

УДК 519.688; 004.89; 004.9

О. Гавриленко, М. Мягкий

АЛГОРИТМ ПРОГНОЗУВАННЯ КУРСУ КРИПТОВАЛЮТИ З УРАХУВАННЯМ ВПЛИВУ РАНЖОВАНОЇ ГРУПИ ЕКСПЕРТІВ В СОЦІАЛЬНИХ МЕРЕЖАХ

Анотація: В даній роботі представлено алгоритм для вивчення рівня впливу дописів ранжованої групи експертів в соціальних мережах на курс криптовалюти.

Алгоритм дозволяє підвищити точність прогнозів, порівняно з однок експертними алгоритмами.

Ключові слова: курс криптовалюти, алгоритм прогнозування, дописи в соціальних мережах, ранжування групи експертів, рекомендаційна система, машинне навчання, статистичний аналіз, інформаційна технологія інтелектуального аналізу.

Вступ

На даний момент наявна величезна кількість публікацій і різної інформації стосовно різноманітних криптовалют, що призводить до постійного зростання популярності даної тематики серед всіх вікових категорій людей. Це обумовлено відносною легкістю входження на криптовалютні біржі для подальшої взаємодії з криптовалютами. Користувачам більшості криптовалютних бірж надається досить широкий спектр фінансових операцій і інших можливостей. Найбільш простими і через це тими, що мають великий попит є: купівля і продаж криптовалюти. Дані операції потребують від користувача мінімальних навичок, однак також вони не мають високої ліквідності. Тобто, біржа не забезпечує користувачів умовами для стрімкого збагачення через використання базових навичок. Втім, розробники і власники криптовалюти зацікавлені в якомога більшому поширенні власної продукції тому біржі, в свою чергу, намагаються стимулювати людей до використання більш «просунутих» технологій і можливостей які передбачені системою. Певні інструменти дозволяють проаналізувати стан ринку відносно відібраної криптовалюти і тим самим сформулювати теорію про те, яким буде її подальший курс. Дані інформаційні маніпуляції після поширення в соціальних мережах заохочують багатьох користувачів до відповідних дій.

Таким чином, актуальність даного дослідження зумовлена зростанням і без того великої кількості інформації відносно криптовалют, яка направлена саме на спонукання людей до відповідних дій відносно певних криптовалют.

Проте, більшість існуючих криптовалютних бірж на поточний момент не мають в своєму арсеналі інструментів які б дозволили перевірити і дослідити інформацію, яка поширюється в соціальних мережах стосовно тієї чи іншої

криптовалюти. Це дозволяє проводити відповідне дослідження з подальшою розробкою відповідного інструменту, який, за умови коректного використання, буде надавати користувачам рекомендаційні поради відносно подальших дій відносно досліджуваної в системі криптовалюти. Базуючись на цих порадах, зацікавлені особи зможуть коригувати свої рішення відносно подальших фінансових кроків.

В основі загальної більшості рекомендаційних систем завжди лежить необхідність встановлення деяких факторів впливу, яким пізніше надаються певні вагові коефіцієнти для сприяння і спрощення формулювання подальших порад для користувачів.

В нашому випадку, досліджується саме вплив публікацій відомих осіб на формування цін за певну криптовалюту в певний момент часу. Важливість і наявність даного впливу було доведено авторами цієї роботи в попередніх дослідженнях.

Аналіз літературних джерел та постановка проблеми

Загалом, збір і обробка даних з публікацій відносно криптовалют з соціальних мереж і інших інформаційних платформ мережі є достатньо значущим завданням, оскільки вони можуть надати багато корисних і вичерпних відомостей котрі будуть необхідними для побудови якісної математичної моделі для подальшого прогнозування.

У роботі [1] розглядається процес комп'ютерного визначення та категоризації думок, висловлених у фрагменті тексту, з метою встановлення, чи є ставлення письменника до певної теми, продукту, тощо: позитивним, негативним чи нейтральним. У наведеному дослідженні, проведено детальне вивчення: аналізу настроїв і їх причинно-наслідкового зв'язку. Також за допомогою аналізу настроїв, визначено узагальнену подію на його основі та з урахуванням часу. Результати аналізу причинно-наслідкового зв'язку можуть бути використаними не тільки для визначення причин і наслідків, але також для подальшого прогнозування настроїв користувачів. Основна частина публікації – огляд комбінування даних підходів, які поєднуються в єдину модель, котра дозволяє визначити настрої під час майбутніх подій, а також створити часовий прогноз щодо довжини інтервалу між певними подіями. Для оцінки точності тут використовувалася - середня відносна похибка.

