

В. Б. Мокін¹
С. О. Жуков¹
Л. М. Куперштейн¹
О. В. Слободянюк¹

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПРОГНОЗУВАННЯ КУРСУ КРИПТОВАЛЮТ НА ОСНОВІ КОМПЛЕКСНОЇ ІНЖЕНЕРІЇ ОЗНАК

¹Вінницький національний технічний університет

Розроблено інформаційну технологію прогнозування курсу криптовалют на основі комплексної інженерії ознак. Особливість цієї технології полягає в системному підході до відбору ознак. Проведено аналіз груп зовнішніх та внутрішніх факторів потенційного впливу на криптовалютний ринок. Аналіз ознак, які характеризують зміни курсу криптовалют, показав, що окрім базових первинних ознак, які доступні на багатьох криптобіржах, важливішими для подальшого прогнозування курсу криптовалют є вторинні ознаки, які отримують з базових первинних шляхом застосування до них різних математичних операцій та/або алгоритмічних трансформацій. Аналіз великої кількості джерел показав, що криптовалюти мають низку характерних особливостей, які й зумовили їх велику популярність і які теж варто враховувати під час формування та вибору ознак. В роботі проведена систематизація ключових таких особливостей, а також запропоновано яким чином їх формалізувати у вигляді ознак. Формалізувати ознаки запропоновано за системним підходом, згідно з постулатами технічної кібернетики, де говориться, що будь-який об'єкт дослідження можна представити у вигляді чорного ящика (ЧЯ), який контактує з зовнішнім середовищем у п'яти точках, що можуть у багатовимірному випадку бути множинами ознак чи змінних. Запропоновано загальну математичну модель формування цих факторів, яке полягає у генеруванні великої кількості вторинних ознак на основі простих математичних, алгоритмічних та статистичних трансформацій з подальшим відбором найрелевантніших з них. Технологія передбачає синтез нових вторинних ознак на основі інших вторинних ознак, за окремими винятками, які формалізовано у вигляді системи правил. Це дозволить зменшити перенавантаження моделі та покращити її узагальнюючу здатність.

Для доведення працездатності розробленої технології розглянуто приклад її застосування на базі криптовалюти біткоїн за щодобовими даними 2020—2021 років. Проведені дослідження та комп'ютерні експерименти показали ефективність та працездатність запропонованої технології.

Ключові слова: криптовалюта, інженерія ознак, прогнозування, інформаційна технологія, біткоїн.

Вступ

Останніми роками суттєво зростає капіталізація світових криптовалют. Якщо на початок 2017 року вона складала \$18 млрд, то станом на 7 квітня 2022 р. вона складає вже більше \$2 трлн. За різними реєстрами існує більше 10 тисяч криптовалют і ця кількість постійно зростає. Все більше як громадян, так і урядів країн, розглядають крипторинок, як можливість вкладення коштів та отримання прибутку. Тому, все актуальнішою стає проблема планування такої стратегії. А у цій проблемі головною задачею є прогнозування курсу криптовалюти у майбутньому — від декількох секунд до декількох місяців.

Задача прогнозування курсу криптовалют суттєво ускладнюється впливом багатьох факторів, а тому рішення на основі часових рядів, які аналізують і враховують тільки попередні значення ряду, не забезпечують високої точності. Для усунення цієї проблеми використовуються методи машинного навчання «з вчителем». Класичний алгоритм розв'язання задачі машинного навчання «з вчителем», тобто на основі тренувального датасету, де є відомою цільова ознака, містить такі етапи: передоброблення:

– *етап 1*. Розвідувальний аналіз даних (англ. — Exploratory Data Analysis (EDA), на якому виділяються ключові закономірності, здійснюється очищення даних, аналіз аномалій, висувуються гіпотези, які перевірятимуться на наступних етапах;

– *етап 2*. Інженерія ознак (англ. — Feature Engineering (FE)) — формування, аналіз цінності та вибір ознак;

– *етап 3*. Побудова моделі за методами машинного навчання (англ. Model Tuning (MT)) — зазвичай використовуються ансамблі моделей чи моделі складнішої архітектури;

– *етап 4*. Аналіз результатів, зокрема аналіз викидів (значень, на яких похибка є найбільшою).

Як правило, цей процес є ітераційним: етапи 1 і 2 доводиться проходити декілька разів (змінити ознаки — проаналізували результат, знов змінили...). Після етапу 4, зазвичай, виконується перехід на етапи 1 чи 2 до тих пір, поки не буде досягнута прийнятна точність та адекватність моделі.

Важливим є те, що етапи 1, 3 і 4 для задачі прогнозування курсу криптовалют вже добре вивчені. Щодо етапу 2 вже опубліковано десятки тисяч статей і ще більше — публічних програмно-утиліт з прикладами застосування різних технологій машинного навчання. Однак, якщо недостатньо системно та якісно виконати етап 2, тоді і на етапі 4 не вдасться досягнути гарних результатів. Отже, найважливішим є удосконалення підходів до інженерії ознак криптовалют, а тоді задача зводиться до варіантів на етапі 3, які вже добре вивчені і мають багато ефективних розв'язків.

Задачею інженерії ознак як етапу інформаційної технології прогнозування курсу криптовалют займається велика кількість науковців провідних країн світу [1]—[4]. Але у їхній переважній більшості властива відсутність системного підходу до відбору ознак. У найкращому випадку, беруть багато технічних показників, виконують різні зсуви в часі та низку комбінованих ознак на їхній основі і далі основну увагу приділяють тільки етапу 3.

Метою дослідження є підвищення точності прогнозування курсу криптовалют за рахунок оптимізації процесу синтезу та відбору їх ознак з урахуванням системного підходу у відповідній інформаційній технології.

Вихідні умови та постановка задачі

Задача прогнозування курсу криптовалют полягає у прогнозуванні цільової ознаки Y_t в часі t на N кроків методами машинного навчання відносно інших ознак. Завичай, вихідною ознакою (поточний курс криптовалюти) є показник «Close» у момент t , обчислений в кінці часового інтервалу, за який беруться значення (за хвилину, годину, добу тощо), позначимо його C_t . Аналіз джерел показав [1], [5]—[9], що в цій задачі найпопулярнішими є такі цільові ознаки Y_t , обчислені за значеннями C_t , зсунутими на N кроків відносно інших ознак: реальні дробові значення показника «Close» $Y_t = C_{t+N}$ заданої криптовалюти (регресійна задача); бінарна ознака «Чи є приріст значень «Close» на 1 крок вперед?» $Y_t = \text{sign}(C_{t+1} - C_t)$ — «Так» або «Ні» (класифікаційна задача); значення дохідності «Return» $Y_t = C_{t+N}/C_t$ (регресійна задача); логарифм дохідності «Log-Return» $Y_t = \ln(C_{t+N}/C_t)$, який, на відміну від дохідності, дозволяє зіставляти різні валюти та перейти від відсотків до абсолютних одиниць, що зручно для аналізу (регресійна задача).

У конкурсі «G-Research Crypto Forecasting» [10] з прогнозування курсу 14 криптовалют на базі платформи Kaggle, який триває з листопада 2021 року по травень 2022 року, 3 з чотирьох авторів цієї статті взяли активну участь серед майже 2000 команд з усього світу. Набутий досвід використано під час роботи над цим дослідженням.

Іноді на етапі розвідувального аналізу даних цільовою ознакою є не останні дні, а випадкові 20...25 % значень з оригінального датасету, але які не використовуються під час тренування моделі — задача передбачення. Однак, в самій задачі прогнозування курсу валют, де потрібно знайти що буде далі, валідаційними та тестовими даними є саме останні дні оригінального датасету.

