

УДК 336.74-028.27:004.75

147. СРАВНИТЕЛЬНЫЙ АНАЛИЗ И ОЦЕНКА УСТОЙЧИВОСТИ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ПРИ ПРОГНОЗИРОВАНИИ КУРСОВ КРИПТОВАЛЮТ В УСЛОВИЯХ АВТОГЕНЕРАЦИИ ДАННЫХ

Салаш А. А., студентка группы 578101

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

*Ефремов А.А. – заведующий кафедрой экономической
информатики, кандидат экономических наук, доцент*

Аннотация. В условиях высокой волатильности крипторынка точное прогнозирование курсов осложняется эффектом накопления ошибок при автогенерации данных. Работа посвящена сравнительному анализу моделей ARIMA, LSTM и XGBoost и оценке их устойчивости на разных временных горизонтах. В качестве экспериментальной базы используются реальные посуточные котировки Ethereum, а точность оценивается по метрике MAPE. Эксперименты показывают, что ARIMA эффективна только краткосрочно, а сеть LSTM запаздывает на пиках. Наибольшую устойчивость и минимальную ошибку в условиях автогенерации данных демонстрирует алгоритм градиентного бустинга XGBoost, доказывая свое превосходство над классическими подходами. Результаты исследования помогают в создании надежных торговых систем.

Ключевые слова. Криптовалюта, прогнозирование курсов, временные ряды, волатильность, автогенерация данных, ARIMA, LSTM, XGBoost, градиентный бустинг, метрика MAPE, Ethereum.

Введение. В эпоху стремительного перехода к концепции Web 3.0 криптовалюты окончательно закрепили за собой статус полноправного элемента глобальной финансовой архитектуры. Согласно актуальной рыночной статистике, доминирующее положение продолжают удерживать Bitcoin (занимающий порядка 58–60% рынка) и Ethereum (около 12%), функционирующие как децентрализованные цифровые активы на базе технологии блокчейн [1]. Однако отсутствие регуляции со стороны центральных банков порождает их фундаментальное свойство – аномально высокую волатильность, при которой котировки активов способны непредсказуемо и кардинально меняться в интервале всего нескольких часов. Принятие некорректных решений в таких условиях неизбежно оборачивается для инвесторов и коммерческих предприятий сокрушительными финансовыми потерями.

В то же время растущая интеграция криптовалют в традиционные финансовые процессы усиливает потребность в точных и устойчивых методах прогнозирования их стоимости. Несмотря на значительное количество исследований, посвящённых анализу динамики цифровых активов, большинство существующих работ сосредоточено на краткосрочных прогнозах и не учитывает особенности реального торгового процесса [7].

Существующий массив научных исследований предлагает широкий спектр инструментов для решения данной проблемы: от классического технического анализа и эконометрических моделей временных рядов (ARIMA) до передовых алгоритмов глубокого обучения и рекуррентных нейросетей (LSTM) [2, 6].

Однако анализ литературы показывает, что в большинстве работ отсутствует комплексная оценка устойчивости моделей в условиях автогенерации данных – режима, при котором каждый следующий шаг прогноза строится на основе ранее предсказанных моделью значений. Именно этот режим наиболее точно отражает реальную биржевую практику, но одновременно является наиболее сложным для математических моделей из-за эффекта накопления ошибки.

Кроме того, в условиях постоянного изменения рыночной структуры многие из ранее предложенных алгоритмов демонстрируют тенденцию к чрезмерному сглаживанию результатов или накоплению ошибок при средне- и долгосрочном прогнозировании [8].

Исходя из выявленного исследовательского пробела, целью данной работы является сравнительный анализ трёх наиболее распространённых подходов – ARIMA, LSTM и XGBoost – с оценкой их устойчивости в условиях автогенерации данных и различного горизонта прогнозирования. Такой подход позволяет объективно определить, какие модели наиболее надёжны при работе с высоковолатильными криптовалютными временными рядами.

Традиционные статистические подходы и эконометрические модели. Исторически первыми инструментами для прогнозирования стоимости цифровых активов стали модели временных рядов [2]. Среди них выделяются:

Модель авторегрессии (AR), опирающаяся на линейную зависимость текущих данных от их прошлых значений. Её ключевое ограничение – требование стационарности исследуемого ряда.

Модель скользящего среднего (MA), ориентированная на учет случайных ошибок прошлых периодов, что помогает снизить влияние рыночного «шума».

Модель ARIMA (и её сезонная модификация SARIMA), объединяющая авторегрессию, интегрирование и скользящее среднее. Она демонстрирует универсальность при работе с нестационарными данными, однако правильный подбор её параметров требует экспертного опыта и значительных временных затрат [2].

