

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет «Києво-Могилянська академія»

Факультет економічних наук

Кафедра фінансів

КВАЛІФІКАЦІЙНА РОБОТА

освітній ступінь – бакалавр

на тему: **«ETHEREUM: НОВА ТЕХНОЛОГІЯ НА ФІНАНСОВИХ РИНКАХ. ВПЛИВ
ЕМОЦІЙНОГО ЗАБАРВЛЕННЯ НОВИН ПРО КРИПТОВАЛЮТУ НА ЇЇ КУРС.»**

Спеціальності:

072 Фінанси, банківська справа та страхування

Сах Артем Ігорович

Керівник: Шпортюк В.Г.

кандидат фіз.-мат. наук, доцент

Рецензент _____

(прізвище та ініціали)

Кваліфікаційна робота захищена

з оцінкою « _____ »

Секретар ЕК _____

« _____ » _____ 2022 р.

Київ 2022

Зміст

| | |
|---|-----------|
| ВСТУП..... | 3 |
| РОЗДІЛ 1 ФУНДАМЕНТАЛЬНИЙ АНАЛІЗ КРИПТОВАЛЮТИ | 6 |
| 1.1. Криптовалюта як новий клас активів | 6 |
| 1.2. Фундаментальний та технічний аналіз активу | 14 |
| 1.3. Огляд досліджень на тему впливу новин на курс криптовалюти | 18 |
| РОЗДІЛ 2 МЕТОДОЛОГІЯ ДОСЛІДЖЕННЯ ТА ЗБОРУ ДАНИХ | 20 |
| 2.1. Огляд Ethereum та його біржових показників..... | 20 |
| 2.2. Збір даних | 23 |
| 2.3. Обробка даних | 25 |
| 2.4. Аналіз настрою..... | 33 |
| 2.5. Описовий аналіз даних | 35 |
| 2.6. Побудова однофакторної лінійної регресії..... | 35 |
| 2.7. Побудова логістичної регресії | 37 |
| РОЗДІЛ 3 ОЦІНКА ТА АНАЛІЗ РЕЗУЛЬТАТІВ..... | 40 |
| 3.1. Результати оцінювання моделі лінійної регресії..... | 40 |
| 3.2. Результати оцінювання моделі логістичної регресії | 42 |
| ВИСНОВКИ | 46 |
| СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ | 50 |

ВСТУП

Наразі складно заперечувати, що криптовалюта – актив майбутнього. Станом на 1 червня 2022 року ринкова капіталізація Біткоїна перетнула межу в 597 мільярдів доларів, а весь ринок криптовалют сягнув відмітки в 1.28 трильйони доларів. Якщо 10 років тому криптовалюти викликали недовіру в інвесторів, то сьогодні в них інвестують найбільші техно-корпорації та банки. Так, деякі банки навіть надають послуги криптовалютного гаманця та дозволяють торгувати активом з особистого кабінету. Одним з прикладів є Revolut, який ще в 2018 впровадив криптовалютний гаманець у своєму банківському додатку, який в даний час використовується понад 15 мільйонами користувачів по всьому світу. Це означає, що клієнти можуть платити у звичайних магазинах карткою, і сума до сплати буде автоматично конвертуватися в криптовалюту і вираховуватися з їхнього гаманця для криптовалют. Іншим прикладом є Coinbase, який пропонує передплачену дебетову картку, яка є повністю підкріплена криптовалютами. Це свідчить про зміщення ринку в бік криптовалютних активів. На відміну від традиційних інвестиційних активів, як то облігації чи акції, ціна яких сильно корелює з макроекономічними новинами (Birz, 2011), криптовалюти не підкріплені цінністю в реальному світі, а отже є більш спекулятивними [8]. Також, сильною відмінністю криптовалюти від звичних активів є її підвищена волатильність, через яку процес прогнозування її вартості є дуже ускладненим. Минулі дослідження також визнали успішним використання Twitter для прогнозування альткойнів (альтернативні, не основні криптовалюти) (Steinert, 2018) [9] та фондового ринку (Kirlic, 2018) [10]. Крім того, є також кілька комерційних постачальників прогнозів курсу криптовалют, які начебто оцінюють настрої ЗМІ; однак точна методологія, яку застосовує будь-яка з цих служб (наприклад, Coinpredictor) незрозуміла і не має документації. Згідно з Alexa.com,

на даний момент Twitter займає 11 місце за рейтингом найбільш відвідуваних веб-сайтів і вже брався як основа для багатьох досліджень із використанням аналізу настроїв

Таким чином, основною метою цього дослідження є дослідити криптовалюту як явище та можливість прогнозування курсу криптовалюти за допомогою сучасних технологій.

Досягнення поставленої мети передбачає вирішення таких завдань:

- Визначити поняття криптовалюти та принцип її роботи;
- Порівняти фундаментальний та технічний аналіз активу;
- Провести короткий огляд технічних та біржових показників обраного криптоактиву;
- Зібрати статистичні дані;
- Обробити та очистити ці дані для подальшої роботи;
- Провести аналіз настрою даних;
- Побудувати регресію та дослідити дані на наявність кореляції між настроєм та вартістю активу

Об'єктом дослідження є сфера криптовалютної торгівлі та окремий актив.

Предметом дослідження є детальне дослідження настрою публікацій про криптовалюту з метою прогнозування її курсу.

Методологічну основу дослідження складає використання мови програмування Python для роботи з даними та побудови економетричних моделей.

У першому розділі цієї наукової роботи досліджено поняття криптовалюти та блокчейну, фундаментального та технічного аналізу, і була розглянута вже існуюча література на схожу тему.

У другому розділі проведено збір статистичних даних, їхня обробка та очистка, аналіз настрою та побудова лінійної та логістичної регресії.

У третьому розділі проведена оцінка та аналіз результатів проведених обрахунків та економетричних моделей.

Для написання даної роботи використовувалися наукові джерела, перераховані у списку літератури.

Особливість нашого дослідження полягає в тому, що дослідженню підлягає значно менш популярна за Біткоїн криптовалюта під назвою Ethereum. Гіпотеза полягає в тому, що ринкова капіталізація Біткоїну складає близько половини всього криптовалютного ринку і на його ціну впливає занадто багато факторів. Вплив соціальних мереж буде не значним серед всіх інших. Натомість, активи, якими володіє значно менша кількість людей, підкорюються меншій кількості факторів і обговорення активу цією меншою спільнотою більше впливає на курс.

Практичне значення роботи полягає в тому, що викладені в ній положення та сформульовані висновки можуть слугувати вихідною науковою базою для нового методу прогнозування курсу активів та створення програми для автоматизованого трейдингу на основі штучного інтелекту та машинного навчання.

РОЗДІЛ 1

ФУНДАМЕНТАЛЬНИЙ АНАЛІЗ КРИПТОВАЛЮТИ.

1.1. Криптовалюта як новий клас активів

Криптовалюта – цифрова заміна звичним грошам, для володіння якою непотрібен банк чи дозвільний орган. Вона зберігається в особистому електронному гаманці. Гаманець може бути як програмним, що знаходиться на вашому комп'ютері чи в хмарному сервісі, так і фізичним пристроєм, схожим на USB-накопичувач, що створений виключно для зберігання криптовалюти. Найбільш безпечним вважається фізичний гаманець з високим рівнем шифрування та біометричним методом авторизації. Працює криптовалюта на базі технології блокчейн. Вона була створена з метою вирішення проблем, притаманних звичним грошам:

- Централізованість;
- Прозорість.

Проблема подвійної витрати. Візьмемо за приклад фото в телефоні. Ми маємо оригінал фотографії в галереї. Ми можемо завантажити її в Facebook і таким чином створимо її копію. Можемо завантажити в Instagram і створити другу копію. Це буде безкоштовно і ні на кого не вплине негативно створення кількох копій однієї фотографії. Тепер уявімо, що замість фотографії в галереї в нас є сто монет криптовалюти. Ми відправляємо своєму другу ці сто монет і, як і у випадку з фотографією, він отримує копію (ще сто монет) без якихось витрат. А потім надсилає іншому другу копію. І так до нескінченності. Різниця від випадку з фотографією полягає в тому, що криптовалюта має цінність і друзі зацікавлені в її копіюванні з метою збагачення. В звичайній фінансовій системі банк та держава слідкують за тим, щоб ми не мали можливості «дублювати» гроші. Але така система – централізована, а стоїть задача створити децентралізовану фінансову

систему. Треба знайти спосіб уникнути втручання держави і унеможливити шахрайство з боку тримачів валюти.

Так, 2008 року Сатоші Накамото опублікував проект під назвою «Біткоїн». Ідея проекту полягає у введенні блокчейну для усунення посередників у вигляді банків та гарантії захищеності від підробок. [5]

“Блокчейн – це децентралізована, розподілена книга, що реєструє походження цифрового активу.” [14] “Іншими словами: блокчейн – це база даних, що складається з впорядкованих блоків і постійно довшас. Щоб зрозуміти, що таке блокчейн, треба зрозуміти, що таке база даних. База даних – набір інформації збереженої та структурованої електронно в комп’ютерній системі. Зазвичай бази даних структуровані у вигляді таблиць, але від звичайних, якими ми звикли користуватися, їх відрізняє об’єм та потужність.” [25] Звичайні таблиці зручні для використання невеликою групою людей і можуть зберігати обмежений обсяг інформації. Зазвичай вони використовуються бізнесом. Бази даних спроектовані таким чином, що їхня місткість може значно перевищувати місткість звичайних таблиць. З ними може проводити маніпуляції надзвичайно велика кількість користувачів одночасно. Це досягається за рахунок утримання баз даних на серверах, що складаються з мережі надпотужних комп’ютерів (тисяч чи десятків тисяч пристроїв).

Різниця між звичайною базою даних та блокчейном полягає в структурі даних. В блокчейні дані збираються в групи, що називаються блоками. Блоки мають визначений розмір і, коли вони заповнюються, то накладаються на попередній заповнений блок, формуючи, таким чином, ланцюг під назвою «блокчейн». Вся нова інформація збирається в новий блок, який після заповнення буде так само об’єднаний з попередніми.