Аналіз існуючих ознак, які характеризують зміни курсу криптовалют

Практично усі криптобіржи та інші сторонні онлайн сервіси дозволяють отримувати через API первинні дані ринку обраної криптовалюти, вказуючи початок та кінець часового проміжку, а також тривалість торгового періоду. Зазвичай надаються такі дані: «timestamp/date» — мітка часу або дата та час торгового періоду, «Open» — ціна на початок торгового періоду, «High» — максимальна ціна за торговий період, «Low» — мінімальна ціна за торговий період, «Close» — ціна на кінець торгового періоду, «Count» — кількість торгових транзакцій за торговий період, «Volume» — обсяги торгів (купівлі/продажу) за торговий період, «AdjClose» — скорегована ціна на кінець торгового періоду.

Зазвичай показники «Open», «High», «Low», «Close» називають «OHLC»-показники.

Для оцінки ринку криптовалют методами машинного навчання окрім первинних показників ринку також можна використати низку ознак з фундаментального аналізу, який використовується інвесторами для визначення «внутрішньої вартості» криптовалюти. Головна мета цього аналізу — шляхом розгляду великої кількості внутрішніх та зовнішніх факторів визначити, переоцінений чи недооцінений криптоактив на цей момент. Тут можна виділити такі ознаки [11]: кількість активних блокчейн адрес, обсяг винагороди за підтвердження блоку, розмір хешрейту, кількість монет в стейкінгу, обсяг ринкової капіталізації, кількість заявок на купівлю/продаж (ліквідність криптовалюти), обсяг оборотної пропозиції (кількість монет, які є загальнодоступними та циркулюють на ринку), розмір інфляції криптовалюти, індикатор NVT (Network Value to Transaction) — відношення цінності мережі до обсягу транзакцій, індикатор MVRV (Market Value by Realised Value) — відношення ринкової вартості до реалізованої, індикатор S2F (Stock-to-Flow) дефіцитності криптовалюти з обмеженою пропозицією.

Серед аналітиків фондових ринків активно використовуються інструменти технічного аналізу (ТА), який орієнтований на передбачення майбутнього руху цін на підставі інформації про минулу прибутковість активів [12], [13]. Низку індикаторів технічного аналізу, які використовуються для характеристики та прогнозування фінансових інструментів, можна використати і для моделювання динаміки криптовалют. При цьому ТА-індикатори є вторинними (похідними) від первинних «OHLC». Можна виділити такі основні ТА-індикатори: індикатор RSI (Relative Strength Index) — індекс відносної сили, який вказує на перекупленість або перепроданість активу. Також на його основі використовують індикатор Stochastic RSI як осцилятор, що відстежує рух ціни, індикатор MA (Moving Average) — ковзна середня (КС), яка згладжує коливання ціни. Ефективнішими є низка модифікацій цього індикатора, а саме: ЕМА (експоненційна КС), WMA та WEMA (зважені КС та ЕКС), DMA та DEMA (подвійна КС та ЕКС відповідно), ТМС та ТЕМА (потрійна КС та ЕКС відповідно), індикатор MACD (Moving Average Convergence Divergence) призначений для визначення майбутнього цінового руху активу за допомогою взаємозв'язку двох ковзних середніх, індикатор Momentum — показник, що вимірює величину зміни ціни фінансового інструменту за певний період, індикатор Williams %R — показник, що вимірює співвідношення ціни закриття з діапазоном цін (макс...мін) протягом певної кількості періодів, індикатор ATR (Average True Range) — показник волатильності ринку, індикатор CCI (Commodity Channel Index) — показник, що вимірює відхилення ціни фінансового інструменту від його середньостатистичної ціни, лінії BB (Bollinger Band) — криві, що вимірюють волатильність ринку, а також визначають рівень перекупленості та перепроданості активу.

До іншої категорії ознак, що можуть впливати на ціну криптовалюти можна віднести топові фондові індекси: S&P500, Dow Jones, Nasdaq, Russel 2000, VIX та ін. [14]. Як ознаки впливу на крипторинки варто врахувати ціни на комодіті, так звані біржові товари [15]. В першу чергу це енергоресурси та цінні метали: нафта, газ, електроенергія, золото, срібло.

В окрему категорію зовнішніх факторів впливу, які напряму не пов'язані з фондовим ринком, але можуть суттєво вплинути на його рух, можна віднести такі: новини, публікації відомих людей у соціальних мережах, рівень захворюваності (наприклад, COVID-19), публікації економічного календаря, військові події, святкові дні, кліматичні та природні фактори тощо.

Крім того, на основі вищеписаних ознак можна отримати низку похідних вторинних шляхом застосування таких операцій:

- обчислення *статистичних* показників (середнє, медіана, дисперсія, середньоквадратичне відхилення, різні квантили тощо) за певні інтервали часу (1, 2, 3, 4, 5, 10, 20, ... кроків) як одного значення на інтервал (проста кластеризація), так і як ковзне значення (згладжування даних) — є спеціальні бібліотеки для автоматизації цього процесу, наприклад Python-бібліотеки TSFRESH [16], TSFEL [17] та інші [18], які генерують сотні таких ознак, а потім відбирають з них найважливіші для заданої цільової ознаки (як правило, тільки для класифікаційної задачі), а просте взяття ковзного середнього із заданим вікном зручно реалізує базова Python-бібліотека Pandas;

- *зсув* у часі на певну кількість кроків назад та/або обчислення *приросту* в абсолютному чи відносному варіанті поточного значення на таке зсунуте чи — навпаки;

- врахування *просторової* прив'язки, якщо вона є, шляхом зміни точки відліку часових інтервалів не від початку доби, а від часу відкриття головної криптобіржи;

- «*one-hot*»-кодування (з англ.: швидке кодування, оскільки це прискорює оброблення, хоча вимагає більше пам'яті, оскільки датасет стає значно більшим), коли замість однієї змінної з *B*

варіантами значень, формується B бінарних ознак (є подія у певний інтервал часу чи немає: 1 або 0) — популярно для частих аномалій (свят, пікових періодів хвиль коронавірусу тощо), але може застосовуватись і до будь-яких змінних, зокрема дробових, якщо перед тим виконати операцію категоризації або групування і трансформувати ці значення у порівняно невелику кількість класів;

– виконання *математичних операцій та/або алгоритмічних трансформацій з одною* заданою ознакою (піднесення у степінь, логарифмування, тригонометричні функції, зворотне значення та ін.) — для автоматизації можна використати бібліотеки типу AutoFeat [19], Featuretools [20] та ін.; також до цієї операції можна віднести різні трансформації часових рядів, які автоматизує Python-бібліотека TSAUG [21], як правило, для збільшення розміру датасетів на основі часових рядів;

– виконання *математичних операцій та/або алгоритмічних трансформацій з декількома* ознаками — цей підхід є досить популярним в задачі прогнозування курсу криптовалют. Для автоматизації таких операцій можна також використовувати різні Python-бібліотеки, наприклад, бібліотека GPLEarn [22] дозволяє з використанням методів генетичного програмування (Genetic Programming Learning) підібрати аналітичний вираз із заданих ознак, який буде найточніше відповідати заданій цільовій ознаці. А інструмент Featuretools [23] дозволяє генерувати нові ознаки, здійснюючи прості математичні операції між певними базовими.

– *кластеризація*, коли декілька (чи усі) ознак кластеризують і додають ознаку у вигляді номера кластера.

Класифікуємо та формалізуємо цю інформацію на основі системного підходу.