Применение алгоритмов машинного обучения и ансамблевых методов. С развитием технологий на смену классическим статистическим методам пришли алгоритмы машинного обучения, способные выявлять нелинейные паттерны в больших массивах данных [5]. В частности:

Случайный лес (Random Forest) – ансамблевый метод, объединяющий множество деревьев решений для минимизации риска переобучения и повышения точности [3, 4].

Градиентный бустинг (в реализациях XGBoost, LightGBM, CatBoost) – метод последовательного построения деревьев, где каждая новая модель компенсирует ошибки предыдущих [5]. Эмпирические исследования на базе данных котировок Ethereum доказывают вычислительное превосходство градиентного бустинга над ARIMA как оптимального инструмента оперативного прогнозирования.

Современные тенденции и глубокое обучение (Нейросети и Трансформеры). Актуальные исследования 2025–2026 годов смещают фокус в сторону гибридных подходов и глубокого обучения (Deep Learning).

Для фиксации временных зависимостей активно применяются рекуррентные нейросети, такие как LSTM и GRU [6]. Кроме того, в практику прогнозирования внедряются модели-трансформеры, одновременно анализирующие множество сигналов. Важным вектором развития стало обогащение исторических ценовых данных внешними предикторами:

Анализ ликвидности – интеграция динамических показателей, таких как средневзвешенная по объему цена (VWAP) и соотношение объема к волатильности (VVR) [7].

Анализ рыночных настроений (Sentiment Analysis) – использование методов обработки естественного языка (NLP) для оценки тональности новостных лент и социальных сетей, что существенно улучшает прогнозирование резких скачков курса.

Несмотря на перспективность гибридных систем и моделей-трансформеров, их внедрение требует колоссальных вычислительных мощностей. В связи с этим в рамках данного исследования фокус сделан на сравнительном анализе классических и наиболее устойчивых к рыночному шуму архитектур (ARIMA, LSTM и XGBoost).

Экспериментальная база и методика исследования. Для проверки эффективности рассмотренных теоретических подходов и проведения их сравнительного анализа на практике, в данной работе использовались реальные исторические данные котировок криптовалюты Ethereum (ETH), извлеченные из открытых биржевых источников [1]. Выбор данного актива обусловлен его высокой ликвидностью и устойчивым вторым местом по капитализации на мировом рынке цифровых активов. Входной массив данных охватывал посуточные замеры цен закрытия, объемов торгов и волатильности за исследуемый период.

На этапе предобработки данных осуществлялось масштабирование признаков для приведения их к единому диапазону и их деление на обучающую и тестовую выборки. Особое внимание уделялось формированию лагов (временных сдвигов), которые выступали в качестве регрессоров для обучения моделей.

Оценка прогностической точности построенных моделей производилась на основе метрики средней абсолютной процентной ошибки (MAPE). Данный критерий позволяет объективно оценивать качество аппроксимации вне зависимости от масштаба абсолютных цен актива. Расчет метрики осуществлялся по следующей формуле (1):

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{\hat{y}_t - y_t}{y_t} \right|. \quad (1)$$

где y_t – фактическое значение в момент времени t ;
 $\hat{y}_t$ – прогнозное значение в момент времени t ;

n – количество наблюдений в тестовой выборке.

Результаты исследования и сравнительный анализ моделей. В ходе апробации моделей на реальных данных были получены различные результаты, подчеркивающие специфику каждого из подходов при работе с нестационарными и волатильными финансовыми рядами. Эконометрическая модель ARIMA продемонстрировала высокую точность на обучающей выборке при краткосрочном шаге прогнозирования (1 день). Однако при попытке построения прогноза на среднесрочную перспективу (7 дней) в условиях автогенерации данных (когда каждое последующее значение рассчитывается на основе ранее предсказанного моделью) ARIMA показала тенденцию к резкому сглаживанию тренда. Модель быстро переходила к прогнозированию средних значений, полностью игнорируя локальные экстремумы и волатильность рынка [2].

Рекуррентная нейронная сеть LSTM (Long Short-Term Memory) успешно справилась с задачей выявления долгосрочных зависимостей [6]. За счет наличия блоков памяти она смогла уловить общую направленность тренда Ethereum. Тем не менее на тестовой выборке модель LSTM продемонстрировала эффект запаздывания: её пиковые значения часто отставали от реальных рыночных скачков на 1–2 временных шага, что снижает её ценность для высокочастотных торговых стратегий [7]. Кроме того, данная модель оказалась наиболее требовательной к объему обучающей выборки и вычислительным ресурсам.

Наилучшие результаты по критерию минимизации ошибки MAPE, рассчитанной по формуле (1), показал алгоритм градиентного бустинга (XGBoost) [5]. Данная модель не только точнее всего описала обучающую выборку, но и сохранила высокую устойчивость на тестовом интервале в 7 дней. В отличие от статистических методов, XGBoost не сглаживал график до прямой линии, а продолжал генерировать колеблющиеся значения, близкие к реальной рыночной динамике. Способность эффективно обрабатывать нелинейные паттерны и устойчивость к переобучению делают ансамблевые методы на основе деревьев решений приоритетным выбором для оперативного прогнозирования курсов цифровых активов.