Для пошуку публікацій котрі б відповідали певним вимогам, як от: кількість, єдиний текстовий формат і інші, потрібно вибрати певну соціальну мережу. Висунутим вимогам добре відповідає месенджер – Twitter (на поточний момент X), у роботі [2] детально розглянуто спеціальний лінгвістичний аналіз та статистику в зазначеній соціальній мережі. Головною метою дослідження, автори зазначили, виявлення злочинних елементів в Сполучених Штатах, шляхом моделювання тем для обговорення та подальшого включення їх у модель прогнозування злочинності.

Проведено ґрунтовний аналіз впливу публікацій у соцмережах відносно потенційної можливості настання в майбутньому певних злочинних дій.

У роботі [3] була розглянута комплексна довідка для дослідників і практиків, а також охоплення всіх сфер, які сприяють побудові та аналізу соціальних мереж.

Робота [4] є досить актуальною на сьогодні у зв'язку зі складною епідеміологічною ситуацією у світі. В ній було проаналізовано мікроблоги в Twitter і запропонував кілька методів надійної класифікації твітів на основі їх настрою та візуалізації даних за допомогою інтерактивної карти в реальному часі.

В роботі [5] було виявлено осіб, які впливають на процеси обміну знаннями через внутрішню соціальну мережу, і спрогнозувати майбутні потоки знань, які можуть перетнути її. Використовуються дослідницькі дослідження та розроблено чотирифазову методологію, яка поєднує аналіз соціальних мереж із структурним моделюванням.

Для бізнесу та фінансового ринку процес аналізу великих об'ємів даних і розуміння потреб більшості людей є дуже важливим, оскільки це безпосередньо впливає на дохід компанії та окремих людей. Дослідження що проводилось в рамках роботи [6] було спрямоване на виявлення домінуючих факторів, які призводять до валютних криз. Також дана публікація призначена для виявлення та характеризування валютних криз, а також прогнозування потенційного настання останніх на ранній стадії. Це заощадить менеджерам деякий час на покращення політики управління кризою та коригування відповідних дій.

У роботі [7] досліджувалася динаміка лінійних і нелінійних, серійних залежностей у фінансових часових рядах, у рамках рухомого вікна. Зокрема, увага зосереджена на виявленні епізодів статистично значущої дво- та трибальної кореляції, в прибутковості кількох провідних валютних курсів, які можуть запропонувати певний потенціал для їх передбачуваності. Підхід рухомого вікна було використано для того, щоб зафіксувати динаміку кореляції для різних довжин вікон і проаналізувати розподіл періодів зі статистично значущими кореляціями. Виявлено, що для досить великих довжин вікон ці розподіли добре відповідають степеневому закону. Виміряно саму передбачуваність за показником попадання, тобто рівнем узгодженості між ознаками фактичної прибутковості та їхніми прогнозами, отриманими за допомогою простого предиктора на основі кореляції.

Слід зазначити, що в усіх наведених роботах дослідження носять загальний характер і не наводилися результати прогнозування саме курсів валют, зокрема криптовалют. Відповідно, не досліджувалися фактори, які впливають на них.

В роботі [8] проведено дослідження основних макроекономічних показників впливу на курс долару США в Україні: купівля/продаж готівкової валюти,

купівля/продаж безготівкової валюти, сальдо купівлі/продажу готівкової та безготівкової валюти, інфляція поточного року, номінальний та реальний ВВП, купівля/продаж клієнтами банків, операції між банками, валові та чисті міжнародні резерви, рівень безробіття, облікова (відсоткова) ставка, сальдо валютних інтервенцій, обсяг угод номінальної вартості. За допомогою методу головних компонент визначено основні економічні складові формування курсу. За допомогою статистичних моделей ARIMA, Exponential Smoothing та SSA спрогнозовано значення обраних факторів впливу. Спрогнозовано значення курсів валют за допомогою регресійних моделей побудованих алгоритмами Fast Tree, Fast Forest, Fast Tree Tweedie та Gam, а також проведено дослідження отриманих значень на точність. В даній роботі не проводилося прогнозування курсів саме криптовалют і не досліджувався вплив такого фактору, як публікації в соціальних мережах.

В роботі [9] проведено аналіз методів, сфер застосування та підходів до аналізу публікацій та прогнозування подій на основі зібраних даних, а також представлено модель оцінки впливу публікацій на зміну курсу криптовалюти з урахуванням дописів лише одного експерта. Подано обґрунтування актуальності теми та описано можливості відповідного застосування результатів роботи. Визначено основні етапи роботи з даними прогнозування подій, а саме: попередня обробка даних, їх подальший аналіз та прогнозування. В даній роботі не досліджувався рівень впливу публікацій групи експертів в соціальних мережах на курс криптовалюти.

В роботі [10] представлено алгоритм оцінки впливу публікацій «основного експерта» на курс криптовалюти. В даній роботі наведено процес визначення найбільш впливового («основного») експерта та отримано прогнози з урахуванням його дописів на курс обраної криптовалюти. В даній роботі не досліджувався рівень впливу публікацій усіх обраних експертів в соціальних мережах на курс криптовалюти. Але наведений алгоритм є перехідним етапом до алгоритмів прогнозування курсу криптовалюти з урахуванням дописів кожного з обраної групи експертів.