Системний підхід до інженерії ознак криптовалют

Проведений аналіз великої кількості джерел показав, що криптовалюти мають низку характерних особливостей, які й обумовили їх велику популярність та які варто враховувати під час формування та вибору ознак, зокрема шляхом порівняння з підходами для традиційних валют та активів [24]—[31]:

– доступ до операцій у будь-який час, тобто не треба чекати відкриття біржі зранку — можна здійснювати операції будь-коли, до речі, зазвичай час у датасетах відраховується від універсального UTC, тобто за Гринвічем — GMT+0;

– деякі криптовалюти орієнтовані більше на ринок США чи Китаю, а відповідно, може бути підвищена активність операцій тоді, коли у відповідній країні відкривається і закривається біржа, тому просторова прив'язка теж може мати значення, хоча й не таке велике, як для традиційних валют та активів;

– вхідні дані, зазвичай доступні через API з криптобирж (і можуть мати відмінності) з усередненням за різний час: для короткострокового прогнозування беруть інтервал 1...15 хвилин, для довгострокового — 1 доба;

– більшість факторів залежить від часу, тобто важливо враховувати їх агреговані показники (середні чи максимальні значення) за різні інтервали часу (15, 30, 60 хвилин, 1, 2, 3, 5, 6, 12, 24 годин або 1, 2, 3, 7, 14, 30 діб та ін., залежно від дискретності вхідних даних);

– операції можна здійснювати з будь-якою частиною монети, оскільки немає необхідності округляти до її сотих частин — можна купити чи продати 0,000241... кількість монет, головне, щоб знайшовся продавець чи покупець, відповідно;

– більшість криптовалют мають обмеження за кількістю монет, наприклад, Біткоїн має 21 млн монет (вже згенеровано більше 19 млн монет), але є й винятки, наприклад, усі монети криптовалюти для Інтернету речей ІОТА (більш як $2 \cdot 10^{15}$ шт.) були згенеровані одразу, для MAKER — межа в 1 млн монет вже досягнута, а щодо криптовалюти Dogecoin, то заявлено, що вона не буде обмежувати свою кількість, а це означає, що моделі для таких валют можуть мати відмінності;

– деякі криптовалюти легалізовані або у стані легалізації урядами країн, але це, зі свого боку, означає, що вони стають найзалежнішими від рішень цих урядів, особливо, якщо вони скасують таку легалізацію; і навпаки, низка криптовалют з дуже сильним рівнем захисту, до прикладу Монего, можуть підпадати під пряму заборону і переслідування як такі, що потенційно можуть використовуватись криміналом та терористами, що також впливає на ринок;

– криптовалютний ринок є сильно волатильним, а тому, зазвичай, страхові компанії не ризикують на ньому здійснювати страхування, а тому немає ніякої стабільності, гарантій того якщо учасник ринку втратить все, через якісь технічні проблеми, то йому це ніхто не компенсує — все на свій страх і ризик;

– операції з криптовалютами можуть проводити як організації, так і звичайні громадяни по всьому світу, які можуть утворювати консорціуми, коаліції, вступати в конфронтацію, можливі різні афери та застосовувати весь апарат теорії ігор, що означає, що значення не є природними, а формуються штучно і залежать не тільки від зовнішніх чи випадкових природних факторів, а й від спланованих чи спонтанних рішень учасників ринку; одним із наслідків є певна залежність таких рішень від аналізу постів у соцмережах (існують дослідження того, як змінювався курс декількох криптовалют після твітів Ілона Маска);

– криптовалюти використовуються часто, як резервний канал зберігання коштів і як альтернатива каналам, які регулюються державами, тому часто активізація має місце під час різних потрясінь, які можуть впливати на державний сектор: різке зростання кількості хворих чи померлих від коронавірусу (рис. 1), активізація бойових дій, масштабні теракти чи стихійні лиха тощо; одним із наслідків такої особливості є, наприклад, зворотна у короткостроковому періоді кореляція курсу біткоїна і курсу золота (рис. 2).

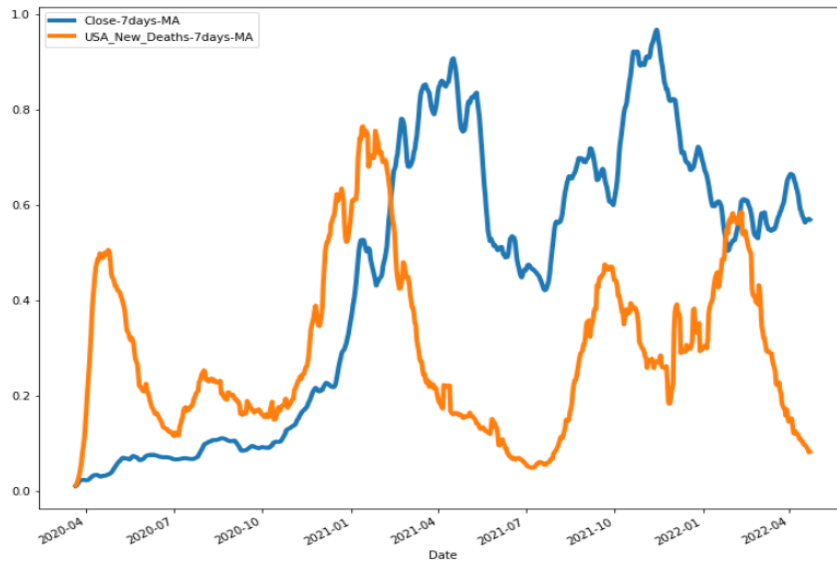


Рис. 1. Порівняння кількості померлих у США від коронавірусу та курсу біткоїну, усереднених за тиждень [18]



Рис. 2. Порівняння курсу біткоїну та курсу золота протягом 2017–2022 рр. [32]

Важливо враховувати в ознаках можливі стратегії дії учасників. Наприклад, популярне застосування стратегії типу «Райдуга», коли, в залежності від потрапляння курсу в одну з 7 зон (4-та зона відповідає середньому тренду), приймається те чи інше рішення (рис. 3).

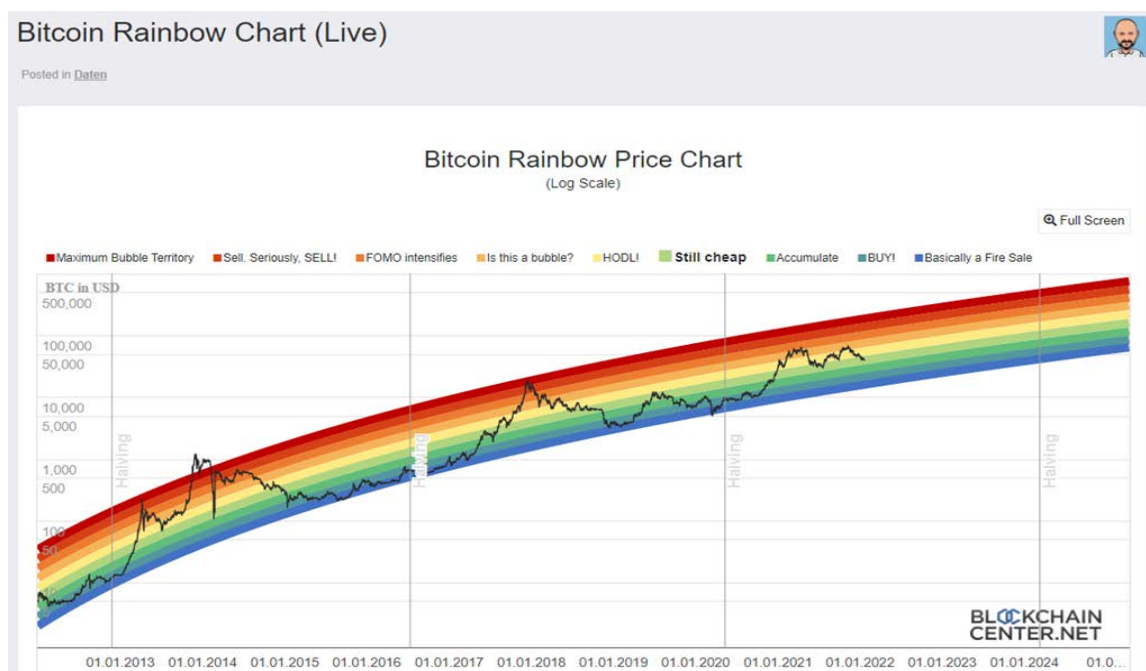


Рис. 3. Візуалізація стратегії типу «Райдуга» для прийняття рішень щодо продажу чи придбання біткоїнів [33]

У таблиці подана систематизація ключових із цих фактів та яким чином пропонується їх формалізувати у вигляді ознак.

