійну регресію, використовуючи розподіли ймовірностей, а не точкові оцінки. Відповідь y не оцінюється як єдине значення, а передбачається, що витягується з розподілу ймовірностей. Вихід y генерується на основі нормального (гаусового) розподілу, що характеризується середнім значенням і дисперсією. Середнім для лінійної регресії є транспонування матриці ваг, помножене на матрицю предиктора. Дисперсія — це квадрат стандартного відхилення σ (помножений на матрицю Identity, оскільки це багатовимірне формулювання моделі). Метою байєсівської лінійної регресії є не знаходження єдиного «найкращого» значення параметрів моделі, а визначення апостеріорного розподілу (тобто умовного розподілу невідомої величини, обумовленим

певним твердженням) для параметрів моделі. Відповідь генерується не тільки з розподілу ймовірностей, але й передбачається, що параметри моделі також надходять із розподілу. Апостеріорна ймовірність параметрів моделі залежить від вхідних і вихідних даних навчання:

$$P(\beta|y, X) = \frac{P(y|\beta, X) * P(\beta|X)}{P(y|X)}$$

Тут $P(\beta|y, X)$ є апостеріорним розподілом ймовірностей параметрів моделі з урахуванням вхідних і вихідних даних. Це дорівнює ймовірності даних $P(y|\beta, X)$, помноженій на попередню ймовірність параметрів і поділеній на константу нормалізації. Це простий вираз теореми Байєса, основи байєсівського висновування.

У випадку, коли ми хочемо прогнозувати значення криптоактиву, дані для тренування моделі необхідно розбити на дві частини, одну з яких використовувати для навчання за класичною лінійною регресією, а другу як аргументи до рівняння регресії за Байєсом. Порівнявши результати та обчисливши Mean Squared Error, можна отримати варіації зміни курсу досліджуваного активу.

РОЗДІЛ 3: Практична частина

3.1 Аналіз проблеми

Стикаючись із криптоактивами, у користувача завжди постає питання того, чи можна, бодай з якоюсь точністю, прогнозувати майбутні коливання курсу. Завдання машинного навчання полягає саме у цьому, тому для візуального відтворення і зрозумілих результатів є сенс створити застосунок, який би допомагав користувачам зрозуміти, якою буде динаміка коливання курсу різних криптовалют за різний проміжок часу.

Саме завдяки вирішенню задачі регресії можна досягти конкретних результатів, що дозволять краще розумітися на динаміці крипторинку.

3.2 Опис запропонованого програмного продукту

Пропонується розробити застосунок, що обраховував би необхідні нам дані, а в якості демонстрації можна було б переглянути веб-сайт, на якому користувач може обрати цікаву йому криптовалюту, побачити дані щодо неї, модель лінійної регресії та прогноз щодо курсу; передбачити можливість обрання криптоактиву з випадного списку.

Задля кращого подальшого аналізу слід додати на сайт дані про криптовалюту за певний період, передбачивши можливість як табличного відображення, так і у вигляді графіка. Передбачається можливість масштабування графіку, щоб розглянути період детальніше.

Також пропонується подати дані у вигляді японських свічок для більшої наглядності. На сторінці нижче пропонується побудова моделі лінійної регресії, яка визначає загальний характер поведінки графіка.

Також приведений графік, що є демонстрацією навчання перцептронну, на основі тих даних, що обробили нейрони: наведено розділ даних на тренувальні і тестові, поведінка останніх на графіку, у свою чергу і є предметом дослідження проекту.

3.3 Огляд факторів, що підлягають аналізу, та математичних інструментів, що для нього застосовуються

Розглядаючи таку комплексну задачу як прогнозування курсу валюти, ми стикаємось із великою кількістю даних які треба обробити, а чинниками, що на них впливають. Також ми маємо справу із часовими рядами – сукупністю проіндексованих та відкладених на графіку у хронологічному порядку точках, стосовно яких задача полягає якраз у тому, щоб передбачити наступне положення точки. Для роботи з ними пропонується використати просту нейронну мережу, яку надає бібліотека Keras. Принцип полягає в тому, що ми порівнюємо часові ряди, які маємо в наявності, і модель згодом навчається самостійно розставляти необхідні точки на графіку.

Після того як ми змодельуємо наші дані та оцінимо навички нашої моделі на наборі навчальних даних, нам потрібно отримати уявлення про навички моделі на нових невидимих даних. Для звичайної задачі класифікації або регресії ми б зробили це за допомогою перехресної перевірки. Для даних часових рядів важлива послідовність значень. Простий метод, який ми можемо використати, — це розділити впорядкований набір даних на набори даних для навчання та тестування. Розділюємо дані на навчальні набори даних із 67% тренувальних спостережень, які ми можемо використовувати для навчання нашої моделі, залишаючи решту 33% для тестування моделі.

Також постає питання, яким чином вимірювати похибку моделі. Для цього пропонується використати метод MSE, середньоквадратичної похибки.

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

Маючи правильну відповідь і прогноз, склавши їх різниці, піднесені до квадрату, для всіх відповідей, та поділити на кількість елементів у вибірці, то ми отримаємо число, що характеризує ефективність роботи моделі.

Іншим важливим фактором є те, як ми будемо визначати точки для часових рядів. Пара ціна-дата не є вичерпною стосовно прогнозування курсу. Тому для аналізу також були взяті дані щодо обсяг торгів у самій валюті та доларі США, максимальної та мінімальної цін за день, ціни відкриття та закриття угоди.