В роботі [11] представлено алгоритм оцінки послідовного впливу дописів групи експертів публікацій на курс криптовалюти АПУДГЕСМ (алгоритм прогнозування з послідовним урахуванням дописів групи експертів в соціальних мережах). Запропонований алгоритм є ітеративним. На кожному його етапі враховувалися дописи одного експерта в порядку слідування у ранжованому списку. Тобто на кожному етапі попередній прогноз доуточнюється у разі отримання незадовільної похибки.

В рамках досліджень наведених в роботах [8-11] було створено відповідні інформаційні системи для реалізації вищеописаних задач інтелектуального аналізу даних.

Наведений вище аналіз свідчить про те, що вплив окремих факторів на курс криптовалют, зокрема вплив дописів відомих людей у соціальних мережах, на сьогоднішній день все ще не достатньо розвинений і потребує подальшого дослідження.

Мета і задачі дослідження

Метою дослідження є розробка алгоритму для вивчення рівня впливу дописів кожного з обраної групи експертів в соціальних мережах на курс криптовалюти.

Даний алгоритм потенційно надасть можливість підвищити надійність прогнозу відносно курсу обраної криптовалюти з метою подальшого формування рекомендацій.

Для досягнення мети були поставлені такі завдання:

- відібрати групу експертів;
- провести ранжування обраної групи експертів, щодо рівня впливу на курс зазначеної криптовалюти;
- отримати прогноз відносно курсу деякої криптовалюти з урахуванням дописів в певній соціальній мережі відібраної групи експертів на основі попередньо проведеного ранжування експертів;
- сформулювати рекомендації щодо фінансових операцій над криптовалютою на основі отриманих прогнозів.

Матеріали та методи дослідження

Об'єктом дослідження є прогноз курсу криптовалют.

Інформація, яка потрібна для того, щоб проаналізувати рівень впливу дописів в соціальних мережах на курс криптовалют – це перелік експертів, рівень впливу яких буде досліджуватися, інтервал часу дослідження, кількість дописів, що зробив кожен з розглянутих експертів за вказаний період часу, а також реальні курси криптовалют за відповідний період.

В якості експертів обиралися відомі особистості, які є як обізнаними в галузі фінансів в цілому та криптовалют зокрема, або діяльність яких так чи інакше пов'язана з певною криптовалютою.

Орієнтовні зразки дата-сетів наведено в табл. 1 та табл. 2.

В табл. 1 наведено курси обраної криптовалюти x_1, x_2, \dots, x_s в обраний проміжок часу $[t_1; t_s]$, які можна взяти з сайту крипто біржі Binance [12].

В табл. 2 m_1, m_2, \dots, m_n – частоти дописів експертів, k_1, k_2, \dots, k_n – частоти дописів експертів у мережі Telegram, які пов'язані з певною криптовалютою.

Таблиця 1. Значення курсу обраної криптовалюти за вказаний проміжок часу

Момент часу	Значення курсу
...	...
t_1	x_1
t_2	x_2
...	...
t_s	x_s
...	...

Таблиця 2. Кількість дописів обраних експертів за вказаний проміжок часу

Експерт	Кількість дописів	Кількість дописів, пов'язаних з криптовалютою
Експерт 1	m_1	k_1
Експерт 2	m_2	k_2
...
Експерт n	m_n	k_n

Сформовані таким чином дані є вхідною інформацією для даного дослідження. В подальшому потрібно:

- проаналізувати кількість дописів обраних експертів в соціальній мережі;
- провести ранжування групи експертів;
- отримати прогноз курсу криптовалюти з урахуванням дописів обраної групи експертів після їх ранжування;
- сформулювати рекомендації щодо фінансових операцій відносно обраної криптовалюти.

Застосування математичного апарату на основі формул повної ймовірності та Байєса надає змогу використати цю інформацію для ранжування обраної групи експертів, в залежності від ймовірності, з якою зробить кожен експерт дописи в соціальній мережі в розглянутий період часу [10].

Прогнозування курсів обраної криптовалюти здійснюється за допомогою алгоритму, який є узагальненням алгоритму АУДСМ (алгоритм прогнозування з урахуванням дописів в соціальних мережах) [9]. Для контролю точності прогнозів обчислюється середня відносна похибка.

Рекомендації щодо фінансових операцій з криптовалютою формуються за допомогою введення критичного значення курсу та обчислення середнього арифметичного курсів криптовалюти за вказаний період часу.

Алгоритм прогнозування курсу криптовалюти з урахуванням дописів ранжованої групи експертів в соціальних мережах (АУДРГЕСМ)

Схему алгоритму прогнозування курсу криптовалюти з урахуванням дописів ранжованої групи експертів в соціальних мережах представлено на рис. 1.