Бази даних структурують інформацію в таблиці, в той час як блокчейн – в блоки, об’єднані між собою. Кожному блоку в ланцюзі призначається точна позначка часу, коли він додався до ланцюжка. [3]

Тож, якщо коротко, «блокчейн – повна і незмінна історія транзакцій децентралізованої спільноти, з якою згідні всі учасники спільноти.» [5] Цей

реєстр оновлюється автоматично в реальному часі, приймається спільнотою і зберігається на комп'ютері кожного учасника. Таким чином відпадає необхідність в центральному контролюючому органі, оскільки жоден учасник не зможе виконати подвійну витрату. Вона одразу призведе до конфлікту в історії кожного учасника.

Децентралізація. Щоб зрозуміти, як працює блокчейн, розглянемо його застосування в Біткоїні. Криптовалюті потрібне місце для зберігання блокчейну з даними про всі транзакції. Для цього використовується децентралізована мережа з комп'ютерів. На відміну від звичайних баз даних, дані про Біткоїн зберігаються на комп'ютері кожного учасника. [25]

Можемо уявити корпорацію-власника приміщення, де знаходиться сервер з базами даних. У неї є доступ і контроль над даними. База даних Біткоїну знаходиться не на комп'ютерах в одному приміщенні, а на пристроях в різних куточках світу. Кожен з таких комп'ютерів називається вузлом.

У цьому випадку блокчейн Біткоїну використовується децентралізовано. Але окремі компанії можуть створювати приватні централізовані блокчейни, де комп'ютери, що складають його мережу, належать і експлуатуються підприємством.

“У блокчейні кожен вузол містить повну історію даних, які зберігались у блокчейні з моменту його створення. Для Біткоїнів дані – це історія всіх транзакцій. Якщо один вузол має помилку в своїх даних, він може використовувати тисячі інших вузлів, щоб виправити себе. Таким чином, жоден вузол у мережі не може змінити інформацію, що зберігається в ній. Через це історія транзакцій у кожному блоці, що становить блокчейн Біткоїна, незворотна.”[25]

Якщо хтось спробує втрутитися в блокчейн, усі інші вузли звіряться між собою і легко визначать вузол з неправильною інформацією. Ця система допомагає встановити точний і прозорий порядок подій. Для Біткоїнів ця інформація є історією транзакцій, але блокчейн може містити таку інформацію, як

юридичні контракти, державні ідентифікаційні дані або інвентаризацію товарів компанії.

Прозорість. В блокчейні кожен може відслідковувати всі коли-небудь зроблені транзакції в реальному часі. Для цього необхідно мати особистий вузол, оскільки на кожному вузлі зберігається копія ланцюга. Або можна скористатися онлайн переглядачем блокчейну.

Були випадки, коли криптовалютні біржі атакували хакери та користувачі втрачали свої активи. Попри те, що особистість хакера лишилася прихованою, вкрадену криптовалюту легко відстежити, оскільки всі бачать номер гаманця, на який вона була виведена. Інша справа, що сам номер не несе жодної інформації про особистість хакера.

Безпека. Основою безпеки технології блокчейну є лінійність запису блоків. Порядок запису блоків важливий. Кожен блок має “висоту”. Висота блоку в блокчейні – це його власна позиція в ланцюгу. Станом на червень 2022 року загальна висота блокчейну Bitcoin досягла 738815 блоків.

Щоб змінити систему необхідна згода 51% учасників системи. Якщо хакер захоче змінити блокчейн і викрасти Біткоїн у інших тримачів, йому потрібно одночасно контролювати і змінити 51% копій блокчейну, щоб їх нова копія стала більшістю копій і, таким чином, узгодженою ланцюжком. Така атака потребує величезної суми грошей та ресурсів, Потрібно було б відредагувати всі блоки, оскільки вони тепер мали б різні мітки часу та хеш-коди. Більше того, вузли розкидані по всьому світу в величезній кількості. Тож такий злам є практично неможливим, що і гарантує безпеку від фальсифікації. Навіть якби таке сталося, члени мережі побачили б такі різкі зміни в блокчейні і перейшли б до нової версії ланцюга, на яку зміни не вплинули. Це призведе до того, що атакована версія Біткоїна різко знизиться в ціні, що зробить атаку безрезультатною, оскільки хакер стане власником знеціненого активу. Те саме відбулося б, якби хакер атакував нову версію Біткоїна. Він побудований таким чином, що участь у мережі набагато краще економічно стимулюється, ніж її атака.

“**Валюта.** Блокчейн формує основу для таких криптовалют, як Біткоїн. Американський долар контролюється Федеральним резервом. Відповідно до цієї централізованої системи, дані та валюта користувача контролюються банком чи урядом. Якщо банк користувача зламаний, особиста інформація клієнтів знаходиться під загрозою. Якщо банк клієнта банкрутує або клієнт живе в країні з нестабільною економікою, вартість їхньої валюти може дуже сильно коливатися. У 2008 році деякі банки, у яких закінчилися гроші, були частково фінансовані за рахунок грошей платників податків. Це ті проблеми, через які Біткоїн був задуманий і розроблений.” [25]

Функціонування криптовалют без втручання центрального контролюючого органу має багато переваг. По-перше, зменшуються операційні витрати. По-друге, криптовалюта інтернаціональна і може слугувати місцем збереження капіталу для жителів країн з нестабільною економікою. По-третє, щоб отримати криптовалютний рахунок не потрібна державна ідентифікація, що сильно допоможе біженцям та іншим людям, що її не мають. [3]

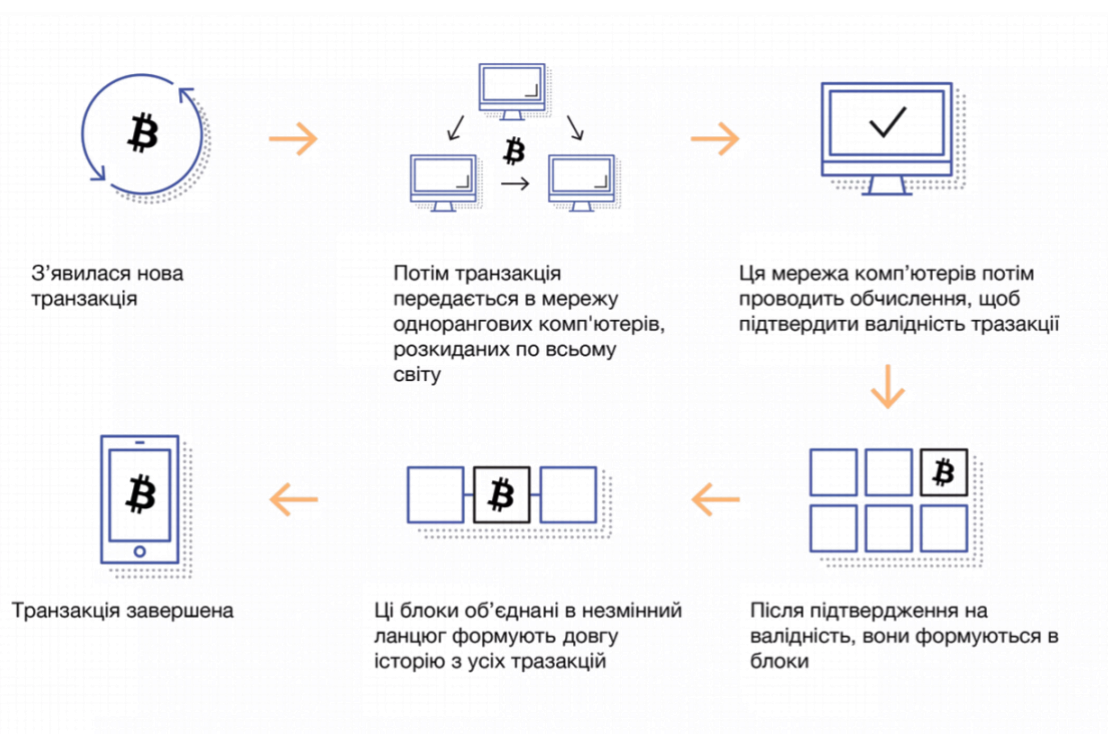


Рис. 1.1 Візуалізація процесу проведення транзакції в мережі блокчейн.

Джерело: Складено автором з на базі даних [25]

Криптовалюта та її цінності виникають на вищому рівні людських потреб. З розвитком технологій наше життя все більше переноситься з реального світу в віртуальний. Це стосується не лише роботи, але і спілкування, перегляду фільмів, ігор. Пандемія COVID-19 остаточно довела, що людство здатне функціонувати в віртуальному світі і часто робить це більш ефективно, ніж в реальному. Тому закономірно, що з перенесенням життя в віртуальний світ, цінності теж переносяться туди. Нематеріальні об'єкти в віртуальному світі набувають цінності на рівні з матеріальними об'єктами. Так, якщо двадцять років тому художники могли продати свої роботи лише на фізичних носіях (полотно, папір тощо), то зараз це не обов'язково. Технологія NFT на базі блокчейну дозволяє художникам продати свій витвір цифрового мистецтва за криптовалюту, а покупцю отримати цифрове підтвердження оригінальності купленої роботи, закарбоване в системі блокчейн. Купівля загальнодоступної картинки може здатися незрозумілою, але популярність даної технології довела, що цінність мистецтва не в носії, на якому воно збережене, а в самому мистецтві. Щоб зрозуміти, в чому полягає цінність цифрового мистецтва, проведемо аналогію з реальним випадком, що стався в Україні навесні 2022. Укрпошта випустила марку з російським кораблем, що тоне. На ній зображена подія, що стала символічною для всього українського народу і відомою на весь світ – відмова українських прикордонників на острові Зміїний здатися в полон екіпажу російського крейсера “Москва”. Цю подію зобразив художник, а Укрпошта випустила марки обмеженим тиражем. Попит на ці марки був надзвичайним, а їхня ціна на вторинному ринку дуже швидко досягла кількох тисяч доларів. На цьому прикладі можемо спостерігати, як щось, що не має матеріальної цінності, набуло її в зовсім інший шлях за рахунок контексту та мистецтва. Звісно, це все має застосування лише після задоволення людиною її найбільш базових потреб, таких

як їжа, житло та одяг, але з плином часу все більше людей переходять на вищий рівень потреб і починають цікавитися віртуальним всесвітом.