3.4 Визначення засобів реалізації

Для реалізації бекенд-частини застосунку пропонується використати мову програмування Python, яка чудово підходить для роботи із статистичними даними та має широкий функціонал бібліотек, що дають змогу спростити велику кількість операцій щодо обробки даних та зекономити час.

Використовуються такі основні бібліотеки:

- ***Dash***

фреймворк з відкритим вихідним кодом для створення інтерфейсів візуалізації даних. Випущена в 2017 році як бібліотека Python, вона розширена, щоб включати реалізації для R і Julia. Dash допомагає науковцям з даних створювати аналітичні веб-додатки, не вимагаючи інтеграції більшої кількості елементів веб-розробки.

- ***Sklearn***

Scikit-learn — це безкоштовна програмна бібліотека машинного навчання для мови програмування Python, яка надає функціональність для створення та тренування різноманітних алгоритмів класифікації, регресії та кластеризації, таких як лінійна регресія, тощо.

- ***Plotly***

Бібліотека графіків Python від Plotly створює інтерактивні графіки високої якості. Користувач може легко створювати лінійні діаграми, діаграми розсіювання, діаграми площ, стовпчасті діаграми, стовпчики помилок,

квадратні діаграми, гістограми, теплові карти, підграфіки, багатоосі, полярні діаграми та бульбашкові діаграми.

- ***Numpy***

NumPy — це бібліотека Python, яка використовується для роботи з масивами та надає можливості зручного представлення та обробки даних. Вона також має функції для роботи в області лінійної алгебри, перетворення Фур'є та матриць. NumPy був створений у 2005 році Тревісом Оліфантом. Це проект з відкритим кодом, і ви можете використовувати його вільно.

NumPy розшифровується як Numerical Python.

- ***Keras***

Keras — це ефективний високорівневий інтерфейс програмування нейронних мереж, написаний на Python. Ця бібліотека нейронних мереж з відкритим кодом розроблена для швидкого експериментування з глибокими нейронними мережами, і вона може працювати поверх CNTK, TensorFlow і Theano.

Keras зосереджується на тому, щоб бути модульним, зручним і розширюваним. Він не обробляє низькорівневі обчислення; замість цього він передає їх до іншої бібліотеки під назвою Backend.

Keras був прийнятий і інтегрований в TensorFlow в середині 2017 року.

Користувачі можуть отримати доступ до нього через модуль `tf.keras`. Проте бібліотека Keras все ще може працювати окремо та незалежно.

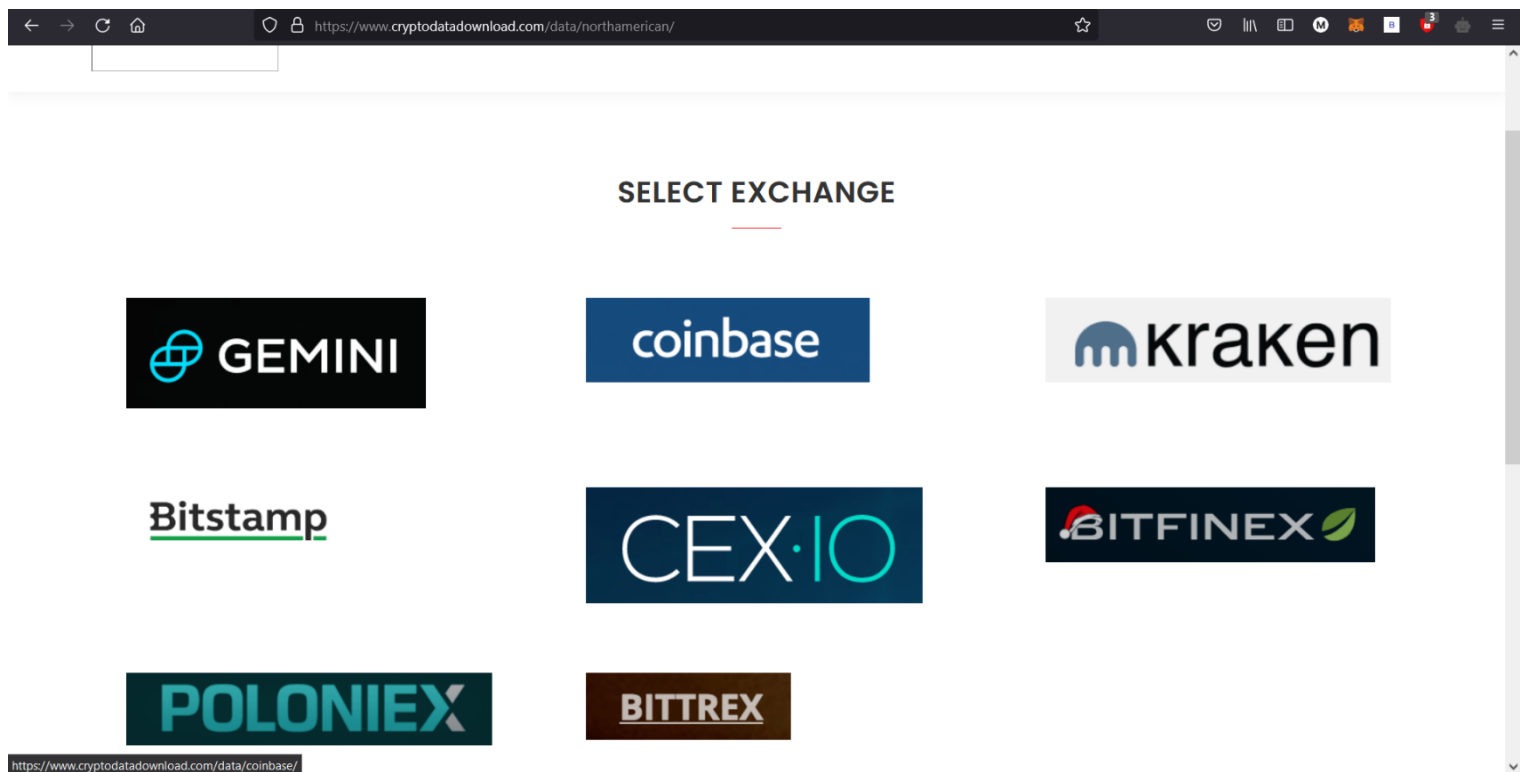
У свою чергу, TensorFlow — це платформа глибокого навчання з відкритим кодом, розроблена Google і випущена в 2015 році. Вона відома підтримкою документації та навчання, масштабованими параметрами виробництва та розгортання, кількома рівнями абстракції та підтримкою різних платформ, таких як Android.

