

АНАЛИЗ ДАННЫХ С КРИПТОВАЛЮТНЫХ БИРЖ ДЛЯ ИСПОЛЬЗОВАНИЯ В СИСТЕМАХ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ДЛЯ АРБИТРАЖНОГО ТРЕЙДИНГА

А.Г. Парамонов, О.В. Оксюта

ФГБОУ ВО «Воронежский государственный лесотехнический университет
имени Г.Ф. Морозова»

Целью данной статьи является определение возможностей и способов анализа данных по существующим и реализованным предложениям на криптовалютных биржах для применения в системе поддержки принятия решений для арбитражного трейдинга. Выделены основные источники информации, на их основе сформирована методика отбора релевантных признаков для математических моделей. Проведена оценка применимости полученных результатов, которая показала пригодность данной модели к использованию в системах поддержки принятия решений.

Ключевые слова: арбитраж, криптовалюта, машинное обучение, анализ данных, прогнозирование, метод наименьших квадратов, рекуррентные нейронные сети.

ANALYSIS OF DATA FOR FORECASTING SALES BY THE MOVING AVERAGE METHOD

A.G. Paramonov, O.V. Oksyuta

Voronezh State University of Forestry and Technologies named after G.F. Morozov

The purpose of this article is to identify the capabilities and methods for analyzing data from existing and implemented offerings on cryptocurrency exchanges for use in a decision support system for arbitrage trading. Key sources of information are identified, and a methodology for selecting relevant features for mathematical models is developed based on these sources. An applicability assessment of the obtained results is conducted, demonstrating the suitability of this model for use in decision support systems.

Key words: arbitrage, cryptocurrency, machine learning, data analysis, forecasting, least squares method, recurrent neural networks.

Арбитраж, по своей сути, является способом заработка, основанном на разнице между покупкой и продажей актива. Положительное сальдо операции есть чистая прибыль трейдера. Арбитраж фондового и криптовалютного рынка во многом схож между собой. Однако на традиционных биржах торговать по данной схеме сложнее, так как разница в котировках очень незначительна. Криптовалютный трейдинг дает больше возможностей для такого вида заработка из-за большей ликвидности.

Криптовалютные биржи предоставляют API для доступа к книгам ордеров. Книга ордеров — это стриминг-система реального времени, построенная на изменениях состояния, где каждая новая заявка, отмена или исполнение — это событие, влияющее на поведение других участников.

На уровне публичных API — например, WebSocket-интерфейса Binance или Bybit — стакан представляет собой последовательность срезов состояний и обновлений. Срез состояния содержит пары цен и объемов на стороне покупки и продажи. Обновления к нему можно получать через WebSocket-интерфейс.

В качестве данных для обучения будем использовать завершённые, а не открытые, сделки, так как именно они отражают реальный курс валюты.

Для сбора, обработки и анализа данных будем использовать язык программирования python.

После предварительной обработки данных применяются статистические модели и модели машинного обучения для выявления возможностей арбитража. Статистический арбитраж основан на принципе возврата к среднему для выявления расхождений цен. Пусть $P_i(t)$ представляет цену криптовалюты на бирже i в момент времени t . Спред между двумя биржами i и j можно выразить как:

$$S_{ij}(t) = P_i(t) - P_j(t)$$

Возможность арбитража возникает, когда абсолютное значение спреда превышает заданный порог τ :

$$|S_{ij}(t)| > \tau$$

Модели машинного обучения, такие как случайные леса и нейронные сети, предлагают более сложную альтернативу, улавливая сложные закономерности в данных. Эти модели обучаются на исторических данных цен, чтобы предсказывать будущие ценовые движения и определять оптимальные

точки для арбитражных сделок. Теория графов предоставляет ещё один мощный инструмент для моделирования криптовалютных рынков, где биржи представлены узлами, а ценовые взаимосвязи – рёбрами. Применяя такие алгоритмы, как кратчайший путь Дийкстры, трейдеры могут определять наиболее эффективные маршруты для проведения арбитражных сделок, минимизируя транзакционные издержки и задержки.