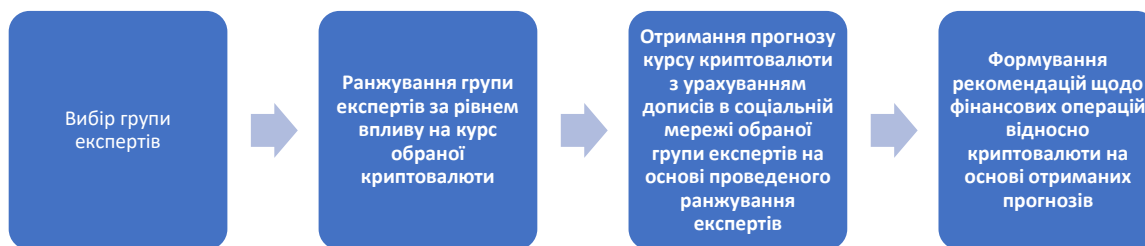


Рисунок 1. Схема АУДРГЕСМ

Відбір групи експертів

Постановка задачі:

З множини користувачів соціальної мережі обираємо підмножину $A = (a_1, a_2, \dots, a_n)$ користувачів, які задовольняють наступним вимогам [13]:

- користувач має бути публічною особою;
- кожен з користувачів має бути активним в обраній соціальній мережі, а також мати значну кількість підписників;
- користувачі мають різну професійну діяльність;
- основна професійна діяльність користувачів так чи інакше пов'язана з використанням криптовалюти;
- кожна пара користувачів a_i та a_j , $i, j = 1, 2, \dots, i \neq j$, не підтримує спілкування в обраній соціальній мережі (не є друзями та не реагують на пости одне одного);
- кожен користувач має достатній кваліфікаційний рівень в фінансовій сфері.

Назвемо цих користувачів експертами. Надалі їх дописи в обраній соціальній мережі будемо враховувати при прогнозуванні курсу певної криптовалюти.

Нехай за деякий період часу експерти зробили m дописів в соціальній мережі, причому з них, k дописів пов'язані з певною криптовалютою. Контекст дописів вважаємо довільним. Причому експерт a_1 за вказаний період часу опублікував m_1 постів, з яких k_1 постів пов'язані з певною криптовалютою, експерт a_2 опублікував m_2 пости, з яких k_2 пости пов'язані з певною криптовалютою і т.д. Останній експерт a_n опублікував m_n постів, з яких k_n постів пов'язані з певною криптовалютою. Тоді:

$$\begin{aligned} m_1 + m_2 + \dots + m_n &= m, \\ k_1 + k_2 + \dots + k_n &= k, \end{aligned} \quad (1)$$

де m_1, m_2, \dots, m_n – частоти дописів експертів, k_1, k_2, \dots, k_n – частоти дописів експертів, які пов'язані з певною криптовалютою.

Необхідно для довільного інтервалу часу підрахувати частоти дописів усіх відібраних експертів [14].

Обґрунтування:

Такий вибір експертів обумовлений потребою у формуванні множини таких експертів, які будуть незалежними один від одного як в просторі обраної соціальної мережі, так і в професійному просторі.

Ранжування групи експертів за рівнем впливу на курс обраної криптовалюти

Постановка задачі:

На основі переліку експертів: $A = (a_1, a_2, \dots, a_n)$, отриманого у пункті 5.1. та враховуючи частоти їх дописів в обраній соціальній мережі за вказаний проміжок часу – m_1, m_2, \dots, m_n (також k_1, k_2, \dots, k_n), наведені в таблиці 2, необхідно провести ранжування експертів на предмет їх впливу на курс обраної криптовалюти.

Сформульована задача легко інтерпретується як класична ймовірнісна задача: за деякий період часу, було написано m постів. Відомо, що за цей період публікували дописи n експертів, причому m_1, m_2, \dots, m_n – частоти дописів експертів, k_1, k_2, \dots, k_n – частоти дописів експертів, які пов'язані з обраною криптовалютою. Подія B полягає у тому, що в довільний період часу один з експертів написав пост, пов'язаний з обраною криптовалютою. Необхідно визначити, який саме експерт з більшою ймовірністю зробив цей допис [14].

Обґрунтування:

Подія B – в довільний момент часу t з інтервалу $[0; T]$ (в таблиці 1 $t_1 = 0$, $t_s = T$) був написаний пост, пов'язаний з обраною криптовалютою.

Гіпотеза H_1 – пост написав експерт 1,

гіпотеза H_2 – пост написав експерт 2,

...

гіпотеза H_n – пост написав експерт n .

Вважаємо, що події H_i та H_j , $i, j = 1, 2, \dots, n$, $i \neq j$ є попарно незалежними. Ці припущення можна зробити на основі переліку вимог, яким повинні відповідати експерти (див. пункт 5.1).