Заключение. В рамках проведенного исследования был осуществлен комплексный сравнительный анализ ключевых математических инструментов, применяемых для прогнозирования стоимости высоковолатильных цифровых финансовых активов. Рассмотренные подходы – от классических моделей временных рядов (ARIMA) до передовых алгоритмов машинного и глубокого обучения – обладают уникальной спецификой и разной степенью применимости в зависимости от горизонта планирования.

Результаты экспериментов на базе реальных котировок Ethereum подтвердили неэффективность чисто эконометрических методов при среднесрочном прогнозировании в условиях автогенерации данных из-за эффекта сглаживания тренда. В то же время было доказано алгоритмическое превосходство градиентного бустинга: модель XGBoost продемонстрировала наименьшую ошибку аппроксимации (MAPE) и максимальное сохранение динамики временного ряда.

Перспективой дальнейших исследований в данной области видится построение гибридных ансамблевых систем. Объединение математической строгости рекуррентных нейросетей с гибкостью градиентного бустинга, а также их обогащение внешними факторами фундаментального анализа (такими как индексы рыночных настроений), позволит минимизировать эффект накопления ошибки и выйти на качественно новый уровень точности прогнозирования динамики крипторынка.

Список литературы

1. Мебония, М.А. Технология блокчейн. Примеры блокчейна и его применение / М.А. Мебония // Вестник науки, и технологии. – 2022. – № 12. – С. 435–438.
2. Сафиуллин, М.Р. Методические подходы к прогнозированию динамики курса криптовалют с применением инструментов стохастического анализа (на примере биткоина) / М.Р. Сафиуллин, А.А. Абдукаева, Л.А. Ельшин // Вестник Института экономики Российской академии наук. – 2018. – № 5. – С. 161–173.
3. Богучаров, В.А. Классификация с обучением на основе информативного дерева решений / В.А. Богучаров // StudNet, Компьютерные и информационные науки. – 2020. – № 7. – С. 593–597.
4. Ахикян, А.И. Алгоритм машинного обучения адаптивный случайный лес и его применение / А.И. Ахикян, С.С. Данилюк // Вестник науки, Компьютерные технологии. – 2024. – № 6. – С. 1393–1401.
5. Дудченко, Д.В. Применение модели градиентного бустинга в задачах регрессии / Д.В. Дудченко // Вестник науки, Компьютерные технологии. – 2024. – № 6. – С. 1344–1347.
6. Кондратьева, Т.Н. Прогнозирование тенденции финансовых временных рядов с помощью нейронной сети LSTM / Т.Н. Кондратьева // Вестник евразийской науки, Информатика, вычислительная техника и управление. – 2017. – № 4. – С. 515–520.
7. Михайлова, С. С. Разработка торговой стратегии криптовалюты с применением методов машинного обучения / С. С. Михайлова, С. А. Сабирова // Computational Nanotechnology. – 2024. – Т. 11, № 2. – С. 11–21.
8. Назарова, В. В. Применение методов машинного обучения для прогнозирования нефтяных котировок / В. В. Назарова, Б. А. Лодягин, Ф. А. Круглов, А. В. Круглов // AlterEconomics. – 2025. – Т. 22, № 3. – С. 482–502.

UDC 336.74-028.27:004.75

COMPARATIVE ANALYSIS AND ASSESSMENT OF THE STABILITY OF MATHEMATICAL MODELS IN FORECASTING CRYPTOCURRENCY RATES IN THE CONDITIONS OF AUTOGENERATION OF DATA

Salash A. A., student gr. 578101

*Belarusian State University of Informatics and Radioelectronics
Minsk, Republic of Belarus*

*Efremov A.A. – Head of the Department of Economic Informatics,
Candidate of Economic Sciences, Associate Professor*

Annotation. In the highly volatile crypto market, accurate price forecasting is complicated by the accumulation of errors during automated data generation. This paper compares ARIMA, LSTM, and XGBoost models and assesses their robustness over different time horizons. Real daily Ethereum quotes are used as the experimental baseline, and accuracy is assessed using the MAPE metric. Experiments show that ARIMA is only effective in the short term, while the LSTM network lags at peaks. The XGBoost gradient boosting algorithm demonstrates the highest robustness and minimal error under automated data generation, proving its superiority over classical approaches. The results of the study can be used to develop reliable trading systems.

Keywords. Cryptocurrency, price forecasting, time series, volatility, data generation, ARIMA, LSTM, XGBoost, gradient boosting, MAPE metric, Ethereum.