

ПРИМЕНЕНИЕ ГРАДИЕНТНОГО БУСТИНГА ДЛЯ ПРОГНОЗИРОВАНИЯ СТОИМОСТИ ЦЕННЫХ БУМАГ

APPLICATION OF GRADIENT BOOSTING FOR STOCK PRICE PREDICTION

A. Solobuto
V. Pavlov

Summary. This study examines the application of gradient boosting as a tool for forecasting the value of securities. This machine learning method has proven to be one of the most effective algorithms for solving regression and classification problems due to its ability to model complex nonlinear relationships and its robustness against overfitting when parameters are properly tuned.

To improve prediction accuracy, a careful selection of informative financial and technical indicators that potentially influence the value of securities was conducted. Feature selection was performed based on both expert evaluations and statistical methods for analyzing variable significance. Additionally, the Parzen window method [1], a type of Bayesian optimization, was used for automated parameter tuning. This approach enables efficient exploration of parameter space.

Keywords: gradient boosting, forecasting, feature engineering, stocks, indicators.

Солобута Алексей Викторович

аспирант,

Московский финансово-юридический университет

29352138@s.mfua.ru

Павлов Валерий Анатольевич

Кандидат экономических наук, доцент,

Московский финансово-юридический университет

29359332@s.mfua.ru

Аннотация. В работе рассматривается применение градиентного бустинга в качестве инструмента для прогнозирования стоимости ценных бумаг. Данный метод машинного обучения зарекомендовал себя как один из наиболее эффективных алгоритмов для решения задач регрессии и классификации благодаря способности моделировать сложные нелинейные зависимости и устойчивости к переобучению при правильной настройке параметров.

С целью повышения точности прогнозирования был осуществлён тщательный отбор информативных финансовых и технических индикаторов, которые потенциально влияют на стоимость ценных бумаг. Отбор признаков осуществлялся как на основе экспертных оценок, так и с применением статистических методов анализа значимости переменных.

Также в работе для автоматизированного подбора параметров использован метод Парзеновского окна [1], являющийся одной из разновидностей байесовской оптимизации. Этот подход позволяет эффективно исследовать пространство параметров.

Ключевые слова: градиентный бустинг, задача прогнозирования, оптимизация параметров, ценные бумаги, индикаторы.

Введение

В задачах экономического прогнозирования может содержаться большое количество нелинейных зависимостей между различными показателями, что значительно усложняет возможность сделать надежный прогноз, однако с ростом вычислительных мощностей и улучшением алгоритмов машинного обучения, стали доступны новые подходы к решению данных задач, одной из которых является прогнозирование поведения на рынке ценных бумаг.

Банки, финансовые аналитики, хедж-фонды, а также другие участники рынка массово начинают использовать существующие системы принятия решений, а также разрабатывать собственные механизмы для предсказания финансовых показателей на бирже. В качестве одного из вариантов таких систем в данной работе будет рассмотрен метод градиентного бустинга, а также произведена калибровка параметров модели.

Анализ актуальных исследований в области

Применение градиентного бустинга рассматривается в работе «Исследование методов прогнозирования стоимости ценных бумаг на основе машинного обучения», в ней автор проводит сравнение нескольких методов машинного обучения: линейная регрессия, метод опорных векторов и градиентный бустинг, где градиентный бустинг показывает самый высокий показатель ошибки MAPE [2]. В качестве улучшения предлагается использовать набор признаков, отличный от описанных в указанной статье, где были выбраны параметры с линейной зависимостью, также было решено произвести калибровку гиперпараметров модели.

Преимущества и недостатки модели градиентного бустинга

Градиентный бустинг основан на последовательном построении ансамбля слабых моделей, что даёт ряд преимуществ и недостатков данному алгоритму. Среди сильных сторон в рамках данной задачи стоит выделить:

- 1) Возможность использовать нелинейные зависимости между данными.
- 2) Данный метод имеет устойчивость к выбросам в данных, что позволяет использовать его даже в случае аномальной ситуации на рынке.
- 3) Устойчивость к недообучению — даже на небольших данных модель улавливает связи между признаками.

К недостаткам относятся:

- 1) Сложная интерпретируемость — данный метод использует последовательный ансамбль моделей, что не дает возможности перевести полученную модель на понятный человеку язык.
- 2) Высокая вычислительная сложность — для обучения градиентного бустинга требуются значительные ресурсы, желательно с использованием GPU.
- 3) Чувствительность к переобучению — без калибровки гиперпараметров модель может адаптироваться под данные, что по итогу даст слабую предсказательную способность на новых данных.

Данный метод следует использовать при наличии серьёзных вычислительных ресурсов, а также при наличии нелинейных связей между признаками [3].

