

Сітайло М.С.

Державний університет «Житомирська політехніка»

ДОСЛІДЖЕННЯ ПРОЦЕСУ ЗБОРУ ДАНИХ З МЕТОЮ ЇХ ВИКОРИСТАННЯ ПРИ ПРОГНОЗУВАННІ КУРСУ КРИПТОВАЛЮТ

Засоби на основі штучного інтелекту продовжують стрімко здобувати популярність у різноманітних галузях та знаходити варіанти використання при вирішенні певних практичних завдань. Незважаючи на широкий спектр задач, які можуть розв'язуватися згаданим інструментом, процес розробки нейронних мереж складається з визначених етапів, які є подібними для майже усіх задач. Збір даних є важливим етапом при розробці рішення з використанням штучного інтелекту. Об'єм і варіативність створеного датасету мають значний вплив на ефективність розробленої нейронної мережі. Хоч процес збору даних складається з певних визначених етапів, особливості поставленого завдання часто вимагають додаткових операцій. Прогнозування вартості активів є актуальним завданням впродовж значного періоду часу. Це сприяє активному пошуку різноманітних методів, які здатні покращити результати прогнозування. Хоч дане завдання є досить подібним до передбачення ціни інших видів активів, особливості технології блокчейн зумовлюють певні особливості під час його вирішення. Доступ до даних блокчейну дозволяє отримувати інформацію про усі операції з певними криптовалютами або токенами, що дозволяє враховувати ці фактори при проведенні фундаментального аналізу.

Дана стаття містить аналіз актуальних досліджень, які стосуються аналізу різних видів даних, що можуть бути використані при прогнозуванні котирувань криптоактивів. На основі проведеного аналізу було визначено особливості роботи з основними категоріями джерел даних, які застосовуються для технічного, фундаментального та соціально-економічного аналізів. Також були проведені опис та проектування процесу збору та обробки даних з криптобірж, блокчейнів та соціальних мереж. Проведений аналіз дозволив провести практичне дослідження процесу збору даних для прогнозування вартості криптовалют. Це полягло у ознайомленні з різноманітними вебресурсами або програмними засобами, які можна застосовувати при вирішенні поставленого завдання. На основі отриманих результатів, було проведено опис процесу взаємодії з криптобіржами, блокчейнами та даними соціальних мереж з метою збору усіх розглянутих категорій даних. Цей процес базується на використанні прикладних програмних інтерфейсів, бібліотек для взаємодії з блокчейнами та бібліотек для аналізу природних мов для оцінки емоційного забарвлення повідомлень у певних соціальних мережах.

Ключові слова: збір та обробка даних, криптовалюти, набір даних, соціально-економічний аналіз, технічний аналіз, фундаментальний аналіз.

Постановка проблеми. Актуальність проблеми прогнозування вартості активів, особливо криптовалют, зумовлює широке використання нейронних мереж для вирішення цього завдання. Проте, ефективність створених мереж тісно пов'язана з наборами даних, які використовуються для їх навчання. Це сприяє широкому дослідженню різноманітних джерел та видів інформація, яка може бути використана для більш ефективного передбачення майбутніх котирувань криптовалют.

Незважаючи на велику кількість наукових робіт присвячених дослідженню різних видів даних, налагодження процесу збору та обробки даних вирішується особисто науковцем, який збирається їх використовувати для навчання власної

нейронної мережі. Так як при прогнозуванні вартості криптовалют зазвичай використовуються досить різноманітні за структурою та особливостями джерела даних, важливим є дослідження процесу реалізації збору та обробки цієї інформації з метою її систематизації та узагальнення.

Аналіз останніх досліджень і публікацій. Дослідження методів прогнозування курсу криптовалют та підходів для їх покращення можна умовно поділити на 2 категорії:

- 1) дослідження покращення нейронних мереж за допомогою яких відбувається прогнозування;
- 2) аналіз різних джерел даних та оцінка їх впливу на котирування криптовалют.