TensorFlow — це символічна математична бібліотека, яка використовується для нейронних мереж і найкраще підходить для

програмування потоків даних у ряді завдань. Він пропонує кілька рівнів абстракції для побудови та навчання моделей.

Фронт-енд частина застосунку де-факто реалізовується за допомогою бібліотеки Dash, яка дозволяє вносити певні зміни на html-сторінку прямо з файлу Python.

Щоб отримати необхідні дані, у потрібному форматі були використані можливості сайту [cryptodatadownload.com](https://www.cryptodatadownload.com), що дозволяє завантажити дані за певний період у вигляді CSV-файлу. Дані були взяті з біржі Gemini. Дані для біткоїна взяті з жовтня 2015 по квітень 2022 року, а для Ethereum – з вересня 2016 по квітень 2022.



3.5 Опис програмної реалізації продукту та демонстрація роботи

Запускаємо застосунок за допомогою PyCharm, попередньо імпортувавши необхідні бібліотеки. Версія Python – 3.9. У нижній частині екрану бачимо, як

програма покроково обраховує значення похибки похибки для порівняння часових рядів.

```

118 elif value == 'ETH':
119     return dft.to_dict('records')
120
121 @app.callback(Output('data-graph', 'figure'), [Input('table-dropdown', 'value')])
122 def update_data_graph(value_dropdown):
123
124     if value_dropdown == 'BTC':
125         df_graph = df
126         text = 'Bitcoin'
127     elif value_dropdown == 'ETH':
128         df_graph = dft
129         text = 'Ethereum'
130
131     return {
132         'data': [go.Scatter(
133             mode='lines',
134             x=df_graph['date'],
135             y=df_graph['price'],
136             text=text,
137             name=text,
138             line=dict(color='red', width=2)
139         )]
140     }
141
142 update_data_graph0
  
```

Run: app x

Dash is running on <http://127.0.0.1:8050/>

- * Serving Flask app 'app' (lazy loading)
- * Environment: production
- WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
- * Debug mode: on

Epoch 1/60
801/801 - 1s - loss: 134523.4688 - 1s/epoch - 1ms/step

Epoch 2/60
801/801 - 1s - loss: 133606.9062 - 679ms/epoch - 848us/step

Epoch 3/60
801/801 - 1s - loss: 132303.8750 - 720ms/epoch - 899us/step

Отримавши у консолі посилання на адресу, на якій запустився сервер, переходимо за цим посиланням: <http://127.0.1.1:8050/>