Способи формалізації особливостей криптовалют у вигляді ознак

№ з/п	Особливість криптовалют	Формалізація для врахування у вигляді ознак
1	Деякі криптовалюти орієнтовані більше на певні біржі у США, Китаю чи ін., а тому можлива активізація операцій під час відкриття чи закриття цих бірж	Створення ознаки, яка містить 1 під час відкриття і закриття відповідної біржі, а решту значень — 0, або брати інтервал часу для формування агрегованих показників, наприклад, за кожні 12 годин не від початку доби за Гринвічем, а від моменту відкриття головної біржі у відповідній країні. Якщо модель будується для декількох валют одночасно, тоді враховувати й просторову прив'язку головної біржі у вигляді часового зсуву
2	Більшість факторів залежить від часу, тобто важливо враховувати їх агреговані показники	Слід передбачити операцію агрегування в часі з різними інтервалами, залежно від дискретності вхідних даних
3	Деякі криптовалюти мають відмінності, які є важливими, якщо модель враховує одночасно більше однієї валюти (обмежена чи необмежена кількість монет, різні програмно-технічні особливості, зокрема майнінгу, тощо)	Формувати ознаку по кожній особливості окремо, але є сенс це робити тільки для мультивалютних моделей або криптовалют, де такі ознаки зазнавали змін протягом періоду моделювання і потенційно могли впливати на курс валюти
4	Аномальні непередбачувані факти впливу — так звані «чорні лелеки»	Зробити ознаки з історією таких впливів, щоб модель розрізняла ситуації, коли такі факти мали місце і коли — ні
5	Учасники можуть приймати рішення за деякими стратегіями («Райдуга» та ін.).	Створити ознаку, яка буде містити номер рішення, яке впливає з відповідної стратегії і яке, ймовірно, багато хто з учасників ринку міг застосувати

Формалізація ознак у задачі прогнозування курсу криптовалют

Пропонуємо формалізувати ознаки за системним підходом. Згідно з постулатами технічної кібернетики, будь-який об'єкт дослідження можна представити у вигляді чорного ящика (ЧЯ), який контактує з зовнішнім середовищем у п'яти точках, що можуть у багатовимірному випадку бути множинами ознак чи змінних, а саме (рис. 4) [34, стор. 9—10]:

– у точці 1 живлення, в якій до об'єкта дослідження з зовнішнього середовища надходить енергія або маса — діяльність із забезпечення функціонування ринку криптовалют, зокрема споживання електроенергії, програмно-технічні показники, майнінг тощо — ці ознаки варто враховувати, якщо будується мультивалютна модель або аналізується тривалий період, коли, наприклад, винагорода за майнінг потенційно може впливати на курс валюти через те, що з часом менше стає

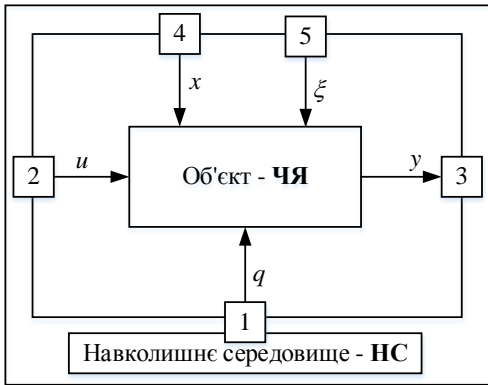


Рис. 4. Графічна інтерпретація виділення об'єкта дослідження з зовнішнього середовища [34, стор. 10]

кожного інтервалу Δt , за який обчислюється вихідна ознака «Close» — «High», «Low», «Volume» тощо;

– у точці 5 прикладення узагальненої завади ξ , в якій зовнішнє середовище намагається протягом цього ж часу додатково до основного, але неконтрольовано, збурити процес в об'єкті дослідження — рішення урядів, стихійні лиха, війни, показники захворюваності на коронавірус, погода, свята чи передсвяткові дні та інші ознаки.

Оскільки класична задача машинного навчання полягає у побудові моделі M , яка дозволяє за множиною вхідних ознак передбачати вихідну Y , то пропонуємо формалізувати її у такому вигляді:

$$Y = M(U, X, \Xi, Q); \quad (1)$$

$$U = \{u, \Omega_u D\}; X = \{x, \Omega_x D\}; \Xi = \{\xi, \Omega_\xi D\}; Q = \{q, \Omega_q D\}; D = \{u, x, \xi, q\}, \quad (2)$$

де U, X, Q, Ξ — множини первинних ознак u, x, ξ, q , відповідно, та вторинних для них, обчислених по усіх вхідних даних $D = \{u, x, \xi, q\}$ з використанням комплексного оператора Ω , а саме — функцій та функціоналів, які застосовуються до даних D у різних комбінаціях.

Відповідно, до вищенаведеної класифікації операцій над первинними ознаками, пропонуємо виділяти такі види операцій над значеннями:

$$\Omega D = \{\{Sd\}, \{\nabla^n d\}, \{R^m d\}, \{Bd\}, \{Fd\}, \Psi D, \Phi D\}, D = \{d\}, \quad (3)$$

де $\{Sd\}$ — множина ознак із статистичними показниками значень кожної ознаки d окремо за різні інтервали часу; $\{\nabla^n d\}$ — множина ознак зі зсувом кожної ознаки d назад у часі на n кроків (для різних ознак, n може набувати різних значень: 1, 2, 3, 5, 10, 20, ...); $\{R^m d\}$ — множина ознак з відношеннями кожної ознаки d до її значень, зсунутих назад у часі на m кроків (для різних ознак, m може набувати різних значень: 1, 2, 3, 5, 10, 20, ...; теоретично, може й набувати від'ємні значення $-1, -2, \dots$, що означатиме відношення минулих значень до поточного); $\{Bd\}$ — множина ознак із «one-hot»-кодуванням, отриманих з кожної ознаки d , яка попередньо була перетворена у скінченну категорійну множину варіантів значень; $\{Fd\}$ — множина ознак, отриманих шляхом виконання математичних операцій та/або алгоритмічних трансформацій зі значеннями кожної ознаки d окремо; ΨD — множина ознак, отриманих шляхом виконання математичних операцій та/або алгоритмічних трансформацій з багатьма чи усіма комбінаціями ознак із множини D ; ΦD — множина ознак, отриманих за допомогою кластеризації усіх ознак D чи їх різних комбінацій, зазвичай це номер кластера, якому відповідає кожна комбінація первинних ознак.