На даний момент є велика кількість архітектур нейронних мереж, які застосовуються для

вирішення даного завдання, проте усі вони використовують певні набори даних при навчанні та прогнозуванні. Отже, удосконалення наборів вхідних даних має позитивний вплив на точність результатів прогнозування для майже усіх видів архітектур мереж.

Задача прогнозування вартості активів завжди була актуальною. Це сприяло пошуку методів та інструментів для прогнозування їх вартості. Так як кожний вид активу має свої особливості, задача прогнозування майбутньої ціни розглядається з їх врахуванням. Якщо систематизувати різні види аналізу вартості активів можна виділити 3 групи:

1. Технічний аналіз.
2. Фундаментальний аналіз.
3. Соціально-економічний аналіз.

Зазвичай передбачення курсу криптовалют базується на використанні технічного методу. Він включає в себе аналіз графіків ціни, попиту та індикаторів з використанням відповідних методів технічного аналізу. Значно рідше застосовується фундаментальний аналіз, що насамперед спричинене відсутністю певних фундаментальних індикаторів. Це зумовлене тим, що часто криптоактиви не прив'язані до певних пасивів та не приносять дохід як компанії, отже складно визначити фундаментальну цінність кожного з них. Проте, даний вид цифрових активів має певні характеристики, які можна враховувати при передбаченні їх вартості, а саме дані про їх використання у мережі блокчейн. Так як модель даних цих валют побудована на основі попиту й пропозиції, дані характеристики можуть бути дуже важливими у передбаченні зміни популярності певного активу, що часто корелює з його ціною.

Окрім згаданих типів аналізу криптовалют треба враховувати й безліч сторонніх факторів. Як було зазначено, поточна ціна криптовалют формується відношенням між попитом та пропозицією. Саме ця особливість підкреслює важливість циклів економіки, відношення суспільства до даного виду активів та інших соціально-економічних факторів під час створення прогнозу для даних активів.

Існує велика кількість досліджень, які як розглядають кожний з видів аналізу окремо, так і об'єднують різні методи між собою.

Найбільш поширеним методом прогнозування вартості криптовалют є передбачення котирувань на основі історичних даних зміни ціни. П. Моді та К. Арші у своїй роботі [1] дослідили прогнозування вартості Bitcoin використовуючи історичні дані зміни ціни на графіку Відкриття-Макси-

муму-Мінімуму-Закриття (Open-High-Low-Close, OHLC) графіку. Цей графік дозволяє отримати інформацію про зміну котирувань активу за певний інтервал часу, а саме: ціну відкриття, ціну закриття, найвищу ціну та найнижчу ціну. Публікація Ш. Лю та К. Ву [2] також базується на використанні історичних даних, проте вони додатково обробляють часові ряди за допомогою визначення тренду і сезонності даних, а також застосування функції автокореляції. Ж. Ву і Х. Чанг у своїй роботі [3] досліджували та порівнювали різні архітектури нейронних мереж, проте частина їх праці була виділена на дослідження впливу Covid-19 на зміну динаміки та значень котирувань основних криптоактивів. Публікація П. Азада та С. Аккора [4] в свою чергу містить аналіз застосування нейронних мереж для аналізу інформації блокчейну. Це є одним із методів отримання фундаментальних даних для прогнозування вартості досліджуваного виду цифрових активів. Наукову роботу Д. Аракі [5] можна віднести до 3 типу аналізу, а саме використання соціально-економічних даних. Дана робота описує та перевіряє ефективність NLP мережі для класифікації економічної інформації та новин, що згодом дозволяє застосовувати отримані результати при передбаченні ціни криптовалют. Праці Ж. Харіса і Н. Сахана [6] та С. Раю і А. Таріфа [7] використовують різні нейронні мережі для прогнозування, проте обидві ці роботи базуються на аналізі емоційного забарвлення твітів й застосуванні отриманої інформації при прогнозуванні майбутньої ціни певних криптовалют. В. Гургул та С. Лесман у публікації [8] провели прогнозування котирувань криптовалютів з використанням всіх згаданих видів аналізу, тобто технічного, фундаментального та соціально-економічного. Окрім аналізу новин та соціальних мереж, науковці досліджують вплив інших економічних факторів на ціну криптоактивів. Наприклад, робота С. Сінха і М. Бата [9] описує взаємозв'язок цін різних криптовалют між собою. Також важливими є дослідження, які направлені на визначення певних особливостей та характеристик, які властиві даному виду активів. А. Зекії і Ф. Амруш у своїй науковій праці [10] дослідили взаємозв'язок між параметрами різних мереж блокчейнів та надійністю відповідних токенів.