За формулою повної ймовірності:

$$P(B) = \sum_{i=1}^n (P(H_i) \cdot P(B/H_i)) \quad (2)$$

де

$$P(H_i) = \frac{m_i}{m}, P(B/H_i) = \frac{k_i}{m_i}, \quad (3)$$

де m_i – кількість публікацій, зроблених i -тим експертом, k_i – кількість публікацій, зроблених i -тим експертом, які пов'язані з обраною криптовалютою, m – загальна кількість публікацій за період $[0; T]$, $P(H_i)$ – ймовірність того, що пост був опублікований i -тим експертом, $P(B/H_i)$ – ймовірність того, що в довільний момент

часу t був написаний пост, пов'язаний з обраною криптовалютою, за умови, що дану публікацію написав i -тий експерт.

Тоді за формулою Байєса: для кожного експерта обчислюємо ймовірності, що саме даний експерт створив допис, пов'язаний з обраною криптовалютою, якщо відомо, що в розглянутий період часу його точно було написано одним з експертів.

$$P(H_i/B) = \frac{P(H_i) \cdot P(A/H_i)}{P(A)}, \quad (4)$$

де $P(H_i)$ – ймовірність того, що пост був опублікований i -тим експертом, $P(B/H_i)$ – ймовірність того, що в довільний момент часу t з інтервалу $[0; T]$ був написаний пост, пов'язаний з обраною криптовалютою, за умови, що публікацію написав i -тий експерт, $P(H_i/B)$ – ймовірність того, що пост написав i -тий експерт, за умови, коли нам відомо, що в довільний момент часу t було створено допис, пов'язаний з обраною криптовалютою.

Отримані апостеріорні ймовірності розміщуємо у порядку спадання. Це означає, що той експерт, якому відповідає найбільша ймовірність $P(H_i/B)$ матиме найбільший вплив на курс обраної криптовалюти на проміжку $[0; T]$. Далі обирається експерт, який матиме наступну за значенням ймовірність $P(H_i/B)$ і т.д. В результаті, отримуємо ранжований перелік експертів згідно впливу їх дописів в соціальних мережах на курс криптовалюти $A^* = (a_1^*, a_2^*, \dots, a_n^*)$.

Слід зазначити, що подібне ранжування також дозволяє оцінити вплив групи експертів на курс певної криптовалюти на прогнозованому проміжку $[T; T + \Delta t]$, оскільки вплив публікацій зроблених за період часу $[0; T]$ розповсюджується також на деякий період часу $[T; T + \Delta t]$.

Окрім вищенаведеного, можна також застосовувати наступні критерії ранжування [15]:

- кількість підписників прямопропорційна до значення впливу даного експерта на певну аудиторію;
- активність – частота та регулярність публікацій на тему криптовалют;
- історія прогнозів – відсоток вдалих прогнозів;
- соціальний капітал – співпраця з фінансовими організаціями;
- регіональний вплив – вплив в конкретному географічному регіоні;
- тематична соціалізація – фокус на конкретних криптовалютах чи блокчейн технологіях;
- часовий фактор – актуальність інформації та швидкість її поширення.

Слід зазначити, що використання кожного з даних альтернативних критеріїв ранжування експертів потребує використання унікального математичного апарату та планується для дослідження у подальшому.

Отримання прогнозу курсу криптовалюти з урахуванням дописів в соціальній мережі обраної групи експертів на основі проведеного ранжування експертів

Постановка задачі:

Нехай є деяка навчальна вибірка $X = (x_1, x_2, \dots, x_s)$, де s – об'єм вибірки, яка складається з справжнього курсу певної криптовалюти за період часу $[0; T]$ (вибірка будується на основі часового ряду з таблиці 1). Необхідно отримати прогнози курсу обраної криптовалюти за наступний період часу $[T; T + \Delta t]$, враховуючи проведене у пункті 5.2 ранжування експертів.

Обґрунтування:

В основі запропонованого алгоритму лежить наступний принцип: кожен обраний експерт має певний вплив на курс криптовалюти. Тобто для того, щоб отримати прогноз із бажаною точністю, необхідно враховувати дописи не тільки «основного» (найбільш впливового) експерта [10], а й інших.

Таким чином, процес урахування дописів кожного експерта від найбільш впливового до найменш впливового складається з наступних кроків:

Крок 1: Встановити поріг для значення середньої відносної похибки прогнозування, а також значення T та Δt .

Крок 2: За допомогою АУДСМ [9] по навчальній вибірці X отримати прогнози курсів криптовалюти за період часу $[T; T + \Delta t]$ з урахуванням дописів кожного з експертів з ранжованого списку $A^* = (a_1^*, a_2^*, \dots, a_n^*)$ (див. пункт 5.2).