Однак, досвід показав, що ефективнішими є не вторинні показники, отримані напряду з первинних, а — вторинні, отримані з інших вторинних, тобто з використанням ще й різних комбінацій вторинних показників, що приводить до суттєвого збільшення кількості можливих ознак:

$$Y = M^*(U, X, \Xi, Q); \quad (4)$$

$$U = \{u, \Omega_u D\}; X = \{x, \Omega_x D\}; \Xi = \{\xi, \Omega_\xi D\}; Q = \{q, \Omega_q D\}; D = \{u, x, \xi, q\}; \quad (5)$$

$$\Omega^* D = \{\Omega d, \{S\nabla^n d\}, \{\nabla^n Sd\}, \{BSd\}, \{BS\nabla^n d\}, \{B\nabla^n Sd\}, \{FSd\}, \dots\}, D = \{d\}, \quad (6)$$

де модель M^* відрізняється від моделі M у виразі (1) тим, що її слід застосовувати мінімум у 2 етапи: спочатку запусити, проаналізувати цінність ознак (їх можуть бути десятки тисяч), відібрати найцінніші, а вже потім будувати модель для прогнозування; а комплекс операцій над вхідними даними Ω^* містить як операції Ω над первинними ознаками, так і операції над вторинними і теж застосовується в декілька ітерацій.

Теоретично, застосування різних перестановок 7 варіантів операцій (порядок має значення, оскільки зсув середньотижневих значень і середньотижневі зсунуті значення — це різні значення, враховують нелінійність і нестационарність процесу) може дати велику кількість комбінацій (це — факторіал від добутку кількості операцій кожного виду). Однак, не усі ці операції доцільно застосовувати у довільному порядку. Варто зазначити такі правила:

1. Не варто будувати ознаки із застосуванням будь-яких операцій над результатами «one-hot»-кодування (B), натомість, це кодування можна застосовувати до усіх інших результатів операцій, окрім результатів іншого (попереднього) кодування (BB).

2. Низьку доцільність мають статистичні показники результатів кластеризації ($S\Phi$) і кластеризація результатів кластеризації ($\Phi\Phi$).

3. Зсув в часі та взяття відношень з таким зсувом є сенс робити тільки щодо тих самих показників, статистичних показників ознак за різний період та результатів математичних операцій та/або алгоритмічних трансформацій з ознаками ($\nabla^n S, \nabla^n \Psi, \nabla^n \Phi, \nabla^n \nabla^n, R^n S, R^n \Psi, R^n \Phi, R^n R^n$).

4. Математичні операції (Ψ, F) доцільно виконувати над усіма результатами операцій, зокрема над результатами цих же операцій, окрім «one-hot»-кодування (B).

Математично це можна представити у такому вигляді:

$$\begin{aligned} & B(\forall(\dots)), \forall \neq B; S(\forall(\dots)), \forall \neq B, \Phi; \Phi(\forall(\dots)), \forall \neq B, \Phi; \\ & \nabla^n(\forall(\dots)), \forall = F, S, \nabla^n, \Psi; R^n(\forall(\dots)), \forall \neq F, R^n, S, \Psi; \\ & F(\forall(\dots)), \forall \neq B; \Psi(\forall(\dots)), \forall \neq B; \quad \forall \subset [B, F, S, R^n, \nabla^n, \Psi, \Phi], \end{aligned} \quad (7)$$

де $\forall(\dots)$ — довільний комплекс операцій із множини (6), який застосовується до значень в дужках.

Схема допустимих комбінацій системи правил (7) подана на рис. 5.

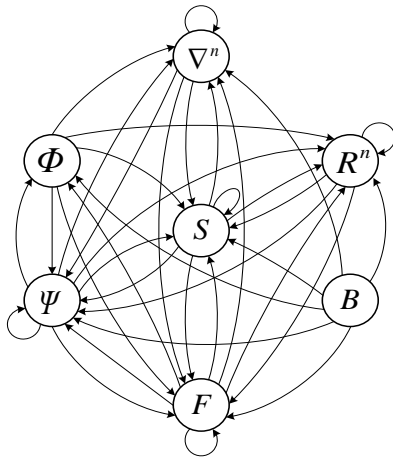


Рис. 5. Схема допустимих комбінацій видів операцій над результатами інших видів операцій (вторинні ознаки у множині (6)) за системою правил (7)

Наприклад, якщо ми вирішимо на етапі EDA (вивчення закономірностей у даних) здійснити прогноз середньодобового курсу біткоїна без зсуву на основі рис. 1 ($Y_t = C_t$ — регресійна задача), виходячи з гіпотези, висунутої на етапі EDA про те, що на цей курс впливає тільки добовий приріст кількості смертей у США (ξ), але у різні періоди часу він зазнає запізнення в часі (операція ∇^n , але вона застосовується до ковзного середнього від ξ — операція типу S), тоді модель (1)—(3) (у розширеній моделі (4)—(6) немає потреби) матиме вигляд:

$$Y = M(\Xi), \Xi = \{\xi, \Omega_\xi \xi\}; \quad \Omega_\xi \xi = \{\{\nabla^n S \xi\}\}, n = 0, \dots, T, \quad (8)$$

де T — максимальна кількість кроків, на яку здійснюється зсув даних по ковіду, для обчислення $T + 1$ нових ознак (значення можуть бути і не послідовними числами).

Побудована модель у режимі передбачення на етапі EDA, коли в тестовий датасет беруться тільки 20 % випадково вибраних даних оригінального датасету, які не використовуються на етапі ідентифікації моделі, щоб перевірити наскільки вдасться ідентифікувати модель, коли $T = 1, 4, 14, 16$ діб — результат вийшов на диво гарним: відносна похибка для тестових даних склала 4,72 %, коефіцієнт детермінації $r2_score = 0,98$ (програма опублікована у платформі Kaggle — [18]). Згенеровані за моделлю (8) ознаки та цільова функція наведені на рис. 6а, а результат ідентифікації моделями «Decision Tree Regressor» і «XGBoost Regressor» — на рис. 6б.

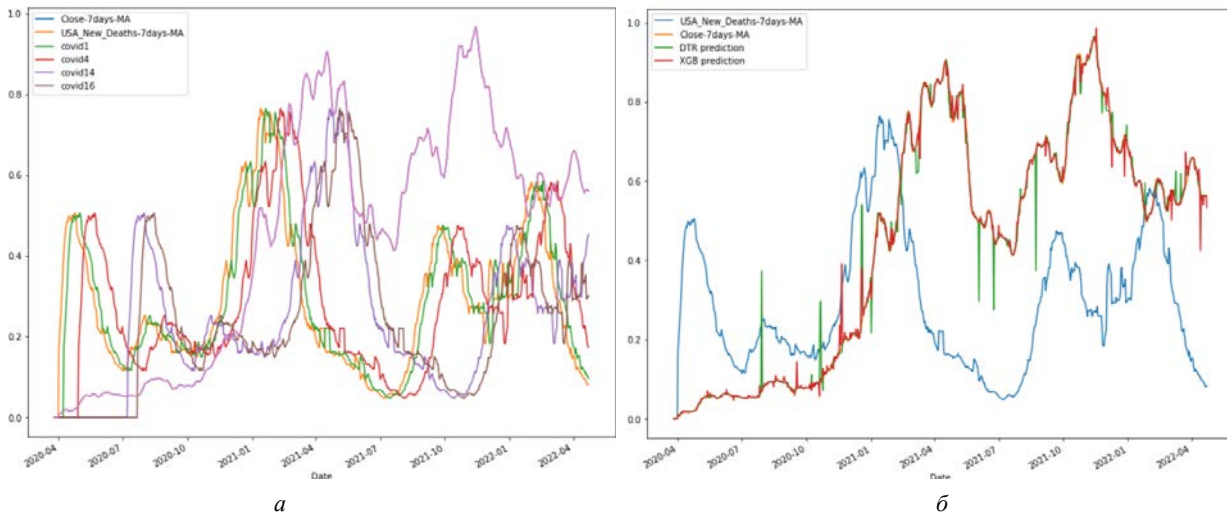


Рис. 6. Результат передбачення курсу біткоїна лише за даними про добовий приріст кількості смертей у США протягом 2020—2022 рр. [18] за моделлю (8): *a* — вхідні ознаки X та цільова змінна Y («Close») за моделлю (8) за $T = 1, 4, 14, 16$ днів; *б* — результат передбачення по всьому діапазону даних за моделями «Decision Tree Regressor» (глибина дерева — 13, відносна похибка на 20 % випадково вибраних тестових даних — 4,72 %, r_2_score — 0,969) і «XGBoost Regressor» (максимальна глибина дерев — 5, відносна похибка — 4,23 %, r_2_score — 0,989)