Постановка завдання. Метою статті є дослідження різних видів даних, які застосовуються для прогнозування вартості криптовалют та процесу їх збору та обробки. Це відбувається шляхом аналізу особливостей взаємодії з криптобіржами,

блокчейнами, а також певними соціальними мережами. В результаті проведеного аналізу необхідно сформулювати та описати особливості основних підходів для отримання та обробки інформації зі згаданих джерел.

Отримані висновки мають дозволити визначити особливості роботи з технічними, фундаментальними та соціально-економічними даними, а також систематизувати методи та засоби для створення датасету на основі цих видів інформації, що буде використовуватися для прогнозування вартості криптовалют.

Виклад основного матеріалу. Початковим етапом навчання майже усіх нейронних мереж є збір та обробка даних. Необхідний набір вхідної інформації залежить від поставленого завдання та цілей, які намагаються досягти. Основною метою прогнозування криптовалют є генерація найбільш точних прогнозів майбутніх котирувань на основі попередніх часових рядів. Остаточний набір даних залежить від методів аналізу, які будуть використовуватися при прогнозуванні ціни криптоактивів. На основі проведеного аналізу наукових досліджень та публікацій було визначено наступні підходи:

- технічний аналіз;
- фундаментальний аналіз;
- аналіз соціально-економічних факторів.

Технічний аналіз криптовалют базується на дослідженні історичних даних зміни вартості певної торгової пари. Зазвичай торговельні операції з криптовалютами відбуваються на відповідних платформах – криптобіржах. Так як поточна ціна криптоактиву формується шляхом відношення між попитом та пропозицією, ціна криптовалюти може відрізнитися в залежності від обраної біржі. Зазвичай ціни на різних біржах є приблизно однаковими, бо вони часто зрівнюються у зв'язку з арбітражем між ними. Для збору фундаментальних даних було обрано платформу Binance. Дана криптовалютна біржа є дуже популярною. Її користувачі проводять велику кількість торгових операцій, що робить її однією з лідируючих у світі за об'ємом.

Зазвичай для прогнозування вартості активів використовується графік OHLC. Він надає більш детальну інформацію про зміну ціни за певний проміжок часу у порівнянні з лінійним графіком. Інформація на цьому графіку представлена у вигляді «свічок». На основі «свічки» можна визначити ціну з якої почався період, ціна на якій закінчився період, а також мінімальну та максимальну позначки яких досягала ціна за дослі-

джуваний часовий проміжок. Відповідно до проведеного аналізу актуальних досліджень можна констатувати використання усіх 4 параметрів при прогнозуванні ціни криптовалют з використанням технічного аналізу. Врахування значень усіх згаданих цін показує вищу точну прогнозів на 10–15 % (в залежності від обраної архітектури мережі) при порівнянні з нейронними мережами, які враховують лише ціну відкриття або закриття.

Отже, для проведення ефективного технічного аналізу необхідно отримати набори даних, які містять інформацію у форматі OHLC графіку, тобто мають значення усіх 4 параметрів. Згадана біржа Binance надає потрібну інформацію, а доступ до неї відбувається за допомогою API, що дозволяє швидко та просто отримувати дані з сервісу Binance. Окрім цього, даний вебсервіс безкоштовно надає доступ до великого часового проміжку історичних даних та дозволяє отримувати оновлення інформації у режимі реального часу за допомогою вебсокетів.

Взаємодія зі згаданим вебсервісом може бути реалізована за допомогою модулю `python-binance`. Даний модуль є обгорткою для офіційного REST API Binance й дозволяє додаткам отримати доступ до історичних даних та додаткових функцій, які надаються користувачам криптобіржі.

