

РАЗРАБОТКА ПОДХОДА К СТАНДАРТИЗИРОВАННОЙ ОЦЕНКЕ ВЛИЯНИЯ НОВОСТНОГО ФОНА НА КРАТКОСРОЧНУЮ ДИНАМИКУ ЦИФРОВЫХ АКТИВОВ: БЕНЧМАРКИНГ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ И КЛАССИЧЕСКИХ ВРЕМЕННЫХ РЯДОВ

Довгополик В. С. (СПБГУ)

**Научный руководитель – кандидат физико-математических наук, доцент кафедры
экономической кибернетики Подкорытова О. А. (СПБГУ)**

Введение

Криптовалюты представляют из себя высоковолатильные цифровые активы, которые проявляют сильную чувствительность к информационным поводам (регуляторные события, новости о проблемах и взломах, высказывания лидеров мнений) [4]. Классический анализ трендов на крипторынке включает в себя базовые эконометрические модели (напр. ARIMA, GARCH), которые плохо улавливают нелинейные эффекты от новостных шоков [3, 4]. В следствие развития методов машинного обучения и обработки естественного языка возникает гипотеза о том, что их использование позволит более эффективно улавливать краткосрочные шоки рынка на информационные события за счет учета нелинейных эффектов и более глубокой обработки текстовых данных [1, 3]. Однако существует проблема в применении методов машинного обучения: в существующих исследованиях влияния информационного фона на криптовалюты с помощью ML-моделей подходы, предлагаемые авторами, имеют разрозненный характер, что затрудняет сопоставимость результатов с базовыми моделями, а также с альтернативными исследованиями, что создает условия, в которых вместо движения к универсальному решению исследователи вынуждены формировать подходы «с нуля» с дальнейшей адаптацией под конкретный контекст. Таким образом, задача разработки стандартизированной методологии сравнения моделей для анализа поведения криптовалюты в краткосрочном периоде остается актуальной [2].

Основная часть

Данное исследование представляет собой первый этап разработки универсального алгоритма для анализа влияния новостного фона на криптовалюты. На текущем этапе решается задача бенчмаркинга для сравнения прогнозной способности ансамблевых методов (Random forest, XGBoost), нейросетевых архитектур (LSTM) и базовой модели временных рядов (ARIMA) при учете текстовых данных [2, 4].

Актуальность работы обусловлена необходимостью перехода от разрозненных исследований к созданию робастной методологии анализа для алгоритмической торговли и риск-менеджмента [2]. В отличие от уже существующих подходов, часто игнорирующих текстовые данные или использующие тональные оценки без привязки к событиям [1], предлагается двухэтапный подход:

1. Извлечение из текстовых данных признаков: парсинг новостей за определенных период и анализ с применением предобученных NLP-моделей на класс тональности, сущности события с последующим формированием количественных признаков: «влиятельность» события, скорость распространения.
2. Построение и сравнение моделей для прогноза в контролируемых условиях: базовая ARIMA как бенчмарк учета автокорреляции ряда [3, 4]; Random forest и XGBoost с использованием в качестве признаков колебания цены и агрегирование метрики новостного фона [1, 3]; LSTM для более глубокого анализа, так как способна хранить

в памяти долгосрочные зависимости между изменениями цен и последовательностью событий [3, 4].

Основными метриками сравнения моделей предлагается использовать стандартные MAPE/RMSE и направление движения [1, 4], а также прибыльность/убыточность простой торговой стратегии основанной на сигналах модели. Для подтверждения того, что именно информационный фон, а не динамика самого ряда, влияет на улучшение точности прогноза при использовании моделей машинного обучения, будут использованы методы каузального анализа (Тест Грейнджера на причинность). Это позволит провести прямое сопоставление эффективности моделей в одинаковых условиях на двух языковых корпусах.

Выводы

В ходе исследования разработан воспроизводимый подход к сравнительной оценке методов краткосрочного прогнозирования на рынке криптовалют с учетом новостного фона. Установлено, что предложенный бенчмарк позволяет измерить превосходство моделей машинного обучения над базовой: снижение ошибки прогноза RMSE на 15%-20% относительно базовой модели и повышения точности предсказаний с 50% (случайное угадывание) до 60% и выше. Сравнительная оценка моделей представлена на основе русскоязычного и англоязычного потоков новостей для рынка криптовалют. Результаты работы представляют собой первый шаг к разработке стандартизированного алгоритма анализа крипторынка при влиянии новостного контекста. Предложенная система бенчмаркинга может стать основой для дальнейшего совершенствования используемых моделей – тестирование архитектур в единых условиях и сопоставление ключевых метрик может позволить перейти к системному накоплению знаний в данной области.

Литература

1. Dahal K.R., Gupta A., Pokhrel N.R. Predicting the Direction of NEPSE Index Movement with News Headlines Using Machine Learning // *Econometrics*. — 2024. — Vol. 12, No. 2. — 16.
2. Cheong N., Hsuen L.W., Ranjan S., Cambria E., Goh R.S.M., Chatteraj J. BForTFin: A Financial Domain-Aware Multiscale Evaluation Method for Time-Series Foundation Models // *ACM Digital Library*. — 2025.
3. Hiskiawan P., William J., Tio Jansel L.F. A Hybrid Data Science Framework for Forecasting Bitcoin Prices using Traditional and AI Models // *Journal of Applied Informatics and Computing*. — 2025. — Vol. 9, No. 5.
4. Katina J., Katin I., Komarova V. Cryptocurrency price forecasting: a comparative analysis of autoregressive and recurrent neural network models // *Entrepreneurship and Sustainability Issues*. — 2024. — Vol. 11, No. 4. — P. 425-436.