Прогнозування курсу криптовалют

Вибрати

Bitcoin

Обрати криптовалюту

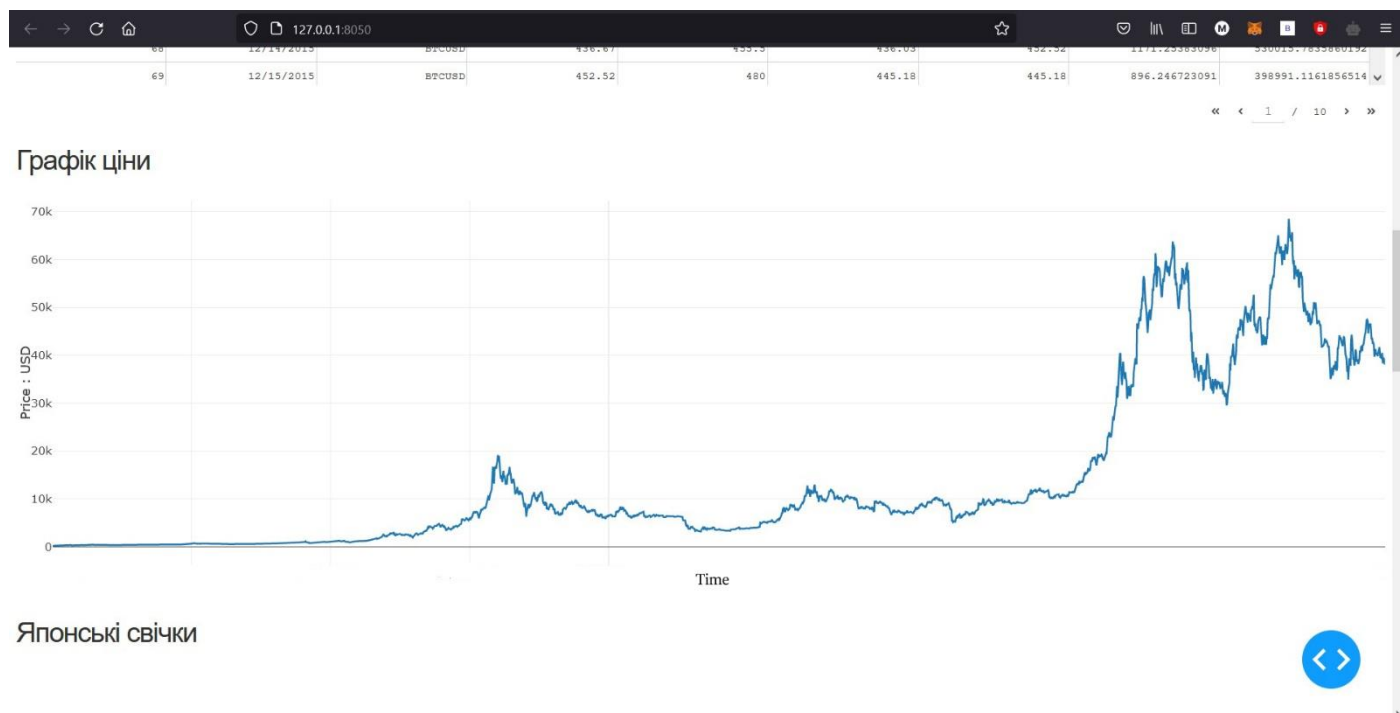
Щоденна статистика криптовалюти протягом останніх років, включаючи максимальну та мінімальну ціну за день, ціну відкриття, закриття позицій, тощо.

Дані були взяті з (<http://www.cryptodatadownload.com/data/northamerican/>). Статистику можна спостерігати у вигляді таблиці та графіку.