Учёт данных блокчейна в процесс обнаружения арбитража добавляет методологии дополнительный уровень сложности. Данные о транзакциях в блокчейне дают представление о загруженности сети, комиссиях за совершение операций и времени подтверждения, которые могут существенно влиять на выполнение арбитражных сделок. Например, в периоды высокой сетевой активности подтверждения транзакций могут задерживаться, что увеличивает риск изменённых цены или упущенных возможностей. Для учета этого в модель можно включить такие метрики блокчейна, как среднее время подтверждения блока и размер мемпула (объём неподтверждённых транзакций). Пусть $T_c(t)$ обозначает среднее время подтверждения в момент времени t , а $G_f(t)$ — среднюю комиссию за транзакцию. Для оценки возможности проведения арбитражной сделки на основе этих факторов можно определить взвешенную функцию $W(t)$:

$$W(t) = \alpha \cdot T_c(t) + \beta \cdot G_f(t)$$

где α и β — весовые коэффициенты, отражающие относительную важность времени подтверждения и транзакционных комиссий. Если $W(t)$ превышает определённый порог, сделка может быть признана слишком рискованной или дорогостоящей для исполнения.

Анализ настроений в социальных сетях и новостных сводках дополнительно повышает надёжность арбитражной модели, предоставляя ранние предупреждения о рыночных тенденциях, таких как изменения цены, регулирований, объемов продаж. Эту оценку можно интегрировать в арбитражную модель в качестве дополнительной входной переменной, позволяя трейдерам корректировать свои стратегии на основе ожидаемого поведения рынка.

В некоторых исследованиях сравнивается производительность различных стратегий машинного обучения, таких как ARIMA, SVR, RF-Regressor, XGBoost, MLP, LSTM, GRU, CNN или их комбинаций, LSTM+CNN, GRU+CNN. Наилучшие результаты дают рекуррентные нейронные сети, когда

прогноз формулируется как задача регрессии. В случае, когда прогноз определяется как задача классификации, лучший результат модели, построенные на алгоритмах с методом опорных векторов.

Для поиска возможностей арбитража мы будем использовать линейную регрессию, метод наименьших квадратов и LSTM алгоритм. Проводились научные исследования, которые определили, что преимущества сложности моделей в криптовалютах ограничены. Простейшие методы оказываются наиболее эффективными. В частности, часто игнорируемый метод наименьших квадратов заметно превосходит деревья регрессии и нейронные сети.

Особенностью LSTM алгоритма является его плохая способность к экстраполяции данных, то есть он показывает плохие результаты, когда приходится предсказывать тренд, выходящий за пределы диапазона значений обучающих данных. Соответственно такая модель хуже будет предсказывать исторические максимумы и минимумы валют.

В качестве набора данных возьмем стоимость биткоина в конце каждого дня на бирже Binance в период с 11.10.2024 по 11.10.2025. Задачей для модели будет построение прогноза на некоторое количество дней. В качестве вектора признаков будем использовать стоимость биткоина в долларах. Результат работы обученной модели на тестовых данных изображён на рисунке 1.