Практична цінність побудованої моделі низька, оскільки ціннішим є прогноз майбутніх значень курсу біткоїна, а тоді треба ще спочатку здійснити прогнозування величини добового приросту кількості смертей у США від ковіду, що теж є дуже складною задачею, враховуючи складності у збиранні вхідних даних. Головна цінність цього дослідження в тому, що зазначена первинна ознака є інформативною і її варто враховувати під час середньо- або довгострокового прогнозування біткоїну.

Запропонуємо інформаційну технологію для розв'язання поставленої задачі та розглянемо адекватніший і складніший приклад.

Створення інформаційної технології прогнозування курсу криптовалют на основі комплексної інженерії ознак

Для розв'язання поставленої задачі пропонується така технологія (схема алгоритму показана на рис. 7). Розроблено комплекс програмних засобів на Python, які дозволяють обробляти дані за таким алгоритмом, з урахуванням змісту етапів, описаних у вступі до цієї статті:

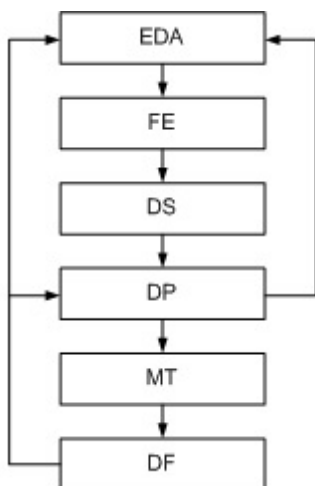


Рис. 7. Схема алгоритму запропонуваної інформаційної технології прогнозування курсу криптовалют на основі комплексної інженерії ознак

1. Розвідувальний аналіз даних (EDA), який включає в себе: аналіз особливостей криптовалюти, виділення впливових зовнішніх факторів та збирання даних по них, аналіз ознак даних та визначення періодичності ряду, підбір правил для його стаціонаризації, статистичний аналіз даних та порівняння з іншими факторами.

2. Конструювання ознак (FE): Комплексна інженерія ознак: формування нових ознак за системою правил (7).

3. Формування датасетів (DS): тренувального (для побудови моделей), валідаційного (для вибору оптимальної моделі) та тестового (для оцінювання точності оптимальної моделі, коли для її побудови використовуються тренувальний і валідаційний датасети разом). Іноді, тестовим датасетом є реальне майбутнє, яке на момент прогнозування є невідомим [10]. У разі довгострокового прогнозування реального майбутнього — прогнозування значень ознак, відібраних у п. 3 для тестового датасету, наприклад з використанням методів аналізу часових рядів та однієї з відомих Python-бібліотек: Facebook Prophet, ARIMA тощо або бібліотек, які використовують та вибирають оптимально одразу серед багатьох моделей: AutoML [35], Darts [36] тощо [37]—[39].

4. Передоброблення даних (DP): як правило, це — стандартизація, тобто центрування випадкових величин (віднімання середнього значення і ділення на середньоквадратичне відхилення), для того, щоб алгоритми машинного навчання

ня працювали ефективніше. Перевірка даних на простих моделях. Видалення малоцінних ознак. Якщо точність надто погана, тоді — перехід до п. 1.

5. Побудова моделей та їх тюнінг (MT). Можливо, побудова ансамблів моделей та їх стакинг (зважене усереднення прогнозів за багатьма моделями).

Найвикористовуванішими моделями машинного навчання в задачі прогнозування цін на криптовалюту є такі: рекурентна неймережа типу «довга короткострокова пам'ять» (Long Short-Term Memory — LSTM), згортова нейронна мережа (Convolutional Neural Network — CNN), багатошаровий перцептрон (Multilayer Perceptron — MLP), випадковий ліс (Random Forest — RF), градієнтний бустінг (Extreme Gradient Boosting — XGBoost), метод опорних векторів (Support Vector Machine — SVM) та ін.

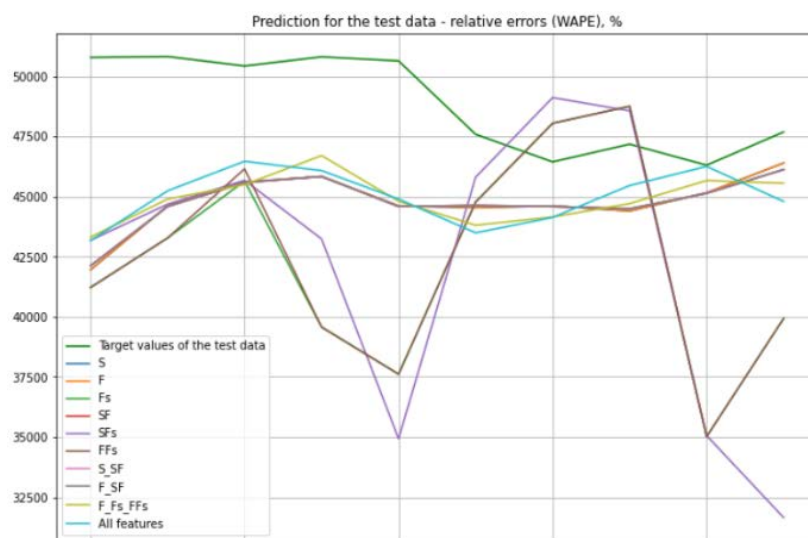
6. Прогнозування даних (DF) та аналіз різних метрик, залежно від типу задачі. В регресійних задачах популярними є метрики RMSE (корінь середнього квадрату похибки, тобто поелементного відхилення прогнозних значень від справжніх), косинус (відхилення вектора прогнозних значень від справжніх), MAE (середня абсолютна похибка), r2_score бібліотеки Sklearn.metrics або відносна похибка — це добуток 100, кількості значень цільової ознаки і значення похибки MAE, поділений на суму реальних значень цільової ознаки, у %). У класифікаційних — accuracy_score або f1_score. Оцінювання викидів, тобто значень найбільших похибок у певні моменти часу. У разі, якщо викиди значні чи їх забагато, то перехід на п. 1 або 4.

Приклад застосування запропонованої інтелектуальної технології

Проаналізуємо точність запропонованої інформаційної технології на даних по курсу біткоїна за щодобовими даними 2020—2021 років та порівняємо точність прогнозування значень «Close» на 10 діб вперед для різних комбінацій видів операцій. Використаємо модель «XGBoost Regressor» та триразову кросвалідацію за допомогою функції GridSearchCV з наперед підібраними теж за її допомогою оптимальними параметрами на усіх ознаках одночасно. Результати розрахунків показані на рис. 8 [40].