Зважаючи на те, що в Україні в війна, варто звернути уваги на те, як веде себе криптовалюта. Перш за все – криптовалюта показала себе як додатковий засіб для збору коштів, який суттєво спрощує цей процес для іноземців. Платежі відбуваються дуже швидко, комісії значно менші, з реквізитів треба лише номер криптовалютного гаманця. Також криптовалюта є єдиним засобом для осіб, які хочуть пожертвувати кошти та залишитися анонімними. Також набув поширення продаж робіт наших митців на NFT платформах для збору коштів на ЗСУ. Таким чином навіть креативна індустрія конвертує свої професійні навички в гроші, які йдуть на допомогу українській армії. Проте, також була помічена схема фінансових махінацій з використанням криптовалют. Люди реєстрували рахунки в іноземних банках, переводили кошти з українського рахунку, купували там долар по курсу 30 грн за 1 \$, за нього купували криптовалюту USDT, ціна якої закріплена за ціною долара, і купували за неї гривню по курсу 33-34 гривні за 1 USDT. Після цього ці кошти виводили з криптовалютної біржі на українську карту і повторювали 20 разів за день, заробляючи на різниці курсів. Станом на 1 червня українські банки навчилися розпізнавати такі махінації то почали блокувати картки громадян, які зловживають такими методами заробітку.

Які основні відмінності між фондовою та криптовалютною біржами?

1. Активи, якими торгують

На фондовій біржі торгують акціями компаній, тоді як на криптовалютній біржі торгують цифровими валютами, такими як Біткоїн, Ethereum та інші.

2. Цінність активу

Цінність акцій на фондових біржах забезпечена власним капіталом компанії. Власник акції є частковим власником компанії-емітента. Чим краще працює компанія, тим багатшою вона стає і тим дорожче коштує акція в володінні покупця.

В багатьох випадках цінність криптовалюти суб'єктивна. Цифрова валюта не завжди підкріплена чимось з реального світу. Часто все, що стоїть за валютою – електроенергія, витрачена на її майнинг. В серйозних проектах монета підкріплена новітньою технологією, яка є потенційно дуже перспективною в майбутньому.

3. Випуск активів

Компанія може випускати акції для залучення капіталу. Кількість акцій та їхня ціна необмежені.

Криптовалюти ж в більшості випадків мають обмежену емісію. Основні економічні науки припускають, що (з урахуванням усіх інших факторів) вартість життєздатних, обмежених криптовалют зростатиме в міру зростання попиту на них.

4. Урегульованість

Фондові біржі доволі старі і мають багату історію. Існує безліч нормативних актів та законів, що регулюють їхню діяльність. Держава забезпечує їх підтримкою, а компанії звітують про свою діяльність перед акціонерами.

Криптовалютні біржі молоді, але розвиваються шаленими темпами. Попри спроби врегулювати криптовалютний ринок більша частина його діяльності залишається нерегульованою. Також, про молодий вік ринку свідчить менший вибір криптовалют для торгівлі, ніж на фондовому ринку.

5. Волатильність

В багатьох волатильність викликає тривогу. Насправді волатильність ринку можна розглядати як в позитивному, так і в негативному ключі.

На фондовій біржі відносно низька волатильність. Це означає, що ринок більш стабільний, але і очікування фінансової винагороди довше. Великі обсяги торгівлі роблять його менш чутливим до впливу торговців-китів. Тим не менше фондова біржа часто зазнає впливу геополітичних подій.

Біржі криптовалют більш волатильні. Менші обсяги торгівлі роблять його вразливим для маніпуляцій торговців-китів. Торговець-кит – це той, хто володіє великою кількістю Біткоїнів. Це означає, що весь ринок може бути вразливим до

торгових рішень тих, хто вкладає значні кошти. Наприклад, завдяки новині про Ілона Маска, який інвестував 1,5 мільярда доларів у Біткоїни у січні 2021 року, ціна Біткоїна раптово підскочила на 17% до нового рекордно високого рівня.

6. Доступність

Зарегульованість фондової біржі призводить до того, що швидко почати торгувати не вийде. Потрібно знайти собі брокера, а потім отримувати схвалення на купівлю та продаж. Крім того, торгівля обмежена графіком роботи біржі.

Натомість, почати торгувати криптовалютою можна за лічені хвилини. Криптовалютна біржа працює в будь-який час і в будь-який день, незалежно від державних свят чи інших подій. Будь-яка людина здатна торгувати криптовалютами, що робить її набагато доступнішою для людей з усіх соціальних щаблів.

7. Комісії та правила

На фондових ринках значно вищі комісії та різні збори, пов'язані з торговими операціями. Також існують правила, що захищають торговців та інвесторів.

Торгівля на біржах криптовалют значно менш витратна. Витрати, пов'язані з транзакціями на блокчейні, є незначними, складаються лише з плати за майнінг. Таким чином, самі біржі несуть менші витрати при купівлі-продажу криптовалют, ніж брокери на фондових біржах. [2], [4]

1.2. Фундаментальний та технічний аналіз активу

Аналіз криптовалюти багато в чому наслідує аналіз більш консервативних активів, таких як акції. Він поділяється на фундаментальний та технічний аналіз.

Протягом довгого часу фундаментальний аналіз був основним методом аналізу активів і використовувався лише для акцій та облігацій. Поступово його методики почали застосовувати на інших активах. Так черга дійшла й до криптовалюти.

Головним компонентом фундаментального аналізу є реальна цінність активу. Замість того, щоб намагатися передбачити зміну ціни активу в майбутньому, фундаментальний аналітик намагається визначити справжню вартість активу. Для цього він враховує багато факторів від економічної ситуації в галузі до ефективності команди управління. Фундаментальний аналіз стверджує, що ринок може в короткостроковій перспективі неправильно оцінити цінний папір, але врешті-решт він виправиться, відображаючи справжню цінність цінного паперу.

Дізнавшись реальну вартість цінних паперів, ми можемо сказати чи цінний папір переоцінений, недооцінений або справедливо оцінений. Залежно від цього ми можемо купувати, продавати чи утримувати його з метою отримання прибутку. Наприклад, ми можемо отримати прибуток від придбання активу, коли він був недооцінений ринком, та його продажу, коли він досягне справедливої ціни або стане переоціненим.

Коли ми застосовуємо фундаментальний аналіз до акцій, ми аналізуємо в першу чергу компанію – емітента. Але якщо криптовалюти не випускаються корпораціями, то лишається аналізувати навколишні фактори, які можуть на неї вплинути. Наприклад, ми можемо оцінити стан криптовалютного сектору, ринок в цілому, внутрішнє та глобальне економічне середовище, попит та пропозицію, щоб прийняти інвестиційне рішення

Ось деякі інші базові показники, які можуть нам допомогти:

Довідковий документ: у криптовалюті - це еквівалент проспекту компанії.. Довідковий документ у криптосекторі містить деталі, необхідні для розуміння передбаченого використання активу, його специфікацій та цілей. Однак, оскільки ці документи не контролюються регуляторними органами, бажано перевіряти твердження в них.

Цільовий ринок: на кого націлений криптовалютний актив? Який розмір цільового ринку? Які є випадки використання активу? Чи намагається вона замінити застарілу технологію?

Розробники: варто дослідити, хто стоїть за проектом, кількість учасників проекту та масштаб діяльності навколо активу. Оскільки деякі криптовалюти все ще перебувають у розробці, кількість розробників, які працюють над проектом, часто може бути ознакою його потенціалу. Це також показник прихильності виконавчої команди до реалізації своїх амбіцій. До проектів з невеликою кількістю розробників слід ставитися обережно.

Чутки: чутки не варто сприймати за чисту монету, але і ігнорувати – теж. Часом корисно знати, що думають колеги про певну криптовалюту. Які загальні настрої щодо цього проекту? Чи здається проект реалістичним для інших? Як це поєднується з обстановкою в секторі? [1], [2]

Технічний аналіз спрямований на розуміння ринкових настроїв, шляхом пошуку шаблонів та тенденцій, передбачення цінових рухів, вивчення історичних даних, таких як ціна та обсяг.

В рамках дисципліни технічного аналізу аналітик може використовувати як кількісні, так і якісні методи для оцінки прибутковості інвестицій. Технічний аналіз базується переважно на кількісних показниках. Оскільки в ньому використовується математичне та статистичне моделювання, щоб зрозуміти ринкову поведінку, технічний аналіз має велике значення при покупці та продажу криптовалюти.

В технічному аналізі криптовалют існують 3 базові поняття:

1. Історія повторюється
2. Ринок враховує все
3. Ціни рухаються в трендах

Озброївшись цими переконаннями, технічні аналітики зважають на безліч чинників, включаючи поведінкові та традиційні економічні принципи, щоб зробити прогноз ринку. На відміну від фундаментального аналізу, технічний аналіз бере до уваги виключно цінові рухи. Але, щоб зрозуміти рухи цін, використовуючи технічний аналіз, ви повинні зважати на такі фактори, як зміна попиту на криптовалюту, глобальні регуляторні режими та останні тенденції у крипто-спільноті. [4]



Рис. 1.2. Зображення застосування графічного технічного аналізу на прикладі акцій International Business Machine Stock Price.

Джерело: Складено автором з на базі даних [18]

1.3. Огляд досліджень на тему впливу новин на курс криптовалют

З метою ознайомлення зі схожими дослідженнями було обрано дві статті на подібну тематику.

1. «Forecasting of the cryptocurrency market through social media sentiment analysis» написана Адамом Салаком. [6]

Його робота має на меті оцінку можливості використання даних із соціальних мереж для прогнозування ціни на Біткоїн на основі аналізу настрою дописів. Ключовою відмінністю від вже існуючих робіт є використання соціальної мережі Reddit як джерела інформації, та порівняння даних з Reddit та Twitter. Автор аргументує вибір Reddit якістю постів. В соціальній мережі відсутні обмеження в кількості символів, а всі публікації належать до теми гілки обговорення (а ті, що не стосуються теми – видаляються модераторами). Також тема криптовалют доволі часто зустрічається на Reddit. Згідно з проведеним дослідженням, кореляція була відсутня.