У результаті використання цього модулю можна завантажити дані про зміну ціни за обрані проміжки часу. Ця інформація може одразу використовуватися у нейронних мережах, проте іноді вона додатково трансформується. На основі отриманих даних можна провести розрахунок різних технічних індикаторів, що застосовуються у трейдингу, наприклад: індекс відносної сили (Relative Strength Index, RSI), ковзна середня (Moving Average, MA), середній дійсний діапазон (Average True Range, ATR). Це дозволяє використовувати значення обчислених індикаторів, що зазвичай базуються на інструментах статистики, разом з інформацією про зміну ціни як вхідні дані нейронних мереж. Відповідно, зібрана та обчислена інформація буде використовуватися при навчанні нейронної мережі й генерації прогнозів щодо майбутніх котирувань криптоактивів.

Аналіз фундаментальних факторів є дуже популярним підходом при дослідженні таких інвестиційних інструментів як акції та облігації компаній. В основу даного виду оцінки вартості активів покладено вивчення та використання інформації, яка містить дані про результати їх операційної діяльності, метрики прибутковості, тощо. Криптоактиви відрізняються від згаданих видів активів

й не можуть аналізуватися з використанням фундаментальних метрик через свої певні особливості. Проте, можна певним чином видозмінити метод фундаментального аналізу й застосувати його для криптовалют.

Усі криптовалюти або токени базуються на технології блокчейн. Ця технологія являє собою децентралізований цифровий реєстр транзакцій, що зберігається протягом тривалого часу і захищений певними методами шифрування. Дані цього реєстру розподіляються за допомогою мережі комп'ютерів. Користувачі цієї технології можуть безпосередньо взаємодіяти зі збереженими даними у режимі реального часу, не потребуючи посередника для підтвердження справжності транзакцій.

Зміни у блокчейн вносяться у вигляді транзакцій. Усі нові транзакції об'єднуються у блоки, після чого додаються до мережі шляхом приєднання блоку до останнього блоку-елемента ланцюга. Так як усі дані є загальнодоступними їх можна переглядати у вільному доступі.

Наступним важливим етапом є розуміння того як дані з блокчейну можуть бути використані для прогнозування вартості криптовалют. Для цього розглянемо приклад, де проведемо аналогію між блокчейном та банком. Чим більше у банку клієнтів, тим більше він починає заробляти на кредитах та комісіях з транзакцій, збільшуючи таким чином свій прибуток. Кожний блокчейн у свою чергу має певну кількість активних користувачів. Чим більше користувачів у мережі, тим більше власників монет, що позитивно впливає на її ринкову капіталізацію. Зміна вартості мережі прямопро-

порційно впливає на вартість криптовалюти, яка є нативною для цього блокчейну.

У наш час існує велика кількість вебсервісів, які дозволяють зручно переглядати блоки, транзакції та іншу інформацію про згадані системи (рис. 1). Ці сервіси є дуже зручними та корисними, проте вони є обмеженими у можливостях отримати дані за великі проміжки часу. Навіть якщо подібний вебресурс надає таку можливість, отримання цих даних вимагає придбання платного доступу.

Вирішити цю проблему дозволяє одна з основних переваг технології блокчейн, а саме – децентралізованість. Ця особливість реалізується можливістю отримати доступ до мережі будь-кому хто має підключення до мережі Інтернет. Отже, мережі блокчейн дозволяють отримувати інформацію про усі зміни, які відбуваються в мережі. Програмна реалізація збору даних може відрізнятися для кожного окремого блокчейну, проте загальний концепт є ідентичним. Розглянемо особливості створення програмного модулю для отримання даних з наступних мереж: Bitcoin, Ethereum та Binance Smart Chain.

Для збору даних мережі Bitcoin можна використати вебсервіс Blockchain Info, який індексує дані усього блокчейну й надає можливість їх отримати за допомогою API. Даний інструмент забезпечує користувачам доступ до блокчейну та виконання пошуку і навігації по його даних. Використання цього вебсервісу значно спрощує завантаження інформації про блоки мережі Blockchain та збір статистики на основі отриманих даних.