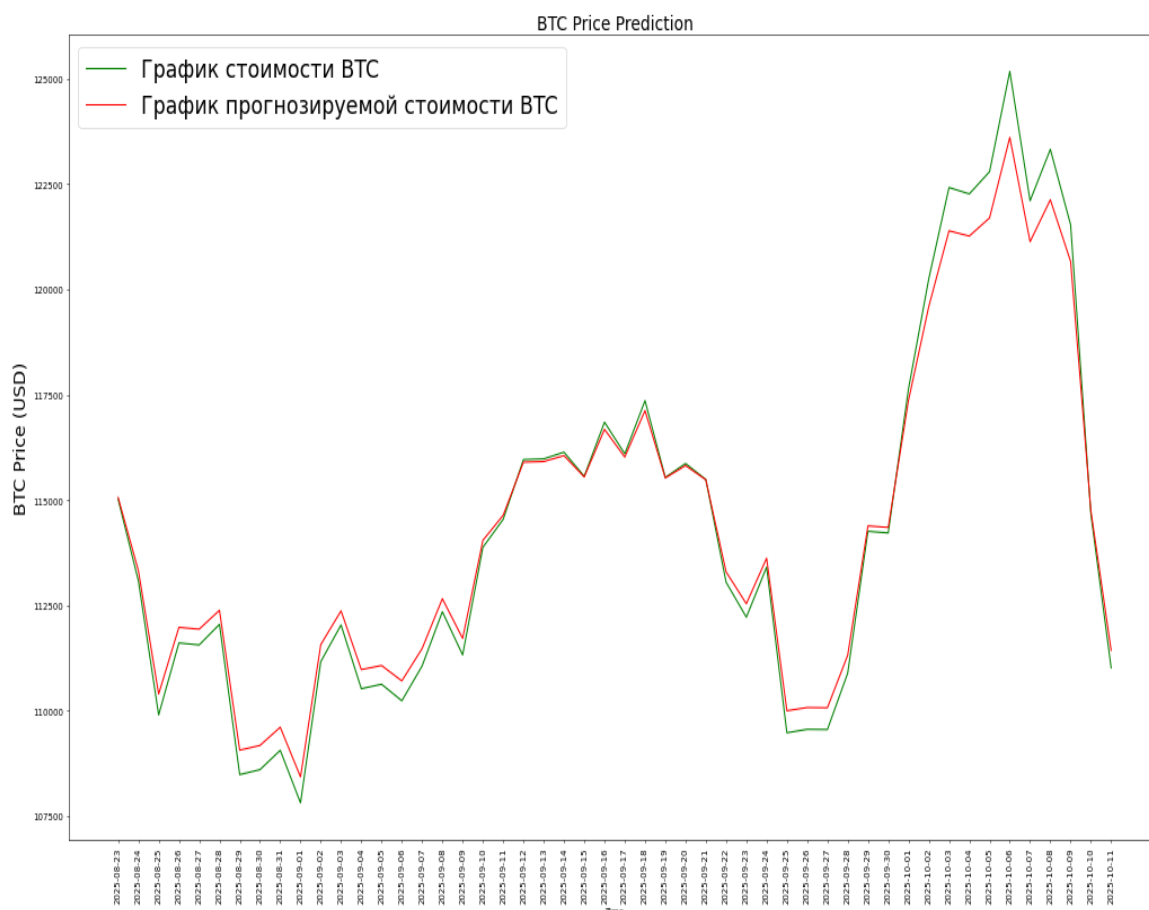


Рисунок 1 – Графики цены BTC

Среднеквадратичная ошибка: 534.393

Средняя абсолютная ошибка: 412.552

Средняя абсолютная процентная ошибка: 0.054

Таким образом, в данной работе разработана процедура отбора данных, переменных и рыночных индикаторов, релевантных для прогнозирования на рынке криптовалют, проведён анализ детализации, временных рамок и признаков данных, используемых для построения прогнозных моделей, реализована модель прогнозирования цен и их соотношений в криптовалютных парах с применением методов машинного обучения, включая линейную регрессию, метод наименьших квадратов и LSTM, а также Выполнена оценка применимости полученных результатов в системах поддержки принятия решений для арбитражного трейдинга. Для повышения точности и надежности модели можно учитывать дополнительные факторы, такие как данные блокчейна (время подтверждения транзакций, размер мемпула), а также анализ настроений в социальных сетях и новостных сводках.

Перспективы дальнейших исследований — интеграция мультимодальных данных, использование гибридных моделей и оптимизация транзакционных издержек с применением методов теории графов. Это позволит создать более адаптивные и устойчивые системы для поддержки автоматизированного арбитражного трейдинга.

Список литературы

1. Виетес, А. Машинное обучение, прогнозирование криптовалюты Ethereum и инвестиционные стратегии, основанные на знаниях / А. Виетес // Системы, основанные на знаниях – 2024 – №299 – DOI 10.1016/j.knosys.2024.112088
2. Генкин, А. Д. Блокчейн. Как это работает и что ждёт нас завтра. / А. Д. Генкин, А. А. Михеев ; Альпина Паблишер – Москва, 2017. – 592 с. – Библиогр. ISBN 978-5-9614-5046-0
3. Володин, С. Н. Сравнение эффективности статистического арбитража на фондовом рынке и рынке криптовалют / С. Н. Володин, М. В. Спиридонов // Валютное регулирование. Валютный контроль – Москва, 2018. – № 9 – С. 23-31

References

1. Genkin, A. D. Blockchain. How it works and what awaits us tomorrow. / A. D. Genkin, A. A. Mikheev; Alpina Publisher – Moscow, 2017. – 592 p. – Bibliography. ISBN 978-5-9614-5046-0
2. Volodin, S. N. Comparison of the effectiveness of statistical arbitrage in the stock market and the cryptocurrency market / S. N. Volodin, M. V. Spiridonov // Currency regulation. Currency control - Moscow, 2018. - No. 9 - P. 23-31
3. Viéitez, A. Machine learning Ethereum cryptocurrency prediction and knowledge-based investment strategies / A. Viéitez // Knowledge-Based Systems – 2024 – №299 – DOI 10.1016/j.knosys.2024.112088