FE_model	WAPE,%	RMSE
All features	7.92	4399.5353
F_Fs_FFs	8.1	4434.0493
S	8.38	4734.6903
F	8.38	4734.6903
SF	8.39	4710.5739
S_SF	8.39	4710.5739
F_SF	8.39	4710.5739
FFs	14.46	8140.8537
Fs	14.56	8169.0531
SFs	15.32	9041.6676

а



б

Рис. 8. Результати прогнозування курсу біткоїна за даними 2020—2021 рр. на 10 діб вперед за допомогою моделі за запропонованою інформаційною технологією [40]: а — похибки моделей для різних комбінацій ознак; б — графіки тестових даних і прогнозів, зробленими за цими моделями

Аналіз результатів на рис. 8а показує, що саме використання і первинних, і вторинних, і вторинних на основі вторинних ознак дає найменшу похибку, а отже, підвищує точність прогнозування. Отже, поставлена мета досягнута.

Висновки

Проведено аналіз груп зовнішніх та внутрішніх факторів потенційного впливу на криптовалютний ринок. В першу чергу, це — первинні дані криптовалюти ОНЛС-показники та обсяги торгів. На основі цих даних будуються індикатори технічного аналізу, які також активно використовую-

ються спеціалістами для прогнозування динаміки фінансових інструментів. В іншій групі внутрішніх факторів — показники фінансового аналізу, які характеризують специфіку крипторинку та самого блок-чейна. Ще одна група зовнішніх факторів включає світові фондові індекси та енергоресурси. В окрему групу факторів можна віднести ті, що напряму не пов'язані з крипторинком, але можуть суттєво на нього вплинути. Це — події і новини щодо них про, наприклад, початок військових дій, зміну кліматичних умов природні катаклізми, ухвали урядів країн тощо.

Запропоновано загальну математичну модель формування факторів для задачі прогнозування курсу криптовалют, яка полягає у генерації великої кількості вторинних ознак на основі простих математичних, алгоритмічних та статистичних трансформацій з подальшим відбором найрелевантніших з них. Більше того, технологія передбачає синтез нових вторинних ознак на основі інших вторинних ознак, за окремими винятками, які формалізовано у вигляді системи правил. Це дозволить зменшити перенавчання моделі та покращити її узагальнюючу здатність.

Запропоновано інформаційну технологію прогнозування курсу криптовалют, яка полягає у комплексному підході до інженерії ознак. Експериментально доведено доцільність запропонованого підходу, оскільки точність моделі з використанням саме комплексу первинних, вторинних і вторинних на основі інших вторинних ознак виявилися найкращою.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] N. P. Patel, et al., "Fusion in Cryptocurrency Price Prediction: A Decade Survey on Recent Advancements, Architecture, and Potential Future Directions," *IEEE Access*, vol. 10, pp. 34511-34538, 2022, <https://doi.org/10.1109/ACCESS.2022.3163023>.
- [2] M. Chen, and N. Narwal, "Predicting price changes in ethereum," *Int. J. Comput. Sci. Eng.*, vol. 4, pp. 975, Apr. 2017.
- [3] T. Phaladisailoed, and T. Numnonda, "Machine learning models comparison for bitcoin price prediction," *Proc. 10th Int. Conf. Inf. Technol. Electr. Eng. (ICITEE)*, Jul. 2018, pp. 506-511.
- [4] *Bitcoin price prediction using Machine Learning*. [Online]. Available: <https://medium.com/@rohansawant7978/forecasting-of-bitcoin-price-using-machine-learning-deep-learning-techniques-93bf662f46ab>. Accessed on: April 7, 2022.
- [5] E. Akyildirim, A. Goncu, and A. Sensoy, "Prediction of cryptocurrency returns using machine learning," *Annals of Operations Research*, no. 297, pp. 3-36, 2021. <https://doi.org/10.1007/s10479-020-03575-y>.
- [6] Zheshe Chen, Chunhong Li and Wenjun Sun, "Bitcoin price prediction using machine learning: An approach to sample dimension engineering," *Journal of Computational and Applied Mathematics*, v. 365, 2020. <https://doi.org/10.1016/j.cam.2019.112395>.
- [7] E. Almasri, and E. Arslan, "Predicting cryptocurrencies prices with neural networks," in *Proc. 6th Int. Conf. Control Eng. Inf. Technol. (CEIT)*, pp. 1-5, Oct. 2018.
- [8] Franco Valencia, Alfonso Gómez-Espinosa, and Benjamín Valdés-Aguirre. "Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning," *Entropy*, 21, no. 6: 589, 2019. <https://doi.org/10.3390/e21060589>.
- [9] Mingxi Liu, Guowen Li, Jianping Li, Xiaoqian Zhu and Yinhong Yao, "Forecasting the price of Bitcoin using deep learning," *Finance Research Letters*, vol. 40, issue C, 2021. <https://doi.org/10.1016/j.frl.2020.101755>.
- [10] *Kaggle Competition "G-Research Crypto Forecasting"*, 2021. [Online]. Available: <https://www.kaggle.com/competitions/g-research-crypto-forecasting>. Accessed on: April 7, 2022.
- [11] *Blockchain & crypto*. [Online]. Available: <https://academy.binance.com>. Accessed on: April 7, 2022.
- [12] Lukas Menkhoff. "The use of technical analysis by fund managers: International evidence," *Journal of Banking and Finance*, vol. 34, issue 11, p. 2573-2586, 2010. <https://doi.org/10.1016/j.jbankfin.2010.04.014>.
- [13] *Bitcoin technical analysis for beginners*. [Online]. Available: <https://www.moneycontrol.com/msite/wazirx-cryptocontrol-articles/bitcoin-technical-analysis-for-beginners/>. Accessed on: April 7, 2022.
- [14] *Major world market indices*. [Online]. Available: <https://www.investing.com/indices/major-indices>. Accessed on: April 7, 2022.
- [15] *Commodities trading: an overview*. [Online]. Available: <https://www.investopedia.com/investing/commodities-trading-overview>. Accessed on: April 7, 2022.
- [16] *Tsfresh documentation*. [Online]. Available: <https://tsfresh.readthedocs.io/en/latest/>. Accessed on: April 7, 2022.
- [17] *TSFEL documentation*. [Online]. Available: <https://tsfel.readthedocs.io/en/latest/>. Accessed on: April 7, 2022.
- [18] Vitalii Mokin. *Kaggle Notebook «BTC & COVID-19 in USA : EDA & Prediction»*, April 2022. [Online]. Available: <https://www.kaggle.com/vbmokin/btc-covid-19-in-usa-eda-prediction>. Accessed on: April 7, 2022.
- [19] *Autofeat library*. [Online]. Available: <https://github.com/cod3licious/autofeat>. Accessed on: April 7, 2022.
- [20] *Featuretools*. [Online]. Available: <https://www.featuretools.com/>. Accessed on: April 7, 2022.
- [21] *Tsaug documentation*. [Online]. Available: <https://tsaug.readthedocs.io/en/stable/>. Accessed on: April 7, 2022.
- [22] *Gplearn documentation*. [Online]. Available: <https://gplearn.readthedocs.io/en/stable/>. Accessed on: April 7, 2022.
- [23] *Featuretools documentation*. [Online]. Available: <https://featuretools.alteryx.com/en/stable>. Accessed on: April 7, 2022.
- [24] R. C. Phillips and D. Gorse, "Predicting cryptocurrency price bubbles using social media data and epidemic modeling," *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, pp. 1-7, Nov. 2017.
- [25] A. Aggarwal, I. Gupta, N. Garg and A. Goel, "Deep learning approach to determine the impact of socio economic factors on bitcoin price prediction," *Proc. 12th Int. Conf. Contemp. Comput. (IC3)*, pp. 1-5, Aug. 2019.