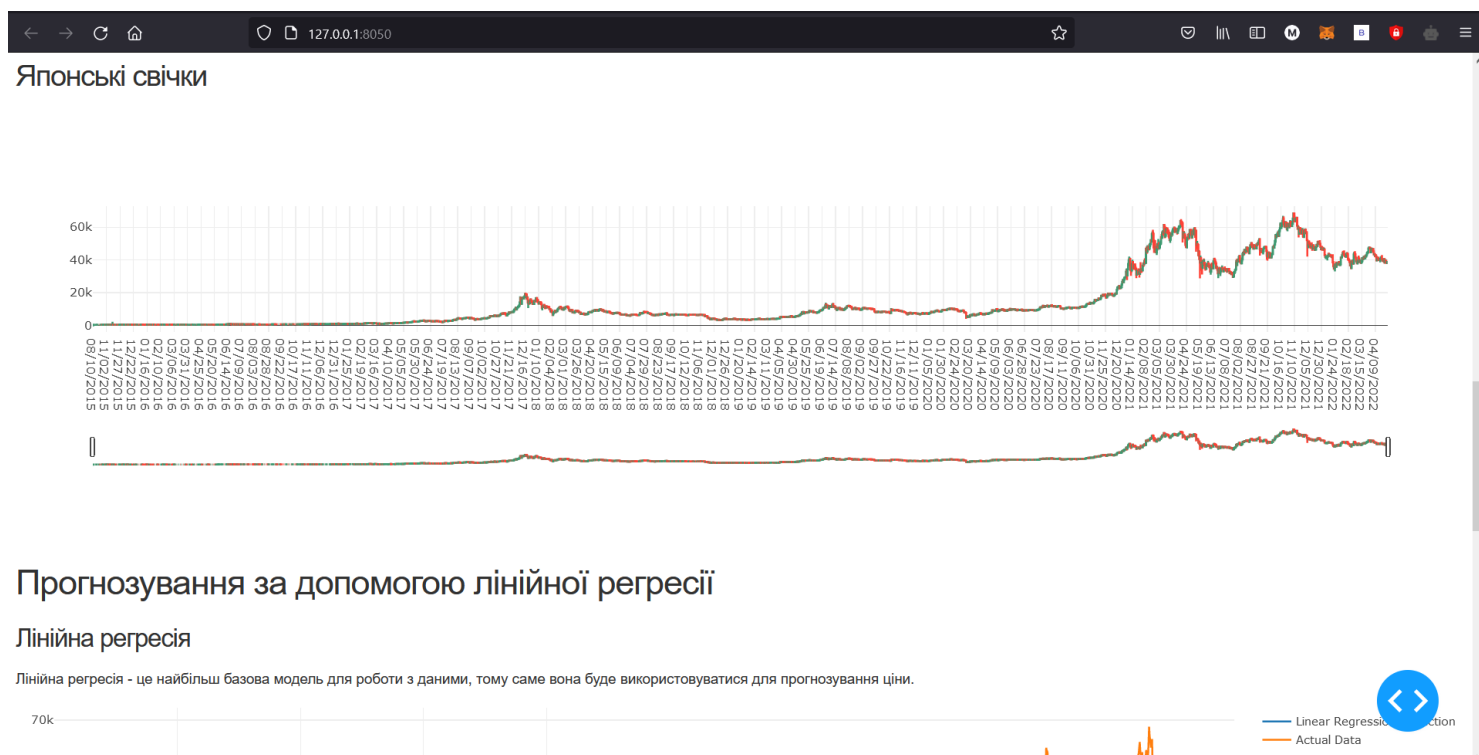
Таблиця ціни

Number	Date	Symbol	Open	High	Low	Close	Volume	Volume USD
54	11/30/2015	BTCUSD	373.87	380.99	367.53	377.32	1346.30344632	507987.2163654624
55	12/01/2015	BTCUSD	377.32	379.99	349	352.52	545.953654593	192459.58231712435
56	12/02/2015	BTCUSD	352.52	366.53	351.79	365.21	521.378642927	190412.6941833697
57	12/03/2015	BTCUSD	365.21	369.99	356	358.5	819.669910265	293851.6628300025
58	12/04/2015	BTCUSD	358.5	369.2	357.56	369.2	962.284415894	355275.4063480648
59	12/05/2015	BTCUSD	369.2	394.99	361.53	392.66	726.0781572310001	285101.8492183245
60	12/06/2015	BTCUSD	392.66	403.33	385.01	392.18	185.137172828	72607.09643968503
61	12/07/2015	BTCUSD	392.18	400	387.46	395	745.261792745	294378.408134275
62	12/08/2015	BTCUSD	395	422.2	390.84	422.2	608.643777036	256969.4026645992
63	12/09/2015	BTCUSD	422.2	424.48	347.5	415	662.142573311	274789.144143
64	12/10/2015	BTCUSD	415	421	412.54	420.79	698.0927540560001	293750.44143
65	12/11/2015	BTCUSD	420.79	473.34	420.79	467.3	922.040497224	430869.5243527752

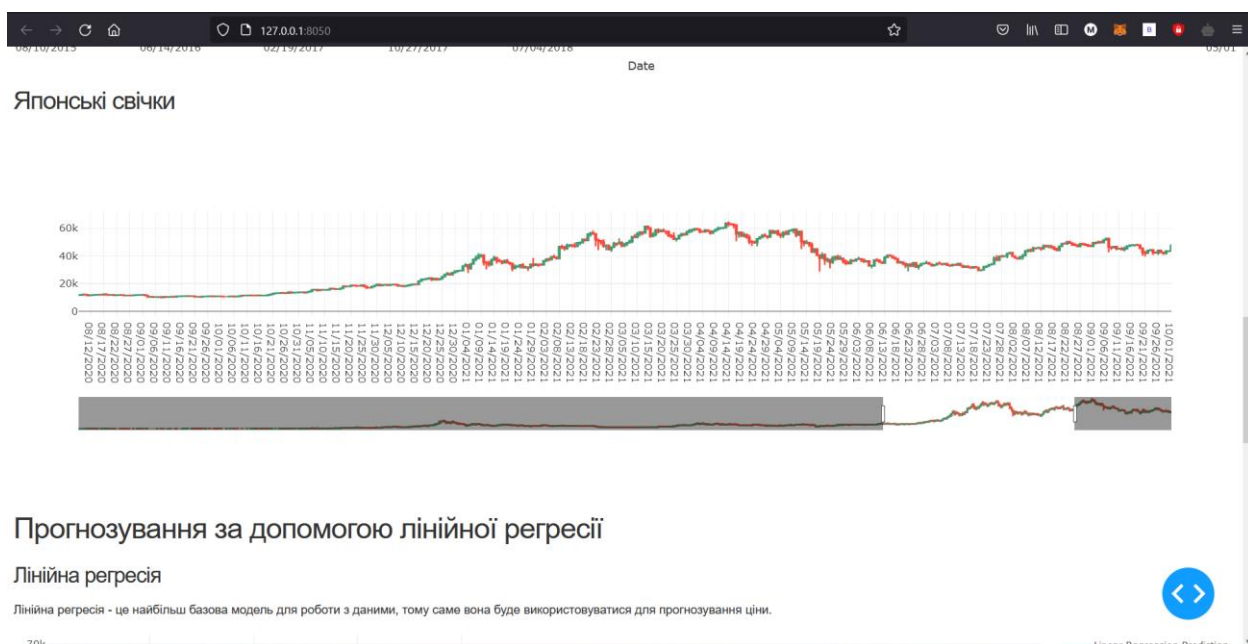
Можемо побачити обрану валюту та таблицю із ціною. Відповідно обираючи іншу валюту – користувач бачить зміну даних.



Наведено ті ж самі дані, але у вигляді графіку.



Для кращої візуалізації інформація дублюється у вигляді японських свічок. Також можна змінити масштаб відображення.



Нижче на сторінці наведено графік лінійної регресії.