За замовчуванням отримання даних з блокчейнів Ethereum та Binance Smart Chain відбувається

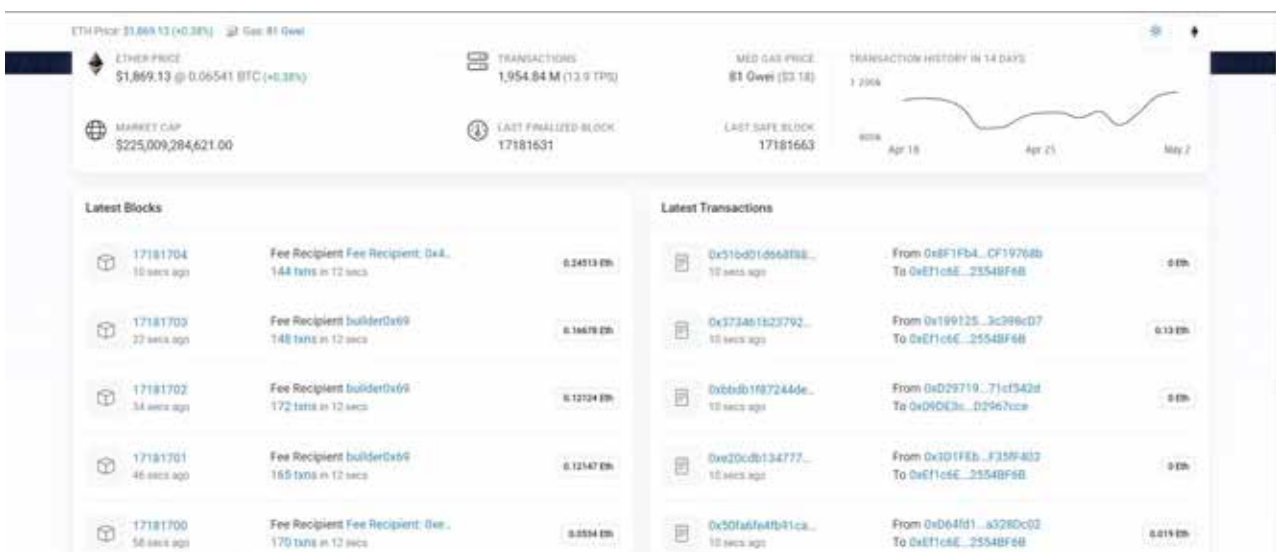


Рис. 1. Перегляд інформації про блокчейн Ethereum за допомогою вебресурсу Etherscan

за допомогою взаємодії з відповідним блокчейном через протокол RPC. Проте, розвиток технології Web3 сприяв використанню особливостей та механізмів даної технології у великій кількості програмних продуктів, що спричинило створення модулю web3. Загалом Web3 представляє собою наступне покоління технології Інтернет, що характеризується децентралізацією як інфраструктурних елементів, так і додатків. Ця інновація надає користувачам більший контроль над їх даними та взаємодією у мережі. Web3 націлена на забезпечення більш децентралізованого та орієнтованого на користувача Інтернету, що сприятиме покращенню безпеки, приватності та інклюзивності при взаємодії у мережі.

Для комунікації з мережею Ethereum необхідно використовувати сервіс Infura (або будь-який інший подібний), що дозволяє отримувати інформацію без розгортання блокчейну на локальному комп'ютері. Даний сервіс забезпечує підтримку вузлів різних мереж і надає користувачам доступ до даних блокчейну, перегляду транзакцій та взаємодії з смарт-контрактами через API. Окрім цього, даний засіб надає доступ до децентралізованої мережі, що дозволяє взаємодіяти з нею за допомогою згаданого модулю web3. На відміну від мережі Ethereum, Binance Smart Chain спрощує процес доступу надаючи відкриті сервіси для взаємодії з мережею. Це позбавляє розробників необхідності розгортати копію блокчейну локально або шукати сервіси, які будуть це робити за них.