Метою методологічного підходу автора є оцінка зв'язку між настроями в соціальних мережах і прогнозованою вартістю криптовалют. Необхідно визначити набір змінних, щоб можна було зробити регресійний аналіз. У цьому випадку залежна змінна була визначена як ціна Bitcoin в доларах США, а незалежна змінна буде представлена оцінкою аналізу настроїв дописів в Reddit.

Незважаючи на те, що існують дослідження, в яких автори використовують лише дані з підтверджених акаунтів Twitter, Адам Салак відмовляється від такого підходу. Головним аргументом є більше неформальна природа крипто-трейдингу порівняно з біржами цінних паперів.

Було визнано, що VADER є підходящим інструментом для аналізу настроїв, і його також використовували в попередніх пов'язаних дослідженнях, проведених Kirlic [10] або Steinert [9]. VADER – Python бібліотека, яка використовується для

аналізу настроїв на основі словника, але протягом усього часу була портована на різні мови програмування та платформи.[6]

2. «Nowcasting the Bitcoin Market with Twitter Signals» написану Jermain C. Kaminski

Ця робота досліджує кореляцію між різними ринковими показниками Біткоїну та емоційною складовою постів на Twitter.

Методологія. Автор зібрав датасет з більш ніж 160 000 твітів від 57 727 користувачів Твіттеру в часовому проміжку з 23 листопада 2013 до 7 березня 2014. Виходячи з розрахунків сервісу для збору датасету ці твіти мали охоплення 300 000 000. Це означає, що сумарно триста мільйонів разів ці твіти потрапляли в стрічки користувачам. Для визначення емоційного забарвлення автор не використовував спеціалізовану Python бібліотеку. Натомість він просто вибрав всі твіти, що містять слово “Bitcoin” і розділив їх на ті, що містять слова “feel”, “happy”, “great”, “love”, “awesome”, “lucky”, “good”, “sad”, “bad”, “upset”, “unhappy”, “nervous”, “hope”, “fear”, “worry”. Потім була проведена чистка даних від твітів, що містять ці слова не в контексті криптовалют (“happy birthday”) чи ті, що не мають однозначної емоції (“not bad”). Ці твіти були розділені на чотири категорії: сума позитивних твітів (131 117 твітів), сума негативних твітів (19 179 твітів), сума емоцій (сума двох попередніх груп) та сума твітів, що відображають невпевненість (10 222 твітів).

Згідно з дослідженням Джермейна, найбільша криптовалютна біржа з точки зору обсягу торгів найбільш чутлива до негативних твітів та тих, що відображають невпевненість. Також дні з малою кількістю негативних твітів корелюють з вищою ціною закриття торгів Bitcoin. Також більше співвідношення позитивних твітів до негативних супурує до вищою ціною закриття. При цьому, сама по собі кількість позитивних твітів не корелює з ціною закриття.

Сума емоційних твітів та твітів-сумнівів протягом дня корелює з волатильністю ціни протягом того ж дня.

В роботі сказано, що «статичні внутрішньоденні вимірювання свідчать про помірну кореляцію настрою постів Twitter із ціною на Біткоїн. Також, модель з додаванням лагової змінної показала, що сума емоційних настроїв (особливо негативних сигналів) позитивно корелює з обсягом внутрішньоденних торгів у межах останніх 48 годин. Це можна трактувати наступним чином: коли обсяг торгів є високим, кількість емоційних висказувань в Twitter зростає. Як такий, Twitter можна трактувати як місце, яке відображає "спекулятивний імпульс".»

[7]

РОЗДІЛ 2

МЕТОДОЛОГІЯ ДОСЛІДЖЕННЯ ТА ЗБОРУ ДАНИХ

2.1. Огляд Ethereum та його біржових показників

Ethereum – це децентралізована програмна платформа з відкритим кодом. В її основі лежить блокчейн, а сама вона використовується для власної криптовалюти під назвою “ETH” або “Етеріум”, “Етер”, “Ефір”. Платформа популярна для створення децентралізованих програм (DApps), а також смарт-контрактів. Ethereum був випущений 2015 року.

“Ethereum - не лише платформа, але й мова програмування, що працює на блокчейні. Додатки, що працюють на Ethereum, функціонують на специфічному для платформи криптографічному токени під назвою Ефір. 2014 року розпочався попередній продаж ETH і зустрів шалений попит. Ефір - це як транспортний засіб для пересування на платформі Ethereum, і його, переважно, використовують розробники, які прагнуть розробляти та запускати програми всередині Ethereum.

Варто розрізняти Ethereum та ЕТН, оскільки Ethereum – це блокчейн, а ЕТН – основний об’єкт блокчейну Ethereum.

ЕТН – це основа Ethereum. Коли ви надсилаєте ЕТН або використовуєте програму Ethereum, то сплачуєте невелику комісію в ЕТН за використання мережі Ethereum. Ця комісія заохочує майнерів до обробки й перевірки того, що ви намагаєтеся зробити.

Майнери – це реєстратори Ethereum. Вони проводять перевірки й попереджають випадки шахрайства. Майнери, які цим займаються, також отримують нові ЕТН у невеликій кількості як винагороду.

Завдяки майнерам мережа Ethereum залишається безпечною та вільною від централізованого контролю. Іншими словами, ЕТН забезпечує роботу мережі Ethereum.

ЕТН широко використовується для багатьох цілей. Перш за все, ЕТН – це дефіцитні цифрові гроші, якими можна користуватися в інтернеті (схожі на Bitcoin). Для цього користувач повинен мати лише доступ до інтернету та криптовалютний гаманець. Він виступає посередником між користувачем та мережею blockchain і зберігає інформацію про активи, вказуючи їх місцезнаходження в blockchain. Кожен може зареєструвати криптогаманець безкоштовно за лічені секунди без необхідності вказувати будь-які персональні дані. Єдиним мінусом такої анонімності є неможливість відновлення гаманця, якщо користувач забув пароль. Розробники не мають доступу до гаманців користувачів.

Друга функція – децентралізовані фінанси (DeFi). DeFi – збірне поняття, що охоплює фінансові продукти та послуги, доступні всім користувачам Ethereum. Ця система дозволяє ринкам бути завжди відкритими і не містить центрального органу влади, який би відмовляв у доступі чи блокував платежі. Також вона усуває людський фактор та автоматизує послуги, які раніше вимагали багато часу та були вразливими до помилок. Це досягається за рахунок того, що код доступний для огляду й аналізу всім бажаючим. Криптоекономіка пропонує

інструменти для позики, придбання страхування, довгих і коротких пропозицій, вкладів під відсотки тощо.

Також на базі Ethereum створюються децентралізовані програми (DApps). Багато програм досі експериментальні, проте вже є успішні приклади у галузях технологій, фінансів, ігор та колекціонування. Так, Aave та Compound дозволяють користувачам позичати свої токени, щоб заробити відсотки та вивести їх в будь-який момент. Tornado cash дозволяє надсилати анонімні транзакції. Zapper дає можливість відстежувати своє портфоліо і проводити оцінку кожного активу окремо. Pooltogether – лотерея, у якій неможливо програти. Учасники купують лотерейний квиток за певну суму, за перебування цих грошей у власності платформи генеруються відсотки і щодня випадковий користувач виграє ці всі відсотки. Foundation та OpenSea дозволяють митцям продавати свої роботи, а колекціонерам – купувати їх. Decentraland – гра з віртуальним світом, в якому можна купувати та продавати земельні ділянки, спілкуватися з іншими користувачами та обмінюватися предметами.

Отже, Ethereum – це багаторівнева платформа, базована на технології блокчейн, створена для обміну цифрової валюти, яка використовується всередині Ethereum для запуску додатків і навіть для монетизації роботи.” [16]

Ринкові показники Ефіру наступні:

- Ціна однієї монети станом на 30 травня 2022 року = 1924 доларів
- Ринкова капіталізація = 232.8 млрд доларів
- Валюти в обігу = 120.98 млн монет (необмежена емісія)
- Обсяги торгів за останню добу = 16 277 951 доларів
- Зміна ціни за останню добу = +6.8%
- Зміна ціни за останній тиждень = -6.3%
- Зміна ціни за останній місяць = -31.5% [11]

На графіку нижче синьою лінією зображена зміна курсу Ефіру до долара США з серпня 2015 по травень 2022. Сіра гістограма – обсяги торгівлі в конкретний момент часу.



Рис. 2.1. Динаміка зміни курсу Ethereum в період з серпня 2015 по червень 2022.

Джерело: Складено автором з на базі даних [11]

2.2. Збір даних

Дані були взяті з відкритого датасету «Cryptocurrency News Tweet» з сайту OpenBlender. [17]

В датасеті зібрано близько тисячі твітів одного з найбільших новинних ресурсів спеціалізованих на криптовалюти – CryptoBoom News [18] Станом на 1 червня 2022 Twitter-акаунт цього ресурсу під назвою «Cryptocurrency News» налічує 419 тисяч читачів, що є дуже хорошим показником для крипто-спільноти. Він був зареєстрований в січні 2011, тож може вважатися надійним. [19]

Стратегія збору саме твітів, а не статей з сайтів має пояснення. По-перше, ринок криптовалют значно менш консервативний за фондовий і на ньому значно більше учасників молодого віку. По-друге, криптовалютний ринок значно більш

волатильний за фондовий, а отже потребує швидкої реакції. Процеси, пов'язані з публікацією статей на сайтах, такі як редагування, затвердження і т.д., роблять ці статті неактуальними для прийняття рішення про операцію з активом на момент виходу статті. Натомість, твіти значно коротші, не потребують довгого редагування і через свою актуальність впливають на прийняття рішення про покупку чи продаж валюти. До того ж, в Twitter публікуються короткі варіанти всіх статей з основного сайту новинного ресурсу, тож вони не залишилися поза дослідженням.

Методологія збору інформації в цьому дослідженні схожа на методологію використану в дослідженні Джермейна Камінського [7]. Його датасет також складається з дописів користувачів Twitter, проте він досліджував зовсім іншу криптовалюту – Bitcoin. В той самий час, методологія збору даних Адама Салака. [6] суттєво відрізняється від нашої, оскільки він збирав тексти дописів користувачів зовсім іншої соціальної мережі – Reddit. Також він досліджував Bitcoin, а не Ethereum.