У коді програми це виглядає наступним чином:


```

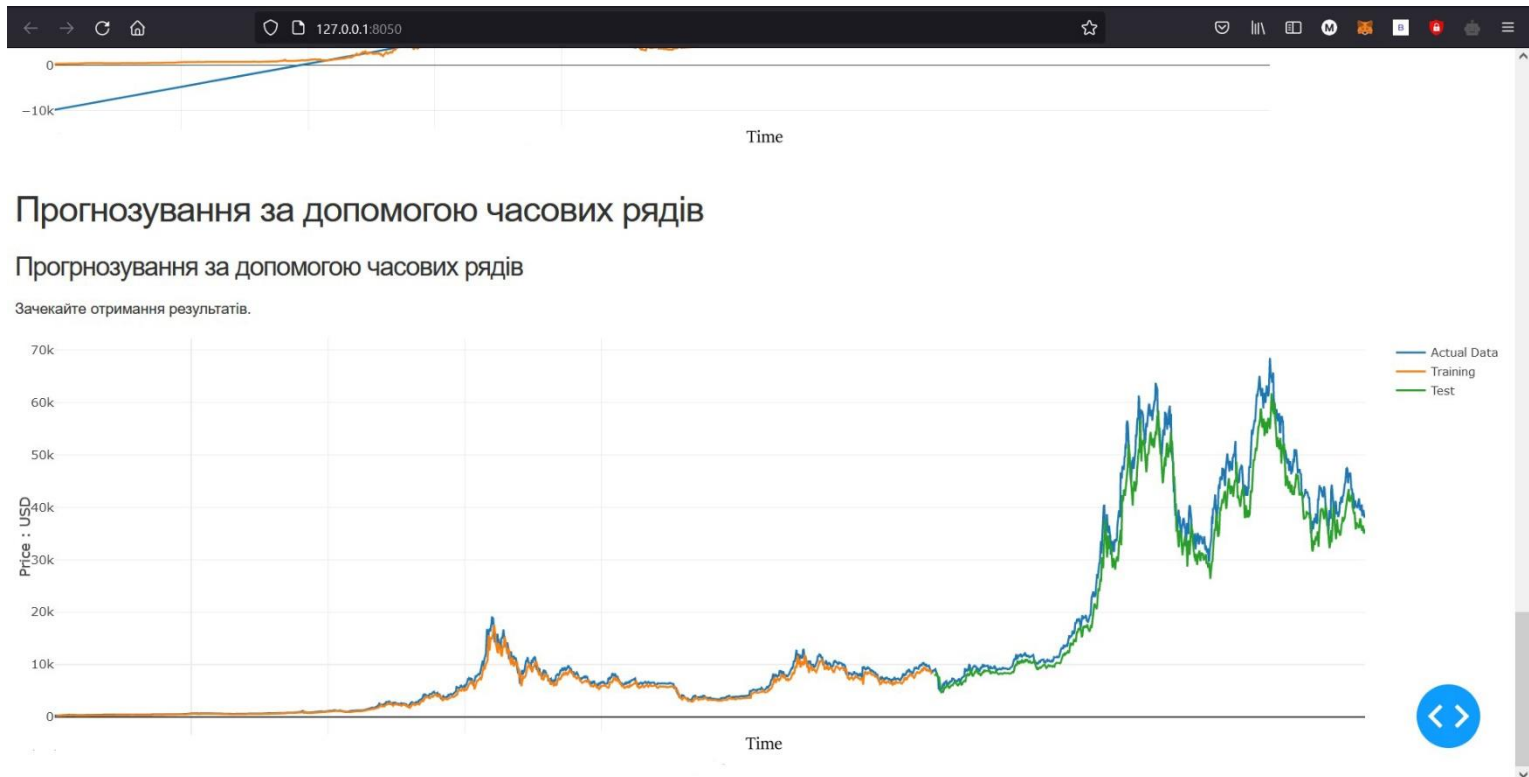
@app.callback(Output('graph', 'figure'), [Input('table-dropdown', 'value')])
def update_graph(value_dropdown):
    if value_dropdown == 'BTC':
        df_to_graph = df
        df_graph = df
        text = 'Bitcoin'
        X = df.iloc[:, 0].values.reshape(-1, 1) # значення перетворюються у масив чисел в numpy
        Y = df.iloc[:, 4].values.reshape(-1, 1)
    elif value_dropdown == 'ETH':
        df_to_graph = dft
        df_graph = dft
        text = 'Ethereum'
        X = dft.iloc[:, 0].values.reshape(-1, 1) # значення перетворюються у масив чисел в numpy
        Y = dft.iloc[:, 4].values.reshape(-1, 1)

    linear_regressor = LinearRegression() # створення об'єкту для класу
    linear_regressor.fit(X, Y) # обчислення показників регресії
    Y_pred = linear_regressor.predict(X) # прогнозування даних

    df_for_graph = pd.DataFrame(Y_pred, columns=['Prediction'])

```

Нижче наведені результати роботи графік прогнозування курсу за допомогою порівняння засобами бібліотеки Keras часових рядів.



Жовтим кольором показані дані для тренування моделі, синім – реальні дані, а зеленим – прогноз, який дає модель.