Доступ до блокчейну дозволяє зчитувати інформацію про блоки та транзакції, які містяться в цих блоках. Кожна транзакція має певні параметри, що можуть бути використані для проведення аналізу, наприклад: адреса відправника, адреса отримувача, сума переказу, сплачена комісія, тощо. Незважаючи на те, що параметри є заздалегідь визначеними, дослідник може визначити та обчислити певні власні метрики, що можуть бути ефективними при вирішенні задачі прогнозування вартості. Наприклад, можна проводити аналіз зміни кількості транзакцій у блоці або сумі сплачених комісій. У випадку якщо кількість транзакцій зростає – це може свідчити про зростання активності у досліджуваній мережі, що у свою чергу може викликати залучення додаткового капіталу, зростання вартості блокчейну й відповідно ціни криптовалюти або певного токена.

Обчислення визначених метрик може відбуватися як на основі інформації блоків, так і шляхом ітеративного перебору кожної транзакції та збору необ-

хідних даних. Однак, при роботі з цим джерелом даних необхідно усвідомлювати важливість оптимізації системи збору. Саме правильна організація роботи модулю з цим великим масивом даних дозволяє ефективно формувати аналітичні дані з мінімальними затратами обчислювальних ресурсів.

Криптовалюти та сфера децентралізованих технологій дуже тісно пов'язана з різними спільнотами. Даний вид сучасних технологій викликає цікавість у суспільства через свій значний потенціал розвитку нового етапу інтернет-технологій. Однак, популярність блокчейну, Web3 й як наслідок вартості криптовалют та токенів зростає за допомогою росту попиту з боку спільнот й оптимістичного погляду на майбутнє даної сфери.

Згадану особливість варто враховувати при аналізі та прогнозуванні курсу криптоактивів. Це можна реалізувати шляхом збору набору даних, який містить статистику щодо зміни настроїв відносно тієї чи іншої криптовалюти або токена. Даний вид аналізу (сентиментальний аналіз) можна проводити за допомогою обробки твітів з мережі X (Twitter). Це відбувається шляхом фільтрації та отримання твітів за певними параметрами та ключовими словами. Усі зібрані твіти зберігаються, після чого відбувається процес їх обробки засобами NLP. Для цього можна використовувати бібліотеку nltk. Вона надає набір інструментів для обробки та аналізу природних мов. Використання даної бібліотеки дозволяє провести аналіз емоційного забарвлення твітів шляхом їх попередньої обробки, аналізу та класифікації. Для прогнозування вартості криптовалют можна розподілити усі твіти у 3 категорії:

- негативні;
- нейтральні;
- позитивні.

Колекція твітів формується шляхом завантаження набору записів, які містять відповідні хештеги, достатню кількість переглядів і підписників у автора, а також відповідають певним іншим критеріям. Розподіл твітів у категорії відбувається за допомогою використання відповідного методу бібліотеки. Він надає можливість отримати оцінку емоційного забарвлення тексту повідомлення, що дозволяє віднести його до відповідної групи. На основі аналізу визначеної кількості твітів за певний проміжок часу формується статистична інформація про кількість позитивних, негативних та нейтральних повідомлень. Далі ця статистична інформація може використовуватися як часовий ряд й відповідно зміна кількості повідомлень у кожній категорії матиме вплив на результат прогнозування ціни криптовалюти.

Висновки. Використання різноманітних видів даних при прогнозуванні вартості криптовалют є досить актуальною тематикою, що досліджується багатьма науковцями. Незважаючи на те, що є велика кількість робіт, які намагаються визначити закономірності між зміною котирувань криптоактивів та певними наборами даних, усі джерела даних можна згрупувати у 3 категорії: технічні, фундаментальні та соціально-економічні. Кожна зі згаданих категорій вимагає окремого процесу для збору та обробки цих даних.

На основі визначених особливостей, був проведений опис методів роботи з кожним дже-

релом інформації та досліджені певних бібліотек, що дозволяють створити універсальні програмні модулі для збору даних кожного виду. Це було реалізовано шляхом більш глибоко дослідження самих джерел даних та засобів взаємодії з ними.

В результаті було наведено приклади створення набору даних з використанням інформації, яка належить усім 3 попередньо згаданим категоріям. Це дозволило дослідити процес на прикладному рівні, що спростить етап програмної реалізації засобів збору інформації для прогнозування вартості криптовалют.

