УДК 004.032.26:519.246.8:336.74

КОНЦЕПЦІЯ ПРОГНОЗУВАННЯ ЦІНОВИХ КОЛИВАНЬ КРИПТОВАЛЮТ НА ОСНОВІ МАШИННОГО НАВЧАННЯ

Захватаєв К.О.[1]; Шибаєва Н.О., к.т.н., доцент[2]; Грішин С.І., к.т.н., доцент[3]
[1, 2, 3] Національний університет «Одеська політехніка»[1, 2, 3]
Україна, Одеса
[1]kirillz925@gmail.com; [2]n.o.shybaieva@op.edu.ua; [3]grishin@op.edu.ua;

Анотація. *Запропоновано концепцію ML-прогнозування цін криптовалют на OHLCV (годинні/денні) з лаговaним індексом настроїв; застосовано лог-доходності, робастне очищення, нормалізацію та технічні/календарні ознаки. У схемі walk-forward порівняно лінійні моделі й LSTM: перші краще відтворюють тренди, другі чутливіші до імпульсів; настрої підвищують якість. Розроблено продукційний конвеєр реального часу для практичного впровадження.*
Ключові слова: *криптовалюти; часові ряди; машинне навчання; LSTM; регресія; Bitcoin.*

Вступна частина. Проблема оцінювання та прогнозування цін криптовалют ускладнюється високою волатильністю, регуляторною невизначеністю та чутливістю до зовнішніх інформаційних сигналів, зокрема новинних потоків, активності соціальних мереж і дій великих гравців. Традиційні фінансові інструменти не завжди коректно відображають нелінійність процесів та періодичні режимні зсуви, що зумовлює інтерес до методів машинного навчання, здатних моделювати складні залежності у багатовимірних просторах ознак.

Мета дослідження є з'ясування можливостей і обмежень методів машинного навчання для прогнозування цінових коливань криптовалют, емпіричне порівняння регресійних підходів і LSTM на реальних даних, а також формування рекомендацій щодо інтеграції таких моделей у сучасні інформаційні системи підтримки рішень.

Основна частина. Емпірична база складалася з OHLCV-рядів із фіксованим кроком (годинним та денним), синхронізованих і очищених від пропусків. Для стабілізації масштабу застосовано лог-доходності та робастні процедури виявлення викидів із подальшим м'яким обрізанням, після чого дані нормалізувалися в межах тренувального вікна, щоб уникнути витоку майбутньої інформації. Простір ознак формувався як поєднання сирих котирувань і обсягів із технічними індикаторами (ковзні середні різних довжин, коливальні показники, оцінки волатильності) та часовими календарними змінними. Для урахування позаринкового впливу сконструйовано індекс настроїв: текстові публікації очищалися, дедуплікувалися, нормалізувалися за мовою й джерелом, після чого кожному повідомленню присвоювався числовий настрій; агрегування у часові кошики виконувалося з вагами за свіжістю й релевантністю, а сам індекс зсували у минуле на один-декілька кроків для коректного причинно-часового зв'язку. Щоб моделювати динаміку, з ряду будувалися ковзні вікна довжини w із горизонтом прогнозу h (один крок уперед і багатокрокові варіанти), причому всі перетворення, зокрема масштабування, фітовалися лише на тренувальних підвибірках у кожному кроці walk-forward.

Математичну основу лінійної частини становило рівняння множинної регресії $y_t = \beta_0 + \sum_{i=1}^{p} \beta_i x_{i,t} + \varepsilon_t$, де у ролі ознак виступали як ринкові, так і текстово-сентиментні компоненти. Нелінійну динаміку й довготривалі залежності моделювала LSTM: на вхід

подавалися тензори $X_{t-w+1:t}$ із багатовимірними ознаками, а вихідним сигналом була точка або вектор прогнозу $\widehat{y_{t+h}}$. Для підвищення стійкості застосовувалися регуляризація, відкидання частини нейронів під час навчання та рання зупинка за валідаційною втратою. Оптимізація проводилася градієнтними методами зі зменшенням швидкості навчання на плато; у багатокроковому режимі розглядалися прямі та рекурсивні схеми формування горизонту. Оскільки фінансові ряди містять режимні зсуви, валідація будувалася як послідовне прокочування з рухомою межею між тренуванням і тестом, зберігаючи часовий порядок і уникаючи перетину інформації між сегментами. Якість оцінювалася метриками похибок на рівнях цін і доходностей, а також перевірялася стабільність знаку прогнозованого руху; окремо відслідковувалася деградація у волатильні періоди[2]. Порівняння моделей показало, що лінійні підходи забезпечують інтерпретованість і добре схоплюють повільні трендові компоненти, проте реагують інерційно на короткі імпульси. LSTM ефективніше вловлює локальні патерни та взаємодії між ознаками на вікні, однак потребує ретельної настройки довжини вікна, глибини й регуляризації, і чутлива до розривів режимів. Додавання індексу настроїв покращує прогноз у «звичайні» ринкові фази, коли текстові сигнали випереджають рух ціни; під час інформаційних шоків і стрибків волатильності спостерігається загальне падіння точності, що вимагає механізмів виявлення режимів та адаптивного перенавчання. З погляду практичного впровадження найбільш надійним виявляється потоковий конвеєр із регулярним оновленням ознак і періодичним walk-forward перенавчанням, реєстром версій моделей, моніторингом похибок і контролем дрейфу даних[3]. для підвищення стійкості до режимних зсувів доцільно поєднувати лінійні моделі як базову компоненту з LSTM або бустингом на резидуалах, а також використовувати згладжені та лаговані текстові сигнали, що зменшують ризик хибних сплесків. Така інтеграція забезпечує відтворюваність експериментів, прозорість життєвого циклу моделей і оперативну візуалізацію невизначеності, що є критично важливим у динамічних криптовалютних ринках.

Висновки. Методи машинного навчання становлять ефективний інструментарій для прогнозування цін криптовалют, однак жоден окремий підхід не забезпечує стабільно високої точності за всіх ринкових режимів. Найбільш перспективними є гібридні рішення, що поєднують статистичні моделі, нейронні мережі та текстово-сентиментні ознаки, працюють у режимі регулярного перенавчання та супроводжуються моніторингом якості. Інтеграція таких рішень у інформаційні системи з розвиненою аналітикою сприяє підвищенню обґрунтованості рішень і зниженню ризиків в умовах динамічних ринків.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
[2] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
[3] López de Prado, M. (2018). *Advances in Financial Machine Learning*. John Wiley & Sons.

Zakhvataiev K.O., Shybaieva N.O., Grishin S.I.

Concept of predicting cryptocurrency price fluctuations based on machine learning

Abstract. *A concept of ML forecasting of cryptocurrency prices on OHLCV (hourly/daily) with a lagged sentiment index is proposed; log returns, robust cleaning, normalization, and technical/calendar features are applied. In the walk-forward scheme, linear models and LSTM are compared: the former better reproduce trends, while the latter are more sensitive to impulses; sentiment improves quality. A real-time production pipeline has been developed for practical implementation.*

Keywords: *cryptocurrencies; time series; machine learning; LSTM; regression; Bitcoin.*