Код для цього графіку:

```
@app.callback(Output('graph-learn', 'figure'), [Input('table-dropdown', 'value')])
def update_graph_learn(value_dropdown):
    if value_dropdown == 'BTC':
        csv_name = 'Gemini_BTCUSD_daily.csv'
        df_to_learn = df
    elif value_dropdown == 'ETH':
        csv_name = 'Gemini_ETHUSD_daily.csv'
        df_to_learn = dft

    # фіксування випадкового числа для відтворюваності результатів
    numpy.random.seed(7)
    # завантаження даних
    dataframe = pd.read_csv(csv_name, usecols=[6], engine='python')
    dataset = dataframe.values
    dataset = dataset.astype('float32')

    # розділення на тренувальні і тестові дані
    train_size = int(len(dataset) * 0.67)
    test_size = len(dataset) - train_size
    train, test = dataset[0:train_size, :], dataset[train_size:len(dataset), :]

    # пристосування датасету до моделі
    look_back = 3
    trainX, trainY = create_dataset(train, look_back)
    testX, testY = create_dataset(test, look_back)

    # створення багатошарової моделі порівняння
    model = Sequential()
    model.add(Dense(1, input_dim=look_back, activation='relu'))
    # генерування прогнозів для тренувальних даних
    trainPredict = model.predict(trainX)
    testPredict = model.predict(testX)

    model.compile(loss='mean_squared_error', optimizer='adam')
    model.fit(trainX, trainY, epochs=60, batch_size=2, verbose=2)

    # Оцінка роботи моделі за допомогою MSE
    trainScore = model.evaluate(trainX, trainY, verbose=0)
    print('Train Score: %.2f MSE (%.2f RMSE)' % (trainScore, math.sqrt(trainScore)))
    testScore = model.evaluate(testX, testY, verbose=0)
    print('Test Score: %.2f MSE (%.2f RMSE)' % (testScore, math.sqrt(testScore)))

    # зміщення тренувальних прогнозів для побудови графіку
    trainPredictPlot = numpy.empty_like(dataset)
```

Обчислення ефективності моделі та прогнозування курсу відбувається за допомогою класу `Sequential()` бібліотеки `Keras`, вона представляє собою лінійний стек шарів. Вказавши для моделі форму очікуваного вводу та кількість шарів, які вона буде обробляти, можна навчити її використовуючи дані з масиву `Numpy` (як це роблять усі моделі `Keras`), для цього використовується функція `fit()`. Навчаючи модель порівнювати часові ряди, на основі цих даних ми і будемо подальші прогнози.

Висновки

У цій роботі було проаналізовано задачу, яка має широке прикладне значення: прогнозування курсу криптовалют. Були досліджені основні криптоактиви, на основі аналізу яких можна розглядати ситуацію на крипторинку в цілому, а також екосистеми застосунків, що використовують цифрові гроші для розрахунку всередині своїх сервісів. Також були описані декілька найбільших криптобірж, а також ті принципи, на основі яких вони взаємодіють із користувачем.

Були розглянуті основні закони торгівлі на крипторинку, зв'язок його аналізу із математикою та статистикою, описані принципи, згідно з якими може коливатися ціна, а також які чинники прямо чи опосередковано на неї впливають.

Також було наведено основні підходи та алгоритми, на яких базується аналіз криптоактивів з точки зору машинного та глибокого навчання, а також як для вирішення цієї задачі використовують регресію та нейронні мережі.

У практичній частині роботи було створено застосунок, що має користувацький інтерфейс у вигляді веб-сторінки та дозволяє подивитись на графіки обраної криптовалюти, а також візуалізує лінійну регресію для заданого набору даних.

Також за допомогою бібліотеки для машинного навчання Keras було створено багатошарову модель, що дозволяє порівнювати дані часових рядів та прогнозувати курс криптоактиву. Попередньо розбиваючи дані на навчальні і тестові, користувач отримує оцінку моделі на вказаний період.

Список використаних джерел

1. G. Bresler, G. H. Chen, and D. Shah, "A latent source model for online collaborative filtering," in Advances in Neural Information Processing Systems: теоретико – прикладне дослідження, 2014. 66 с.
2. What is bitcoin? [Електронний ресурс]: Money CNN, стан на 15 квітня 2022 – Режим доступу: <http://money.cnn.com/infographic/technology/what-is-bitcoin/> - Назва з екрану.
3. Blockchain explained [Електронний ресурс]: Investopedia, стан на 20 квітня 2022 – Режим доступу: <https://www.investopedia.com/terms/b/blockchain.asp> - Назва з екрану.
4. R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society. Series B (Methodological). 1996. Vol. 58, №1, с. 267.
5. Andrew W. Lo & A. Craig MacKinlay, 1987. "Stock Market Prices Do Not Follow Random Walks: Evidence From a Simple Specification Test," NBER Working Papers 2168, National Bureau of Economic Research, Inc. NJ: Princeton University Press, 1999. 50 с.
6. What is Linear Regression [Електронний ресурс]: Complete Dissertation, стан на 16 квітня 2022 – Режим доступу: <https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/what-is-linear-regression/> - Назва з екрану.
7. Simple Linear Regression [Електронний ресурс] : Scribbr, стан 20 квітня 2021 - Режим доступу: <https://www.scribbr.com/statistics/simple-linear-regression/> - Назва з екрану.
8. The Complete Guide to Time Series Analysis and Forecasting [Електронний ресурс] : Marko Peixeiro, Towards Data Science, стан 22 квітня 2022 – Режим доступу: <https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775> - Назва з екрану.
9. Long Short-Term Memory Networks [Електронний ресурс] Math works, стан на 21 квітня 2022 – Режим доступу: <https://www.mathworks.com/help/deeplearning/ug/long-short-term-memory-networks.html;jsessionid=332c404a0fd4687f02ad7a309e3c> - Назва з екрану.
10. Introduction To Long Short-Term Memory (LSTM) [Електронний ресурс]: Analytics Vidhya, стан 30 квітня 2021 - Постанова - Режим доступу: <https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/> - Назва з екрану.
11. LSTM Recurrent Neural Networks — How to Teach a Network to Remember the Past [Електронний ресурс]: Saul Dobilas, Towards Data Science, стан на 1 травня 2022 – Режим доступу: <https://towardsdatascience.com/lstm-recurrent-neural-networks-how-to-teach-a-network-to-remember-the-past-55e54c2ff22e> - Назва з екрану.
12. New ACM Journal on Distributed Ledgers to Explore Latest Trends in Blockchain, Cryptocurrencies and Related Technologies [Електронний ресурс]: ACM, the Association for Computing Machinery: інтернет-журн. – 2022. – Режим доступу: <https://www.acm.org/media-center/2021/september/dlt-cfp> - Назва з екрану.
13. Sparse PCA from Sparse Linear Regression G. Bresler, S.M. Park, M. Persu in Advances in Neural Information Processing Systems: теоретико-прикладне дослідження, 2018. 4 с.