Список літератури:

1. MODI, Parth Daxesh, et al. A Data-driven Deep Learning Approach for Bitcoin Price Forecasting. In: 2023 24th International Conference on Digital Signal Processing (DSP). IEEE, 2023. p. 1–4.
2. LIU, Shun, et al. Financial time-series forecasting: Towards synergizing performance and interpretability within a hybrid machine learning approach. arXiv preprint arXiv:2401.00534, 2023.
3. WU, Jingyang, et al. Review of deep learning models for crypto price prediction: implementation and evaluation. arXiv preprint arXiv:2405.11431, 2024.
4. AZAD, Poupak; AKCORA, Cuneyt Gurcan; KHAN, Arijit. Machine Learning for Blockchain Data Analysis: Progress and Opportunities. arXiv preprint arXiv:2404.18251, 2024.
5. ARACI, Dogu. Finbert: Financial sentiment analysis with pre-trained language models. arXiv preprint arXiv:1908.10063, 2019.
6. HARITHA, G.; SAHANA, N. Cryptocurrency price prediction using twitter sentiment analysis. In: CS & IT conference proceedings. *CS & IT Conference Proceedings*, 2023.
7. RAJU, S. M.; TARIF, Ali Mohammad. Real-time prediction of BITCOIN price using machine learning techniques and public sentiment analysis. arXiv preprint arXiv:2006.14473, 2020.
8. GURGUL, Vincent; LESSMANN, Stefan; HÄRDLE, Wolfgang Karl. Forecasting Cryptocurrency Prices Using Deep Learning: Integrating Financial, Blockchain, and Text Data. arXiv preprint arXiv:2311.14759, 2023.
9. SINGH, Shubham; BHAT, Mayur. Transformer-based approach for Ethereum Price Prediction Using Crosscurrency correlation and Sentiment Analysis. arXiv preprint arXiv:2401.08077, 2024.
10. ZEKIYE, Abdulrezzak, et al. AI-Assisted Investigation of On-Chain Parameters: Risky Cryptocurrencies and Price Factors. In: 2023 Fifth International Conference on Blockchain Computing and Applications (BCCA). IEEE, 2023. p. 52–59.

Sitailo M.S. RESEARCH ON THE DATA MINING PROCESS FOR USE IN CRYPTOCURRENCY PRICE PREDICTION

Artificial intelligence-based tools continue to rapidly gain popularity across various fields and find applications in solving specific practical problems. Despite the wide range of tasks that can be addressed with these tools, the process of developing neural networks consists of defined stages that are similar for almost all tasks. Data mining is an important stage in developing an AI-based solution. The volume and variability of the created dataset have a significant impact on the effectiveness of the developed neural network. Although the data collection process consists of certain defined stages, the specifics of the task often require additional operations. Asset price prediction has been a relevant task for a significant period of time. This encourages the active search for various methods that can improve prediction results. While this task is quite similar to predicting the prices of other types of assets, the features of blockchain technology introduce certain peculiarities during its resolution. Access to blockchain data allows for obtaining information about all transactions with certain cryptocurrencies or tokens, enabling the consideration of these factors in conducting fundamental analysis.

This article contains an analysis of current research related to the analysis of various types of data that can be used in predicting cryptocurrency quotations. Based on the conducted analysis, the features of working with the main categories of data sources used for technical, fundamental, and socioeconomic analyses were determined. The article also includes the description and design of the process of mining and processing data from crypto exchanges, blockchains, and social networks.

The obtained results contributed to conducting a practical study on the data mining process for cryptocurrency price prediction. This involved familiarizing with various web resources or software tools that can be applied to solve the given task. Based on the results, a description of the process of interacting with crypto exchanges, blockchains, and social media data was conducted to collect all considered categories of data. This process is based on the use of application programming interfaces, libraries for interacting with blockchains, and natural language processing libraries to assess the emotional tone of messages in certain social networks.

Key words: *data collection and processing, cryptocurrencies, dataset, socioeconomic analysis, technical analysis, fundamental analysis.*