2.3. Обробка даних

Оригінальний датасет твітів має такий вигляд:

Таблиця 1.1 Візуалізація оригінального датасету з новинами на прикладі одного семплу

| | re_tweeter | author | author | text | Hashtags | reply_count | associated_tweet | timestamp | mentions | author_id | retweet_count | type | id | favorite_count |
|---|------------|--------|---------------------|---|----------|-------------|------------------|------------|----------|-----------|---------------|-------|---------------------|----------------|
| 0 | | | CryptoCurrency News | Good Morning #Crypto Twitter! A new week ahead and the #Cryptocurrency market has recovered nicely heading back over a \$2 Trillion market cap with #Bitcoin back over \$53,000 | | | | 1619436842 | | | | Tweet | 1386644736686104576 | |

Він непридатний для дослідження в такому вигляді, оскільки містить багато зайвих ознак, незаповнених полів та незрозумілий людині формат часу UNIX TimeStamp. Тому було прийняте рішення провести чистку датасету.

1. Видалення зайвих стовпців.

З датасету для дослідження необхідні лише дві змінні – [text] - текст твіту, який буде пізніше аналізуватися, та [timestamp] - дата публікації, по якій можна буде об'єднати цей датасет з датасетом ціни на Ethereum. За допомогою наступних команд на Python з застосуванням відкритої бібліотеки Pandas [22] було залишено в датафреймі лише дві ознаки – [timestamp] та [text]

```
import pandas as pd
```

```
df = df[['timestamp', 'text']]
```

```
df.head()
```

| | timestamp | text |
|---|------------|---|
| 0 | 1619436842 | Good Morning #Crypto Twitter! A new week ahead... |
| 1 | 1619403087 | Well that 13 day bear market was fun. Now it's... |
| 2 | 1619399970 | FTX had the largest market share growth of all... |
| 3 | 1619398093 | The market is GREEN again #Bitcoin #Cryptocur... |
| 4 | 1619383095 | #Ethereum vs #BSC active addresses. Pretty ama... |

Рис. 2.2 Змінні з датасету, необхідні для дослідження

2. Адаптація часу публікації

Нам необхідно перевести час публікації з машинного формату UNIX TimeStamp в звичний dd/mm/yyyy. За допомогою функції «datetime.fromtimestamp» з модуля datetime ми конвертуємо дату публікації в читабельний вигляд. Потім видаляємо [timestamp], лишивши лише [date] та [text].

```
from datetime import datetime
```

```
df['date'] = (df['timestamp']).apply(datetime.fromtimestamp)
```

```
df = df[['date', 'text']]
```

```
df.head()
```

| | date | text |
|---|---------------------|---|
| 0 | 2021-04-26 13:34:02 | Good Morning #Crypto Twitter! A new week ahead... |
| 1 | 2021-04-26 04:11:27 | Well that 13 day bear market was fun. Now it's... |
| 2 | 2021-04-26 03:19:30 | FTX had the largest market share growth of all... |
| 3 | 2021-04-26 02:48:13 | The market is GREEN again #Bitcoin #Cryptocur... |
| 4 | 2021-04-25 22:38:15 | #Ethereum vs #BSC active addresses. Pretty ama... |

Рис. 2.3 Конвертований формат часу з машинного формату UNIX TimeStamp в dd/mm/уууу

Лишилася ще одна маніпуляція з часом. Як видно з попереднього зображення, крім дати публікації в нас є час. Для об'єднання двох датасетів різних розмірів нам потрібен 1 стовпчик з однаковими даними в обох датасетах. Час публікації унеможливить таке об'єднання, тому нам треба його позбутися і привести дату до вигляду уууу-mm-dd. Для цього просто відрізаємо все, що йде після десятого символу в кожній клітинці за допомогою команди функції `str.slice()`.

```
df_clean['day'] = (df_clean['date']).str.slice(stop=10)
#df_clean = df_clean.drop(['date', 'timestamp', 'Unnamed: 0'], axis=1)
df_clean = df_clean[['day', 'text', 'compound']]
df_clean.head()
```

| | day | text |
|---|------------|---|
| 0 | 2021-04-26 | Good Morning Crypto Twitter! A new week ahead ... |
| 1 | 2021-04-26 | Well that 13 day bear market was fun. Now it's... |
| 2 | 2021-04-26 | FTX had the largest market share growth of all... |
| 3 | 2021-04-26 | The market is GREEN again Bitcoin Cryptocurre... |
| 4 | 2021-04-25 | Ethereum vs BSC active addresses. Pretty amazi... |

Рис. 2.4 Видалення точного часу публікації зі змінної [day]

3. Чистка тексту від зайвих символів

Оскільки ми проводимо `sentiment analysis` кожного твіта, то треба максимально очистити текст від символів та буквосполучень, які не зчитуються людиною, але можуть збивати алгоритм. До них належать: «@», «#», «https://» та інші. Очищення відбувалося за допомогою регулярних виразів (re – regular expressions)

```

import pandas as pd
import re
from tqdm import trange, tqdm_notebook, tqdm

d = pd.read_csv(tweets_clean_file)
for i,s in enumerate(tqdm(d['text'])):
    text = d.loc[i, 'text']
    text1 = str(text)
    text = text1.replace("#", "")
    text = re.sub('https?:\/\/(?:[-\w.]|(?:%[\da-fA-F]{2}))+', '', text,
flags=re.MULTILINE)
    text = re.sub('@\w+ *', '', text, flags=re.MULTILINE)
    d.loc[i, 'text'] = text
d.head()

```

| | Unnamed: 0 | date | timestamp | text |
|---|------------|---------------------|------------|---|
| 0 | 0 | 2021-04-26 13:34:02 | 1619436842 | Good Morning Crypto Twitter! A new week ahead ... |
| 1 | 1 | 2021-04-26 04:11:27 | 1619403087 | Well that 13 day bear market was fun. Now it's... |
| 2 | 2 | 2021-04-26 03:19:30 | 1619399970 | FTX had the largest market share growth of all... |
| 3 | 3 | 2021-04-26 02:48:13 | 1619398093 | The market is GREEN again Bitcoin Cryptocurre... |
| 4 | 4 | 2021-04-25 22:38:15 | 1619383095 | Ethereum vs BSC active addresses. Pretty amazi... |

Рис. 2.5 Зовнішній вигляд перших 5 семплів датасету з твітами після очищення тексту твітів від зайвих для алгоритму символів.

Як бачимо, в перших семплах з тексту зникли гештеги - #. Аналогічним чином всі інші семпли були очищені від вищезазначених символів та підготовані до sentiment analysis.

4. Видалення рядків з пустими клітинками

Останній етап очищення датасету з твітами – видалення рядків, що не містять значень. Вони викривлюють результат майбутньої моделі і не дозволяють коректно працювати алгоритму аналізу тексту. Ця операція була виконана за допомогою відповідного інструменту в пакеті Pandas.

```
df_clean = df_clean.dropna()
```

На цьому етапі підготовка датасету з твітами завершена. Наступний крок – підготувати датасет ціни на Ethereum до подальшого об'єднання.

Таблиця 1.2 Оригінальний датасет з ціною на Ethereum.

| S N o | Nam e | Sy mb ol | Dat e | High | Low | Open | Close | Volu me | Market cap |
|-------------|--------------|----------------|--|------------------------|------------------------|------------------------|------------------------|--------------|-------------------|
| 1 | Ethe reum | ET H | 201 5- 08- 08 23:5 9:59 | 2.79881000 51879883 | 0.71472501 75476074 | 2.79376006 1264038 | 0.75332498 55041504 | 6741 88.0 | 454868 94.2408 |
| 2 | Ethe reum | ET H | 201 5- 08- 09 23:5 9:59 | 0.87980997 56240845 | 0.62919098 1388092 | 0.70613598 82354736 | 0.70189702 51083374 | 5321 70.0 | 423995 73.4991 |
| 3 | Ethe reum | ET H | 201 5- 08- 10 23:5 9:59 | 0.72985398 76937866 | 0.63654601 57394409 | 0.71398901 93939209 | 0.70844799 28016663 | 4052 83.0 | 428183 64.3945 |

Для нашого дослідження достатньо лишити лише нам потрібна лише дата ['date'], та відсоткова зміна курсу за день. Перша змінна в нас є, а другу треба створити

1. Розрахунок відсоткової різниці ціни закриття та ціни відкриття

В нас є стовпець ['close'] з ціною закриття дня та ['open'] з ціною відкриття дня. Ми хочемо отримати відсоткову різницю ціни закриття та ціни відкриття і звести її в одну змінну. Щоб отримати цю змінну, ми обчислимо логарифмічну

різницю між ціною закриття та відкриття. Для цього використали команду `np.log()`.

```
df['log_diff'] = np.log(df['close']) - np.log(df['open'])
```

```
df['diff'] = df['close'] - df['open']
```

```
df
```

Логарифмічна шкала цін будується таким чином, що ціни в масштабі НЕ розташовані на однаковій відстані, в рівній мірі один від одного. Натомість міра побудована таким чином, що два рівні відсотки змін нанесені як однакові вертикальні відстані на шкалі.

Більшість технічних аналітиків і трейдерів використовують логарифмічні шкали цін. Звичайно повторювані відсоткові зміни представлені рівним інтервалом між числами у шкалі. Наприклад, відстань від 10 до 20 доларів дорівнює відстані від 20 до 40 доларів, оскільки обидва сценарії представляють 100% зростання ціни.

Логарифмічні шкали цін є кращими, ніж лінійні шкали цін, коли демонструють менш серйозні підвищення чи зниження цін. Вони можуть допомогти уявити, наскільки далеко повинна рухатися ціна, щоб досягти цілі купівлі чи продажу.