Крок 3: На основі отриманих на кроці 2 прогнозів сформувати вибірки $Y_i = (y_{i1}, y_{i2}, \dots, y_{is})$, $i = \overline{1; n}$, де s – об'єм вибірок.

Крок 4: Для перевірки точності прогнозів, обчислити середню відносну похибку прогнозування:

$$MAPE_i = \frac{1}{s} \cdot \sum_{l=1}^s \left(\frac{|x_l - y_{il}|}{x_l} \right) \cdot 100\%, \quad (5)$$

де x_l – елементи вибірки X , y_{il} – елементи вибірки Y_i , $i = \overline{1; n}$, $l = \overline{1; s}$, n – кількість експертів, s – об'єм вибірок X та Y [16].

Крок 5: Якщо значення $MAPE_i$ не перевищує заданого порогу, то перейти до наступного етапу. Якщо значення $MAPE_i$ перевищує заданий поріг, то повернутися до кроку 2 з метою отримання більш якісного прогнозу. Підвищення якості прогнозу може бути здійснено за рахунок збільшення тренувального датасету або зміни списку експертів.

Крок 6: Обчислити прогноз y_l курсів криптовалюти з урахуванням дописів усіх обраних експертів за формулою:

$$y_l = \sum_{i=1}^n (y_{il} \cdot \omega_i), \quad (6)$$

де y_{il} – прогноз курсу обраної криптовалюти в довільний момент часу t_l з інтервалу $[T; T + \Delta t]$, $l = \overline{1; s}$, отриманий на кроці 2, ω_i – рівень впливовості дописів в соціальній мережі експерта i , $i = \overline{1; n}$.

$$\omega_i = P(H_i/B), \quad (7)$$

$$\sum_{i=1}^n \omega_i = 1, \quad (8)$$

де ймовірність $P(H_i/B)$ обчислюється за формулою (4) (див. пункт 5.2).

Крок 7: На основі отриманих на кроці 2 прогнозів сформувані вибірку $Y = (y_1, y_2, \dots, y_s)$, де s – об'єм вибірок.

Крок 8: Для перевірки точності прогнозу, обчислити середню відносну похибку прогнозування $MAPE$.

$$MAPE = \frac{1}{s} \cdot \sum_{l=1}^s \left(\frac{|x_l - y_l|}{x_l} \right) \cdot 100\%, \quad (9)$$

де x_l – елементи вибірки X , y_l – елементи вибірки Y , $l = \overline{1; s}$, n – кількість експертів, s – об'єм вибірок X та Y [16].

Крок 9: Якщо значення $MAPE$ не перевищує заданого порогу, то роботу алгоритму завершено. Якщо значення $MAPE$ перевищує заданий поріг, то повернутися до кроку 6 з метою отримання більш якісного прогнозу. Підвищення якості прогнозу може бути здійснено за рахунок збільшення тренувального датасету або зміни списку експертів.

Запропонований алгоритм аналогічний методу лінійної згортки [17, 18].

Варто підкреслити, що вибірка X формується на відповідному часовому інтервалі на основі часового ряду з таблиці 1, а вибірки Y_i та Y будується з прогнозів курсу криптовалюти на цьому ж часовому інтервалі у ті самі моменти часу, що і відповідна вибірка X . Окрім того, всі вибірки є неранжованими, що обумовлюється особливістю поставленої задачі. Це дає змогу обґрунтувати правомірність переходу від часових рядів до статистичних вибірок і таким чином коректно обчислювати всі $MAPE$.

Слід також зазначити, що даний алгоритм дозволяє врахувати вплив абсолютно всіх обраних експертів, що наближає його до реальної картини процесу, порівняно з алгоритмом АУДСМ [9] та алгоритму [10], які враховують дописи лише одного експерта та є більш ідеалістичними. Постійний контроль точності отриманих результатів дозволяє гарантувати високу точність отриманих прогнозів. Порівняно з ітераційним алгоритмом АПУДГЕСМ, який представлено в роботі [11], наведений алгоритм дозволяє одночасно враховувати вплив дописів кожного з експертів на курс криптовалюти. Це унеможливило накопичення похибки обчислень на кожному ітераційному кроці.

Формування рекомендацій щодо фінансових операцій відносно криптовалюти на основі отриманих прогнозів

Постановка задачі:

На основі прогнозу курсів обраної криптовалюти, отриманого в пункті 5.3, сформувані рекомендації щодо можливих подальших фінансових операцій відносно цієї криптовалюти.

Обґрунтування:

Отримання достатньо точних прогнозів курсів обраної криптовалюти (див. пункт 5.3) дає змогу відстежити тенденції їх зміни на проміжку часу $[T + (n - 1) \cdot \Delta t; T + n \cdot \Delta t]$. Базуючись на цьому, можна зробити рекомендації щодо доцільності купівлі або продажу криптовалюти, з метою отримання прибутку.