- [26] J. Abraham, D. Higdon, J. Nelson and J. Ibarra, "Cryptocurrency price prediction using tweet volumes and sentiment analysis," *SMU Data Sci. Rev.*, vol. 4, pp. 1, Apr. 2018.
- [27] M. M. Patel, S. Tanwar, R. Gupta and N. Kumar, "A deep learning-based cryptocurrency price prediction scheme for financial institutions," *J. Inf. Secur. Appl.*, vol. 55, Dec. 2020.
- [28] S. Khuntia and J. Pattanayak, "Adaptive market hypothesis and evolving predictability of bitcoin," *Econ. Lett.*, vol. 167, pp. 26-28, Dec. 2018.
- [29] C. Gurdgiev, and D. O'Loughlin, "Herding and anchoring in cryptocurrency markets: Investor reaction to fear and uncertainty," *J. Behav. Exp. Finance*, vol. 25, Mar. 2020.
- [30] V. L. Tran, and T. Leirvik, "Efficiency in the markets of crypto-currencies," *Finance Res. Lett.*, vol. 35, Aug. 2020.
- [31] O. Angela, and Y. Sun, "Factors affecting cryptocurrency prices: Evidence from ethereum," *Proc. Int. Conf. Inf. Manage. Technol. (ICIMTech)*, pp. 318-323, Aug. 2020.
- [32] Vitalii Mokin, *Kaggle Notebook "BTC & Gold : EDA,"* 2022. [Online]. Available: <https://www.kaggle.com/code/vbmokin/btc-gold-eda>. Accessed on: April 7, 2022.
- [33] *Bitcoin rainbow price chart.* [Online]. Available: <https://www.blockchaincenter.net/bitcoin-rainbow-chart/>. Accessed on: April 7, 2022.
- [34] Б. І. Мокін, В. Б. Мокін, і О. Б. Мокін. «Практикум для самостійної роботи студентів з навчальної дисципліни «Методологія та організація наукових досліджень», Частина 1: від постановки задачі до синтезу та ідентифікації математичної моделі». Вінниця, Україна: ВНТУ, 2018, 179 с. [Електронний ресурс]. Режим доступу: https://ecopy.posibnyky.vntu.edu.ua/txt/2018/Mokin_Pos_%D0%A1%D0%A0%D0%A1_%D0%9C%D0%9E%D0%9D%D0%94%20-%20p008.pdf. Дата звернення: Квітень 7, 2022.
- [35] *AutoML documentation.* [Online]. Available: <https://www.automl.org/automl/>. Accessed on: April 7, 2022.
- [36] *Time series made easy in Python.* [Online]. Available: <https://unit8co.github.io/darts/>. Accessed on: April 7, 2022.
- [37] В. Б. Мокін, О. В. Слободянюк, О. М. Давидюк, і Д. О. Шмундяк, «Інформаційна технологія пошуку можливих джерел підвищеного забруднення річки з використанням моделі Prophet.» *Вісник Вінницького політехнічного інституту*, № 4, с. 15-24, 2020. <https://doi.org/10.31649/1997-9266-2020-151-4-15-24>.
- [38] О. Б. Мокін, В. Б. Мокін, і Б. І. Мокін, «Метод ідентифікації моделі авторегресії-ковзного середнього АРКС(Р, Q) з довільними значеннями порядків Р, Q, який узагальнює методику Юла–Уокера.» *Наукові праці Вінницького національного технічного університету*, № 2, 2014.
- [39] Pratik Gandhi, *7 libraries that help in time-series problems, 2021.* [Online]. Available: <https://towardsdatascience.com/7-libraries-that-help-in-time-series-problems-d59473e48ddd>. Accessed on: April 7, 2022.
- [40] Vitalii Mokin, *Kaggle Notebook. "BTC Growth Forecasting with Advanced FE for OHLC."* [Electronic resource]. Available: <https://www.kaggle.com/code/vbmokin/btc-growth-forecasting-with-advanced-fe-for-ohlc>.

Рекомендована до друку кафедрою системного аналізу та інформаційних технологій ВНТУ

Стаття надійшла до редакції 21.04.2022

Мокін Віталій Борисович — д-р техн. наук, професор, завідувач кафедри системного аналізу та інформаційних технологій, e-mail: vbmokin@gmail.com ;

Жуков Сергій Олександрович — канд. техн. наук, доцент кафедри системного аналізу та інформаційних технологій, e-mail: sazhukov@gmail.com ;

Куперштейн Леонід Михайлович — канд. техн. наук, доцент кафедри захисту інформації, e-mail: kupershtein.lm@gmail.com ;

Слободянюк Олена Валеріївна — канд. пед. наук, доцент кафедри опору матеріалів, теоретичної механіки та інженерної графіки, e-mail: olenas8@gmail.com .

Вінницький національний технічний університет, Вінниця

V. B. Mokin¹
S. O. Zhukov¹
L. M. Kupershtein¹
O. V. Slobodianiuk¹

Information Technology for the Cryptocurrency Rate Forecasting on the Basics of Complex Feature Engineering

¹Vinnitsia National Technical University

The paper is devoted to the development of information technology for cryptocurrency exchange rate forecasting based on complex feature engineering. The peculiarity of this technology is a systematic approach to the feature selection. The

analysis of external and internal groups of factors of potential influence on the cryptocurrency market was carried out. The analysis of features that characterize changes in cryptocurrency rates showed that in addition to the basic primary features, which are available on many cryptocurrencies, more important for the further forecasting of cryptocurrency rates are secondary features derived from basic primary ones by applying various mathematical operations and/or algorithmic transformations to them. The analysis of a large number of sources showed that cryptocurrencies have several characteristics that have caused their great popularity and which should also be taken into account when forming and choosing features. The systematization of such key characteristics are carried out in the paper, and also it is offered how to formalize them in the form of features. It is suggested to formalize the features according to a systematic approach, according to the postulates of technical cybernetics, which state that any object of study can be represented as a black box (BB), which is in contact with the environment at five points, which can in a multidimensional case be sets of features or variables. A general mathematical model of the formation of these factors is given, which consists in generating a large number of secondary features based on simple mathematical, algorithmic, and statistical transformations with subsequent selection of the most relevant of them. The technology involves the synthesis of new secondary features based on other secondary features, with some exceptions, which are formalized as a system of rules. This will reduce the overfitting of the model and improve its generalizing ability.

To prove the efficiency of the developed technology, an example of its application based on the cryptocurrency bitcoin according to the daily data of 2020—2021 is considered. Studies and computer experiments have shown the efficiency of the suggested technology.

Keywords: cryptocurrency, feature engineering, forecasting, information technology, bitcoin.

Mokin Vitalii B. — Dr. Sc. (Eng.), Professor, Head of the Chair of System Analysis and Information Technologies, e-mail: vbmokin@gmail.com ;

Zhukov Serhii O. — Cand. Sc. (Eng.), Associate Professor of the Chair of System Analysis and Information Technologies, e-mail: sazhukov@gmail.com ;

Kupershtein Leonid M. — Cand. Sc. (Eng.), Associate Professor of the Chair of Information Protection, e-mail: kupershtein.lm@gmail.com ;

Stobodianiuk Olena V. — Cand. Sc. (Educ.), Associate Professor of the Chair of Strength of Materials, Theoretical Mechanics and Engineering Graphics, e-mail: olenas8@gmail.com