2. Створення змінної для логістичної регресії

Додаємо стовпчик `['target']`, який знадобиться для логістичної регресії. В ньому будуть лише значення 0 або 1. Якщо ціна виросла - 1, якщо впала або залишилася незмінною— 0. Виконаємо наступну команду і додаємо бажаний стовпчик до датасету.

```
df['target'] = [1 if log_diff > 0 else 0 for log_diff in df['log_diff']]
```

```
df
```

3. Конвертація дати до уууу-mm-dd

Так само, як і в датасеті з твітами, тут ми маємо час з точністю до хвилини. Аналогічно, за допомогою команди `str.slice()` ми вкоротили значення клітинок. Але є ще одна проблема. Нараз і датасеті з Ethereum дані в форматі `dd-mm-yyuu`, що є оберненим до формату датасету з твітами. Ми провели конвертацію за допомогою наступної команди :

```
df_eth['day'] = pd.to_datetime(df_eth['day'])
```

Отримали наступний датасет:

| | day | log_diff | close |
|-------------|------------|-----------------|--------------|
| 0 | 2021-04-24 | -0.067281 | 2213.350000 |
| 1 | 2021-04-23 | -0.012822 | 2367.390000 |
| 2 | 2021-04-22 | 0.017195 | 2397.940000 |
| 3 | 2021-04-21 | 0.011435 | 2357.060000 |
| 4 | 2021-04-20 | 0.075255 | 2330.030000 |
| ... | ... | ... | ... |
| 1117 | 2017-01-05 | -0.021484 | 78.447885 |
| 1118 | 2017-04-30 | 0.133368 | 80.337893 |
| 1119 | 2017-04-29 | -0.039785 | 70.375681 |
| 1120 | 2017-04-28 | 0.140298 | 73.491575 |
| 1121 | 2017-04-27 | 0.161382 | 63.775229 |

Рис. 2.6 Зовнішній вигляд частини датасету з ціною Ethereum з конвертованими датами.

2.4. Аналіз настрою

Щоб шукати кореляцію між емоційним забарвленням новин про криптовалюту та її курсом треба проаналізувати ті новини. Ручний аналіз великих масивів даних містить людський фактор і не є можливим. Для цього була використана технологія NLP – Natural Language Processing. Його найпростіша версія полягає в тому, що алгоритм аналізує текст на наявність слів-маркерів, які мають певну вагу від -1 до 1 (-1 – дуже негативний, 1 – дуже позитивний). Наприклад такі слова, як “love”, “enjoy”, “happy”, “like” мають позитивний настрій, а “hate”, “dislike”, “deteste” – негативний. Натомість, використана бібліотека розпізнає такі фрази, як “did not love”, як негативного характеру. Також вона розуміє, що слово “ENJOY” має більше ваги, ніж “enjoy”, оскільки користувачі інтернету використовують верхній регістр для підсилення значення окремих слів чи фраз. Ці ваги сумуються і видається середнє зважене для всього тексту твіта. Однією з найбільш відомих відкритих бібліотек NLP на Python є VADER (Valence Aware Dictionary and sEntiment Reasoner). Вона була застосована в наступній ділянці коду для аналізу зібраних твітів та виведення значення емоційного забарвлення в змінну [‘compound’].

Якщо порівнювати методи аналізу тексту використані в цій роботі з роботами інших авторів, то Адам Салак [6] також використав Python бібліотеку VADER, а Джермейн Камінський просто вибрав твіти, що містять слова-маркери і поділив їх на кілька груп: позитивні, негативні та ті, що відображають сумнів.

```
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
df_clean = df_clean.dropna()

analyzer = SentimentIntensityAnalyzer()
compound = []
for i,s in enumerate(tqdm(df_clean['text'])):
    if type(df_clean['text']) is float:
        continue
```

```
vs = analyzer.polarity_scores(s)
compound.append(vs['compound'])
df_clean['compound'] = compound
df_clean.head()
```

| | day | text | compound |
|---|------------|---|----------|
| 0 | 2021-04-26 | Good Morning Crypto Twitter! A new week ahead ... | 0.7263 |
| 1 | 2021-04-26 | Well that 13 day bear market was fun. Now it's... | 0.7627 |
| 2 | 2021-04-26 | FTX had the largest market share growth of all... | 0.9516 |
| 3 | 2021-04-26 | The market is GREEN again Bitcoin Cryptocurre... | 0.0000 |
| 4 | 2021-04-25 | Ethereum vs BSC active addresses. Pretty amazi... | 0.9081 |

Рис. 2.7 Огляд початку датасету після аналізу настрою

Останнім етапом підготовки даних є зведення двох датасетів в один. `df_tw` – датасет з твітами, `df_eth` – датасет з ціною на Ethereum. `on='day'` – параметр, в якому вказується спільна для двох датасетів ознака, за якою буде здійснюватися об'єднання.

```
df_merge = pd.merge(df_tw, df_eth, on='day')
df_merge.head()
```

| | day | text | compound | log_diff | close | target |
|---|------------|---|----------|-----------|---------|--------|
| 0 | 2021-04-24 | The Ethereum Transaction Volume is at a 3-year... | 0.0000 | -0.067281 | 2213.35 | 0 |
| 1 | 2021-04-24 | What do Visa, Deutsche Bank, ING, VanGuard, Ho... | 0.1860 | -0.067281 | 2213.35 | 0 |
| 2 | 2021-04-24 | This bull run is following the same patterns a... | 0.0000 | -0.067281 | 2213.35 | 0 |
| 3 | 2021-04-24 | Bitcoin Ethereum MassAdoption | 0.0000 | -0.067281 | 2213.35 | 0 |
| 4 | 2021-04-24 | Gemini users can now buy Bitcoin with Apple Pa... | -0.2714 | -0.067281 | 2213.35 | 0 |

Рис. 2.8 Зовнішній вигляд частини фінального датасету після об'єднання попередніх двох датасетів в один

2.5. Описовий аналіз даних

[day] – дата публікації твіту. Одночасно – дата, в яку була обрахована зміна ціни на Ethereum.

[text] – текст однієї публікації в Twitter акаунті «Cryptocurrency News» в день [day].

[compound] – математична інтерпретація емоційного забарвлення новин з [text]. Число в діапазоні від -1 до 1.

[log_diff] – логарифмічна різниця ціни закриття та ціни відкриття Ethereum в день [day]. Характеризує зміну курсу в обраний день.

[close] – ціна закриття Ethereum в день [day]

[target] – бінарна інтерпретація [log_diff]. Якщо ціна виросла = 1, якщо впала або залишилася незмінною = 0.

У роботі ми використовуємо дані в період з 01.01.2020 по 24.04.2021. Всього було взято 661 семплів.

2.6. Побудова однофакторної лінійної регресії

Отже, в даному розділі ми проведемо специфікацію для кінцевої моделі.

Кінцева модель містить наступні змінні:

- Залежна змінна [log_diff]

[log_diff] – логарифмічна різниця ціни закриття та ціни відкриття Ethereum в конкретний день, що вимірює дохідність активу.

- Незалежна змінна [compound]

Математична інтерпретація емоційного забарвлення тексту твіта в діапазоні від -1 до 1.

Загальний вигляд кінцевої моделі описується наступним рівнянням:

$$[\log_diff] = \beta_0 + \beta_1 * [\text{compound}]$$

Для побудови однофакторної лінійної регресії була використана Python-бібліотека scikit-learn. Це безкоштовна бібліотека створена ще в 2007 році для роботи з машинним навчанням та data science. [23]

Для побудови моделі обмежимо датасет до трьох стовпчиків, з яких нам потрібні два – [compound] [log_diff]

```
data = pd.read_csv('/Users/artemsakh/Documents/Jupyter/merge.csv')
```

```
data = data[['compound', 'log_diff', 'target']]
```

```
data
```

| | compound | log_diff |
|---|----------|-----------|
| 0 | 0.0000 | -0.067281 |
| 1 | 0.1860 | -0.067281 |
| 2 | 0.0000 | -0.067281 |
| 3 | 0.0000 | -0.067281 |
| 4 | -0.2714 | -0.067281 |

Рис. 2.9 Дві змінні необхідні для побудови лінійної регресії

За X візьмемо перший стовпчик зі змінною [compound], а за y – другий зі змінною [log_diff].

```
X = data.iloc[:, 0].values.reshape(-1, 1)
```

```
y = data.iloc[:, 1].values.reshape(-1, 1)
```

Далі, за допомогою команди **train_test_split** розіб'ємо датасет на дві частини – тренувальну та тестову. Під тестову частину виділимо 20% датасету.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,  
random_state=0)
```

Почнемо процес тренування моделі лінійної регресії на тренувальному датасеті наступною парою команд:

```
regressor = LinearRegression()
```

```
regressor.fit(X_train, y_train)
```

І спробуємо використати її для передбачення курсу Ethereum

```
y_pred = regressor.predict(X_test)
```

Варто зазначити, що дослідження Адама Салака [6] також базувалося на однофакторній лінійній регресії.

2.10. Побудова логістичної регресії

- Залежна змінна [target]

[target] – бінарна інтерпретація [log_diff] (логарифмічної різниці ціни закриття та ціни відкриття Ethereum). Якщо ціна виросла = 1, якщо впала або залишилася незмінною = 0.

- Незалежна змінна [compound]

Математична інтерпретація емоційного забарвлення тексту твіта в діапазоні від -1 до 1.

Загальний вигляд кінцевої моделі описується наступним рівнянням:

$$[\text{target}] = \beta_0 + \beta_1 * [\text{compound}]$$

Для побудови моделі обмежимо датасет до трьох стовпчиків, з яких нам потрібні два – [compound] [target]

```
data = pd.read_csv('/Users/artemsakh/Documents/Jupyter/merge.csv')
```

```
data = data[['compound', 'target']]
```

```
data
```

| | compound | target |
|----------|-----------------|---------------|
| 0 | 0.0000 | 0 |
| 1 | 0.1860 | 0 |
| 2 | 0.0000 | 0 |
| 3 | 0.0000 | 0 |
| 4 | -0.2714 | 0 |

Рис. 2.10 Дві змінні, необхідні для побудови логістичної регресії

За X візьмемо перший стовпчик зі змінною [compound], а за y – другий зі змінною [target].