Слід зазначити, що при формуванні рекомендацій необхідно враховувати потреби користувачів. Для цього необхідно встановити таке критичне значення, що: якщо $y_{cp} \geq y_{кр}$, то криптовалюту слід продавати; $y_{cp} < y_{кр}$, то криптовалюту слід купляти. В даних нерівностях y_{cp} – середнє арифметичне вибірки Y , що обчислюється за формулою:

$$y_{cp} = \frac{\sum_{l=1}^s y_l}{s}, \quad (10)$$

y_l – елементи вибірки, $l = \overline{1; s}$, s – об'єм вибірки.

Значення $y_{кр}$ може бути встановлено на розсуд користувача або, керуючись думкою експерта, згідно ситуації на ринку криптовалют.

Слід також зазначити, що $y_{кр}$ також можна отримати зі статистичної таблиці значень розподілу χ^2 за рівнем значущості α та степенем свободи $r = s - 3$: $y_{кр} = \chi_{кр}^2(\alpha; r)$. Це зумовлено тим, що згідно Центральної Граничної Теорема вибірка Y повинна наближатися до нормального розподілу, оскільки на формування курсу криптовалюти, взагалі кажучи, може впливати багато різних факторів. Дану гіпотезу можна перевірити за допомогою чисельних статистичних критеріїв, зокрема за допомогою критерію χ^2 -Пірсона.

Висновки

В даній роботі наведено модифікацію алгоритму АУДСМ [9], та розширено перелік вимог щодо вибору експертів [15].

Запропонований алгоритм дає змогу врахувати рівень впливу кожного експерта з відібраної групи. При цьому, використання АУДСМ [9] на кожному етапі алгоритму дозволяє значно спростити процес отримання прогнозів.

Даний підхід дозволяє підвищити точність прогнозів курсів обраної криптовалюти за рахунок врахування дописів кожного з відібраних експертів, на відміну від алгоритмів, представлених в роботах [9] та [10], в котрих, або обирався єдиний експерт, або з групи експертів визначався «основний». Тобто в обох випадках потенційним впливом інших експертів нехтували.

Запропонований алгоритм є більш наближеним до реальної картини процесу, порівняно з алгоритмом АУДСМ [9] та алгоритму з визначенням «основного» експерта (АУДОЕСМ) [10], які враховують дописи лише одного експерта та є більш

ідеалістичними. Постійний контроль точності отриманих результатів дозволяє гарантувати високу точність отриманих прогнозів.

На відміну від алгоритму АПУДГЕСМ, представленого в роботі [11], запропонований алгоритм дає можливість одночасно враховувати дописи групи експертів, що підвищить точність прогнозів шляхом унеможливлення накопичування похибки на кожному ітераційному кроці.

В рамках даного підходу обчислюються апостеріорні ймовірності того, що пост, пов'язаний з обраною криптовалютою і був написаний конкретним експертом на інтервалі прогнозування. На їх основі проводилося ранжування експертів.

Слід зазначити, що для різних проміжків часу можуть бути отримані різні результати ранжування експертів.

Для використання даного підходу рекомендується розглядати невеликі часові інтервали (до тижня), на кожному з яких більш точно можна проводити ранжування. Це потенційно має підвищувати точність прогнозів курсів обраної криптовалюти на вказаному часовому інтервалі.

Використання частот дописів в соціальних мережах, в якості параметру для визначення впливу експертів, дає змогу застосувати класичний апарат теорії ймовірностей, що гарантує коректність отриманих результатів.

Отримані прогнози використовуються для формування рекомендацій щодо купівлі або продажу обраної криптовалюти в залежності від потреб користувача та ситуації на ринку криптовалют.

До недоліків слід віднести, що на точність прогнозу може негативно вплинути невдало підібраний інтервал часу, на якому робиться прогноз, оскільки визначення тривалості впливу прогнозу певного експерту виходить за рамки розв'язуваної задачі.

Також якість прогнозу залежить від обраної групи експертів, адже невдало підібрана експертна група може негативно вплинути на роботу алгоритму.

Дані особливості свідчать про необхідність постійного моніторингу як курсів криптовалюти, так і дописів експертів в соціальних мережах.