Додатки

Додаток А. Software Design Description застосунку для прогнозування курсу

Застосунок для роботи прогнозом курсу криптоактивів

1. Вступ

Машинне навчання є важливим інструментом у аналізі даних, тому його використання у прогнозуванні курсу криптоактивів якнайкраще підходить під саме цю задачу. Розглядаючи цю проблему як задачу регресії можна побудувати модель, яка на основі статистики за певний період зможе спрогнозувати поведінку показників на іншому часовому проміжку.

1.1 Мета

Провести дослідження, яке б доводило ефективність застосування лінійної регресії, а також нейронних мереж у контексті криптовалют та прогнозування їхнього курсу.

1.2 Предметна область

Крипторинок – сукупність торгівельних майданчиків, на яких проходять торги криптовалютами. Саме він стане предметом дослідження.

1.3 Загальна ідея

Застосувати методи машинного і глибокого навчання, щоб навчити модель, використовуючи наявні дані, а потім на основі цих даних побудувати графік курсу цікавого нам криптоактиву за вказаний період.

2. Загальний огляд застосунку

2.1 Основні проблеми, вирішення яких є на меті проекту

- Необхідно з усіх методів машинного навчання обрати саме ті, які б у повному обсязі підходили до поставленого завдання і водночас зберігали свою доцільність та ефективність
- Зібрати необхідні дані за вказаний період, щоб проаналізувати їх. Уникнути помилок у обробці даних, похибок у виді викидів значень, тощо
- Оптимізувати обробку даних таким чином, щоб велика кількість циклів навчання моделі не перевантажувала IDE і пам'ять комп'ютера

2.2 Запропоновані рішення

- Вирішено сформулювати задачу проекту як вирішення задачі регресії. Виконуючи цю задачу, після моделювання необхідних нам даних на основі показників-предикторів та оцінки навичок нашої моделі на наборі навчальних даних, нам потрібно отримати уявлення про навички моделі на нових даних, які ще не оброблялись. Після необхідного розподілу даних на тренувальні і тестові, необхідно провести обробку тренувальних за допомогою нашої моделі.
- Усі дані збираються з сайту <https://www.cryptodatadownload.com/data/northamerican/>, що підтверджує їх коректність та відсутність помилок, адже це сертифікований сервіс, що збирає дані з восьми різних бірж і подає їх у зручному для обробки .csv форматі
- Для досліджень буде використана бібліотека Keras, що дозволяє швидко працювати з мережами глибинного навчання майже будь якої глибини. Модульність дозволить ефективно тестувати проект, а бібліотека TensorFlow допоможе у обробці даних

3. Реалізація

- Пропонується розробити застосунок, роль користувачького інтерфейсу в якому відігравала би веб-сторінка, на якій користувач може обрати цікаву йому криптовалюту, побачити дані щодо неї, модель лінійної регресії та прогноз щодо курсу
- На сайті наявне обрання криптоактиву з випадного списку
- Нижче приведені дані за вказаний період у вигляді таблиці
- Також наведені дані у вигляді графіку
- Передбачається можливість масштабування графіку, щоб розглянути період детальніше
- Також пропонується подати дані у вигляді японських свічок для більшої наглядності
- На сайті нижче побудована модель лінійної регресії, яка визначає загальний характер поведінки графіка
- Також приведений графік, що є демонстрацією навчання перцептронну, на основі тих даних, що обробили нейрони: наведено розділ даних на тренувальні і тестові, поведінка останніх на графіку, у свою чергу і є предметом дослідження проекту.

4. Посилання

The Complete Guide to Time Series Analysis and Forecasting [Електронний ресурс] : Marko Peixeiro, Towards Data Science, стан 22 квітня 2022 – Режим доступу: <https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775> - Назва з екрану.

Long Short-Term Memory Networks [Електронний ресурс] Math works, стан на 21 квітня 2022 – Режим доступу: <https://www.mathworks.com/help/deeplearning/ug/long-short-term-memory-networks.html;jsessionid=332c404a0fd4687f02ad7a309e3c> - Назва з екрану.

Introduction To Long Short-Term Memory (LSTM) [Електронний ресурс]: Analytics Vidhya, стан 30 квітня 2021 - Постанова - Режим доступу: <https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/> - Назва з екрану.