```
X = data.iloc[:, 0].values.reshape(-1, 1)
```

```
y = data.iloc[:, 1].values.reshape(-1, 1)
```

Тепер, командою **train_test_split** розбиваємо датасет на дві частини – тренувальну та тестову. Під тестову частину виділяємо 20% датасету.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,  
random_state=0)
```

Запустимо процес тренування моделі логістичної регресії на тренувальному датасеті наступною парою команд:

```
regressor = LinearRegression()
```

```
regressor.fit(X_train, y_train)
```

І спробуємо використати її для передбачення курсу Ethereum

```
y_pred = regressor.predict(X_test)
```

РОЗДІЛ 3

ОЦІНКА ТА АНАЛІЗ РЕЗУЛЬТАТІВ

3.1. Результати оцінювання моделі лінійної регресії

Для оцінювання результату нашої моделі порівняємо передбачені дані зі справжніми і виведемо в таблицю, де в стовпчику Actual знаходяться фактичні дані, а в стовпчику Predicted – передбачені. Використаємо наступні команди:

```
df = pd.DataFrame({'Actual': y_test_f.tolist(), 'Predicted': y_pred_f.tolist()})
```

```
df.head()
```

| | Actual | Predicted |
|---|-----------|-----------|
| 0 | -0.044232 | 0.003307 |
| 1 | -0.042333 | -0.002838 |
| 2 | 0.017195 | 0.006740 |
| 3 | -0.009903 | 0.007766 |
| 4 | -0.086306 | -0.004239 |

Рис. 3.1 Результат прогнозування лінійної регресії у вигляді таблиці з двома змінними: перший стовпчик – справжні дані, другий – прогнозовані

Виведемо базові характеристики моделі за допомогою команд на Python.

```
print(regressor.intercept_)
```

Коефіцієнт перетину = 0.00469303

```
print(regressor.coef_)
```

Коефіцієнт нахилу = 0.00712672

```
print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, y_pred))
```

```
print('Mean Squared Error:', metrics.mean_squared_error(y_test, y_pred))
```

```
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))
```

Mean Absolute Error: 0.03950816814402355

Mean Squared Error: 0.0031496364228106896

Root Mean Squared Error: 0.056121621705103014

Коефіцієнт детермінації

```
r_sq = regressor.score(X, y)
```

```
print('coefficient of determination:', r_sq)
```

coefficient of determination: 0.00612556861013025

Для візуалізації даних використаємо бібліотеку `matplotlib`. [24] Нижче бачимо зображення двох графіків. Синій графік – справжні значення, а оранжевий – передбачені.

```
import matplotlib.pyplot as plt
```

```
df.plot(figsize=(15,10)).get_figure().savefig('linear-actual-predicted.png')
```

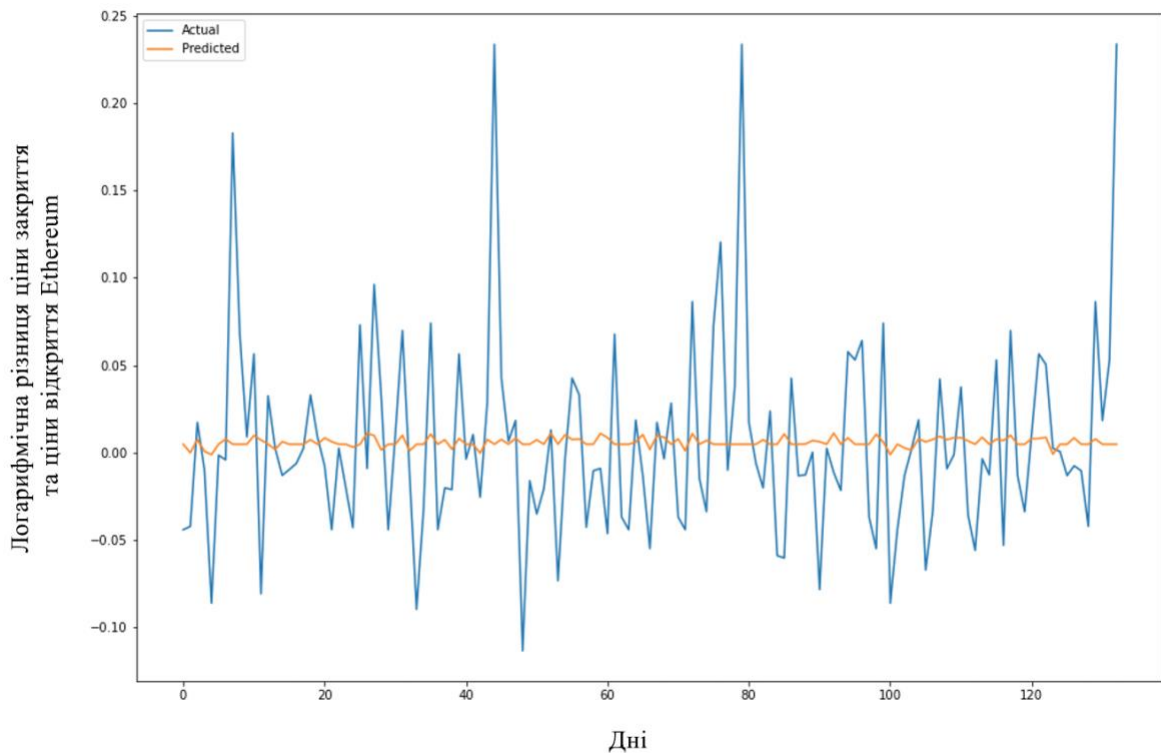


Рис. 3.2 Результат прогнозування лінійної регресії у вигляді графіку: синя лінія – справжні дані, оранжева – прогнозовані

З отриманих показників можемо сказати, що якщо змінна настрою зростає на 1%, то ціна на Ethereum зростає на 0.7 %. Якщо новини в Twitter нейтральні, тобто настроїв = 0, то на основі коефіцієнту перетину ринок все одно зростає на 0.47%. Оскільки коефіцієнт детермінації нашої моделі вкрай низький, то можемо зробити висновок, що кореляції між емоційним забарвленням новин про криптовалюту та її курсом немає.

Варто зазначати, що такі результати збігають з висновками, наведеними в роботі Адама Салака [6]. В його роботі також була використана однофакторна лінійна регресія і кореляція була відсутня.

3.2. Результати оцінювання моделі логістичної регресії

Щоб оцінити якість нашої моделі порівняємо передбачені дані зі справжніми і виведемо в таблицю, де в стовпчику Actual знаходяться фактичні дані, в стовпчику Predicted – передбачені, а в стовпчику predicted-polarized – поляризовані значення зі стовпчику Predicted. Використаємо наступні команди:

```
df = pd.DataFrame({'Actual': y_test_f.tolist(), 'Predicted': y_pred_f.tolist()})
```

```
df_pol = df
```

```
df_pol['predicted_polarized'] = [1 if Predicted > 0.5 else 0 for Predicted in df_pol['Predicted']]
```

```
df_pol.head()
```

| | Actual | Predicted | predicted_polarized |
|---|--------|-----------|---------------------|
| 0 | 0 | 0.494072 | 0 |
| 1 | 0 | 0.518429 | 1 |
| 2 | 1 | 0.480470 | 0 |
| 3 | 0 | 0.514533 | 1 |
| 4 | 0 | 0.523981 | 1 |

Рис. 3.3 Результат прогнозування лінійної регресії у вигляді таблиці з двома змінними: перший стовпчик – справжні дані, другий – прогнозовані

Виведемо базові характеристики моделі за допомогою наступних команд:

```
print(regressor.intercept_)
```

Коефіцієнт перетину = 0.4940723

```
print(regressor.coef_)
```

Коефіцієнт нахилу = -0.03577637

```
print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, y_pred))
```

Середнє відхилення: 0.4999434848842031

```
print('Mean Squared Error:', metrics.mean_squared_error(y_test, y_pred))
```

Середнє квадратичне відхилення 0.2503870768902956

```
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))
```

Корінь з середнього квадратичного відхилення: 0.5003869271776548

Знайдемо коефіцієнт детермінації за допомогою наступної команди:

```
r_sq = regressor.score(X, y)
```

```
print('coefficient of determination:', r_sq)
```

coefficient of determination: -0.00042896641676426484

Для візуалізації даних використаємо бібліотеку matplotlib. Нижче бачимо зображення двох графіків. Синій графік – справжні значення, оранжевий – передбачені, а зелений – передбачені поляризовані.

```
import matplotlib.pyplot as plt
```

```
df.plot(figsize=(15,10)).get_figure().savefig('logit-actual-predicted-pol.png')
```

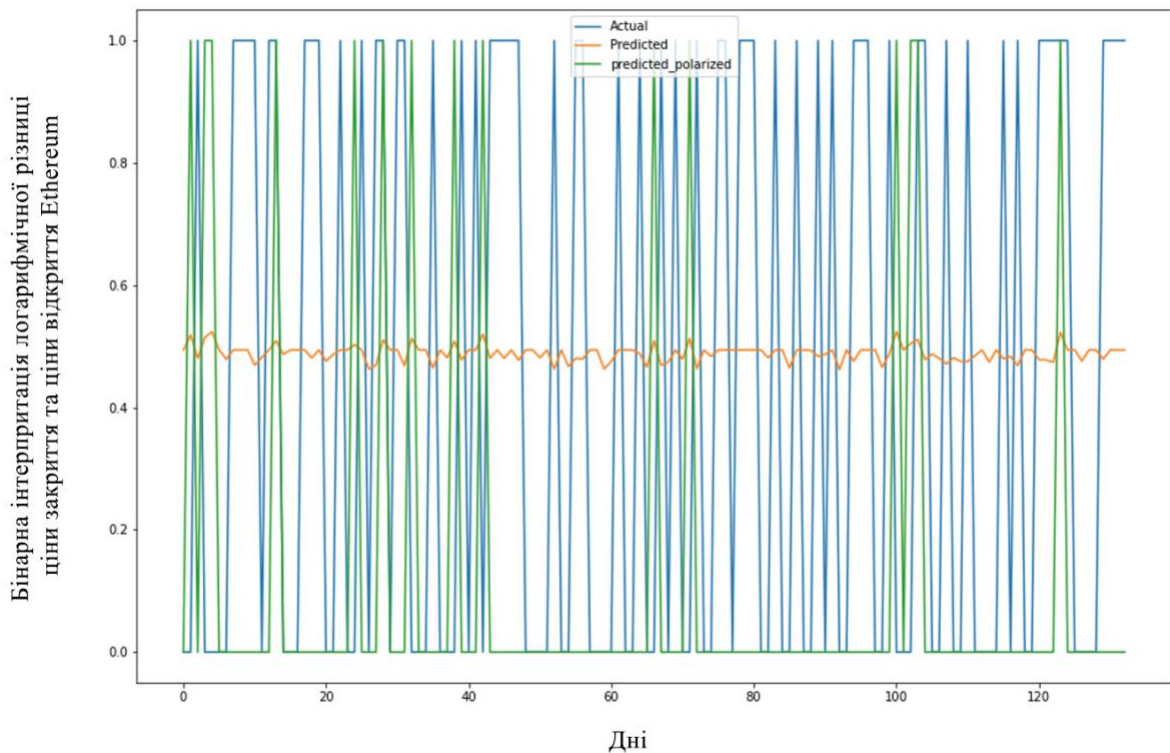


Рис. 3.4 Результат прогнозування лінійної регресії у вигляді графіку: синя лінія – справжні дані, оранжева – прогнозовані, зелена – прогнозовані поляризовані

Згідно з підрахунками модель змогла передбачити курс Ethereum в 49,62% випадків, що не є достатнім показником, щоб вважати її вдалою.