З метою ще більшого підвищення точності прогнозів курсів криптовалюти та релевантності рекомендацій щодо її купівлі або продажу, в подальшому планується вдосконалення отриманого алгоритму шляхом використання альтернативних критеріїв ранжування експертів [15], розроблення альтернативних алгоритмів та їх включення в загальну інформаційну технологію визначення впливу дописів в соціальних мережах на курси криптовалют.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. P. Preethi, V. Uma, A. Kumar. (2015). «Temporal Sentiment Analysis and Causal Rules Extraction from Tweets for Event Prediction». *Procedia Computer Science*. №48. pp. 84–89. DOI: <https://doi.org/10.1016/j.procs.2015.04.154>.
2. Matthew S. Gerber. (2014). «Predicting crime using Twitter and kernel density estimation». *Decision Support Systems*. №61. pp. 115–125. DOI: <https://doi.org/10.1016/j.dss.2014.02.003>.
3. R. Alhajj, J. Rokne, (eds). (2018). «Encyclopedia of Social Network Analysis and Mining». *Analysis and Mining*. Springer, New York, NY. pp. 2699. DOI: <https://doi.org/10.1007/978-1-4614-7163-9>.
4. K. Byrd, A. Mansurov, O. Baysal. (2016). «Mining Twitter data for influenza detection and surveillance». *SEHS '16: Proceedings of the International Workshop on Software Engineering in Healthcare Systems*. pp. 43–49. <https://doi.org/10.1145/2897683.2897693>.
5. Ramona-Diana Leon, Raúl Rodríguez-Rodríguez, Pedro Gómez-Gasquet, Josefa Mula. (2017). «Social network analysis: A tool for evaluating and predicting future knowledge flows from an insurance organization». *Technological Forecasting and Social Change*. Vol. 114. pp. 103–118. <https://doi.org/10.1016/j.techfore.2016.07.032>.
6. D. Karahoca, A. Karahoca, Ö. Yavuz. (2013). «An early warning system approach for the identification of currency crises with data mining techniques». *Neural Comput & Applic*. Volume 23. pp. 2471–2479. DOI: <https://doi.org/10.1007/s00521-012-1206-9>.
7. M. Žukovič. (2012). «Dynamics of episodic transient correlations in currency exchange rate returns and their predictability». *centr.eur.j.phys*. Volume 10. pp. 615–624. DOI: <https://doi.org/10.2478/s11534-011-0120-6>.
8. Gavrilenko, O., Novakivska, K., Shumeiko, O. (2022). «Select the most influx economic factors for forecasting the US dollar exchange rate». *Bulletin of NTU*. No. 54. pp. 26–35. DOI: <https://doi.org/10.33744/2308-6645-2022-4-54-026-035>.
9. Gavrylenko, O., Miahkyi, M., Zhurakovskiy, Y. (2022). «The task of analyzing publications to build a forecast for changes in cryptocurrency rates». *Adaptive automatic control systems*. Volume 2. No. 41. pp. 90–99. DOI: <https://doi.org/10.20535/1560-8956.41.2022.271349>.
10. Bidyuk, P., Gavrilenko, O., Myagkyi, M. (2023). «The algorithm for predicting the cryptocurrency rate taking into account the influence of posts of a group of famous people in social networks», *System research and information technologies: an international scientific and technical journal*. № 2. pp. 22–34. DOI: <https://doi.org/10.20535/SRIT.2308-8893.2023.2.02>.

11. *O. Gavrilenko, M. Myagkyi.* (2023). «Forecasting the cryptocurrency exchange rate based on the ranking of expert opinions». Innovative technologies and scientific solutions for industries. No. 4 (26). pp. 24-32. <https://doi.org/10.30837/ITSSI.2023.26.024>.
12. Binance cryptocurrency exchange. URL: <https://www.binance.com/en/strategy/spot/grid/DOGEUSDT>. (accessed: 13.10.2023).
13. *Kartashov, M.* (2007). «Internationality, processes, statistics». Publishing and printing center Kiev University, Kiev, 504 p.
14. *Denisyuk, V., Bobkov, V., Pogrebetska, T., Repeta, V.* (2009). «Higher mathematics. Part 4. The theory of probabilities and mathematical statistics». Edition «NAU-druk», Kiev, 256 p.
15. *Gavrilenko O., Myagkyi M.* (2023). «Study of the influence of the expert group's posts on the course of cryptocurrencies». XIX International Scientific Conference named after academician Mykhailo Kravchuk. Abstracts of reports. P. 162–164.
16. Time series. URL: <https://kstat.pnu.edu.ua/wp-content/uploads/sites/63/2018/04/%D0%A7%D0%B0%D1%81%D0%BE%D0%B2%D1%96-%D1%80%D1%8F%D0%B4%D0%B8.pdf>. (accessed: 14.09.2024).
17. *Michkivsky S.M., Prigunov O.V., Rymar P.V.* (2019). «Decision-making systems and methods: methodological guidelines». Vinnytsia, Vasyl' Stus DonNU. P. 76. URL: <https://r.donnu.edu.ua/bitstream/123456789/> (accessed: 14.09.2024).
18. *Bidyuk P.I., Tymoshchuk O.L., Kovalenko A.E., Korshevnyuk L.O.* (2022). «Systems and methods of decision support. Textbook». Igor Sikorsky Kyiv Polytechnic Institute. P. 610. URL: <https://ela.kpi.ua/server/api/core/bitstreams/6958f683-fbac-4506-9c85-5115c8f8b4c6/content>. (accessed: 14.09.2024).