При зміні настрою на 1% курс Ethereum зменшується на 3%.

Оскільки коефіцієнт детермінації нашої моделі вкрай низький, то можемо зробити висновок, що кореляції між емоційним забарвленням новин про криптовалюту та її курсом немає.

ВИСНОВКИ

Отже, Ethereum – це децентралізована програмна платформа з відкритим кодом. В її основі лежить блокчейн, а сама вона використовується для власної криптовалюти під назвою “ETH”. Платформа використовується для створення децентралізованих фінансових систем (DeFi), децентралізованих програм (DApps), а також смарт-контрактів. Ethereum був випущений 2015 року.

“Блокчейн – це децентралізована, розподілена книга, що реєструє походження цифрового активу.” [3]

«Блокчейн – повна і незмінна історія транзакцій децентралізованої спільноти, з якою згодні всі учасники спільноти.» [5] Цей реєстр оновлюється автоматично в реальному часі, приймається спільнотою і зберігається на комп’ютері кожного учасника. Таким чином відпадає необхідність в центральному контролюючому органі, оскільки жоден учасник не зможе виконати подвійну витрати. Вона одразу призведе до конфлікту в історії кожного учасника.

Криптовалюта - це цифрова валюта, базована на децентралізованій системі блокчейну. Може слугувати як платіжним засобом, так і інвестиційним інструментом. Інтернаціональна, не регулюється жодною державою. Не потребує державної ідентифікації для використання.

Криптовалютна біржа

1. Тип активу: на крипто-біржі торгують цифровими валютами, такими як Біткоїн, Ethereum та інші.
2. Цінність: цінність суб'єктивна. Іноді підкріплена перспективною технологією.
3. Випуск активів: в більшості мають обмежену емісію.
4. Урегульованість: більша частина діяльності криптовалютного ринку залишається нерегульованою
5. Волатильність: ключова відмінність ринку цифрової валюти - висока волатильність та малі об'єми.

6. Доступність: почати торгувати можна за лічені хвилини. Брокер не потрібен.

Фундаментальний аналіз. Головний компонент - реальна цінність активу. Фундаментальний аналітик звертає увагу на новини про актив, стан криптовалютного сектору, внутрішнє та глобальне економічне середовище, щоб визначити реальну цінність активу.=

Технічний аналіз. Головний компонент - історія ціни активу. Технічний аналітик використовує математичне та статистичне моделювання, щоб зрозуміти ринкову поведінку і передбачити цінові коливання.

Ethereum - Платформа для створення децентралізованих онлайн-сервісів на базі блокчейну (децентралізованих додатків), що працюють на базі смарт - контрактів. Ефір - цифрова валюта на базі Ethereum.

Ринкові показники Ефіру наступні:

- Ціна однієї монети станом на 30 травня 2022 року = 1924 доларів
- Ринкова капіталізація = 232.8 млрд доларів
- Валюти в обігу = 120.98 млн монет (необмежена емісія)
- Обсяги торгів за останню добу = 16 277 951 доларів
- Зміна ціни за останню добу = +6.8%
- Зміна ціни за останній тиждень = -6.3%
- Зміна ціни за останній місяць = -31.5% [11]

Датасет. В датасеті зібрано близько тисячі твітів одного з найбільших новинних ресурсів спеціалізованих на криптовалюті – CryptoBoom News.

Була проведена чистка датасету та аналіз настрою кожної новини. Після цього були побудовані дві моделі: лінійної та логістичної регресії залежності ціни на Ethereum від настрою новин.

Лінійна регресія.

Рівняння: $[\log_diff] = \beta_0 + \beta_1 * [compound]$

Результат: з отриманих показників можемо сказати, що якщо змінна настрою зростає на 1%, то ціна на Ethereum зростає на 0.7 %. Якщо новини в Twitter нейтральні, тобто настрої = 0, то на основі коефіцієнту перетину ринок все одно зростає на 0.47%. Оскільки коефіцієнт детермінації нашої моделі вкрай низький, то можемо зробити висновок, що кореляції між емоційним забарвленням новин про криптовалюту та її курсом немає. Такі висновки сходяться з результатами дослідження Адама Салака [6]. В його роботі також була використана однофакторна лінійна регресія і кореляція була відсутня. Це дозволяє стверджувати, що в двох найпопулярніший криптовалютних соціальних мережах не була помічена залежність між емоційним забарвленням публікацій та курсом ЕТН. Але в роботі Джермейн Камінського [7] було з'ясовано, що дописи користувачів Twitter впливають на курс Bitcoin.

Логістична регресія.

Рівняння: $[\text{target}] = \beta_0 + \beta_1 * [\text{compound}]$

Результат: згідно з підрахунками модель змогла передбачити курс Ethereum в 49,62% випадків, що не є достатнім показником, щоб вважати її вдалою. При зростанні настрою на 1% курс Ethereum зменшується на 3%. Оскільки коефіцієнт детермінації нашої моделі вкрай низький, то можемо зробити висновок, що кореляції між емоційним забарвленням новин про криптовалюту та її курсом немає.

З огляду на роботи інших авторів на подібну тему, можна сказати, що висновки неоднозначні. В дослідженнях одного автора кореляція була присутня, а в дослідженні іншого – відсутня. Це ще раз підтверджує доцільність нашого дослідження.

Отже, ми визначили, як працює блокчейн, що таке криптовалюта та в чому різниця між фондовою та криптовалютною біржами.

Провівши дослідження та побудувавши лінійну та логістичну регресії ми визначили, що кореляція між настроєм новин про криптовалюту та її курсом відсутня.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Alan T. Norman, Cryptocurrency Investing Bible: The Ultimate Guide About Blockchain, Mining, Trading, ICO, Ethereum Platform, Exchanges, Top Cryptocurrencies for Investing and Perfect Strategies to Make Money, 2018
2. Stephen Satoshi. Cryptocurrency: Ultimate Beginners Guide to Making Money with Cryptocurrency like Bitcoin, Ethereum and altcoins , 2017
3. Daniel Drescher. Blockchain Basics (apress 2017)
4. Marcus Demaria. A Beginner's Guide to Profiting from Cryptocurrencies, 2017
5. Джулиан Хосп. “О криптовалюте просто. Биткоин, эфириум, блокчейн, децентрализация, майнинг, ICO & Co”
6. Adam Salač, Forecasting of the cryptocurrency market through social media sentiment analysis
7. Jermain C. Kaminski, Nowcasting the Bitcoin Market with Twitter Signals
8. Birz, G., & Lott Jr, J. R. (2011). The effect of macroeconomic news on stock returns: New evidence from newspaper coverage. *Journal of Banking & Finance*, 35(11), 2791-2800.
9. Steinert, Lars, and Christian Herff. "Predicting altcoin returns using social media." *PloS one* 13.12 (2018): e0208119.
10. Kirlić, A., Orhan, Z., Hasovic, A., & Kevser-Gokgol, M. (2018). Stock market prediction using Twitter sentiment analysis. *Invention Journal of Research Technology in Engineering & Management (IJRTEM)* , 2(1), 01-04.
11. Ethereum – CoinMarketCap. URL: <https://coinmarketcap.com/currencies/ethereum/>

12. Bitcoin – CoinMarketCap. URL: <https://coinmarketcap.com/currencies/bitcoin/>
13. Number of social network users worldwide from 2017 to 2025. URL: <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>
14. Що таке блокчейн технологія: це не лише про криптовалюту. URL: <https://futurenow.com.ua/shho-take-blokchejn-tehnologiya-tse-ne-lyshe-pro-kryptovalyutu/>
15. Must-know: A glossary of technical analysis terms. By Gordon Kristopher. URL: <https://marketrealist.com/2014/12/must-know-glossary-technical-analysis-terms/>
16. Ethereum – Investopedia. URL: <https://www.investopedia.com/terms/e/ethereum.asp#:~:text=Launched%20in%202015%2C%20Ethereum%20is,interference%20from%20a%20third%20party.>
17. Cryptocurrency News Tweet – OpenBlender. URL: <https://www.openblender.io/#/dataset/explore/5ea209c495162936348f13ebjhjhjh>
18. CryptoBoom News. URL: <https://cryptoboomnews.com>
19. Cryptocurrency News – Twitter. URL: https://twitter.com/CryptoBoomNews?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor
20. Python. URL: <https://www.python.org>
21. Vader (Valence Aware Dictionary and sEntiment Reasoner). URL: <https://pypi.org/project/vaderSentiment/>
22. Pandas Python. URL: <https://pandas.pydata.org>
23. Scikit-learn - Machine Learning in Python. URL: <https://scikit-learn.org/stable/>
24. Matplotlib. URL: <https://matplotlib.org>

25. Blockchain explained – Investopedia. URL:
<https://www.investopedia.com/terms/b/blockchain.asp>