Отбор индикаторов

Для отбора индикаторов было принято решение, разделить все данные на три группы: планируется рост,

планируется незначительное колебание цены и планируется падение [4], после чего построить гистограммы для всех индикаторов и отобрать те, где явно выделяются представители отобранных групп. По итогу были выбраны следующие индикаторы: rsi, aroon oscillator, adx, TRIX, MFI, ultimate oscillator, cci, далее будут приведены описания данных индикаторов и их гистограммы:

- Индекс относительной силы (RSI) — индикатор технического анализа, определяющий силу тренда и вероятность его смены [5]. Данный индикатор может принимать значения от 0 до 100. Для задачи классификации используется строгая версия индикатора, в которой значение 80 означает перекупленность бумаги, а значение 20 недокупленность. Гистограмма изображена на рисунке 1.
- Ароон-осциллятор — это технический индикатор, используемый в анализе финансовых рынков для оценки силы и направления тренда, а также для определения возможных точек разворота, основан на двух линиях: Aroon Up и Aroon Down. Ароон-осциллятор представляет собой разницу между этими двумя линиями [6]. Гистограмма изображена на рисунке 2.
- Индикатор ADX (Average Directional Index, средний индекс направленного движения) — технический индикатор, используемый в анализе финансовых рынков для оценки силы тренда, независимо от его направления (восходящего или нисходящего) [7]. Гистограмма изображена на рисунке 3.

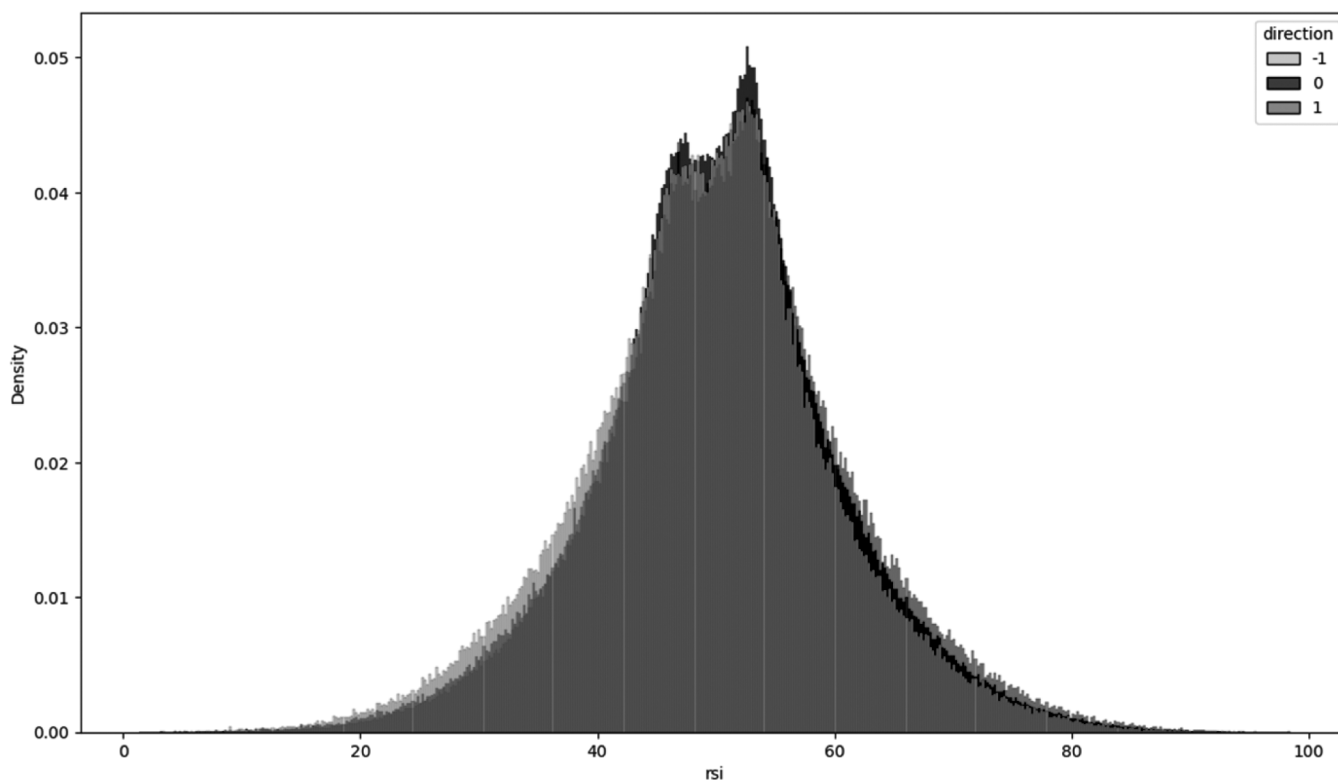


Рис. 1. Гистограмма индикатора rsi

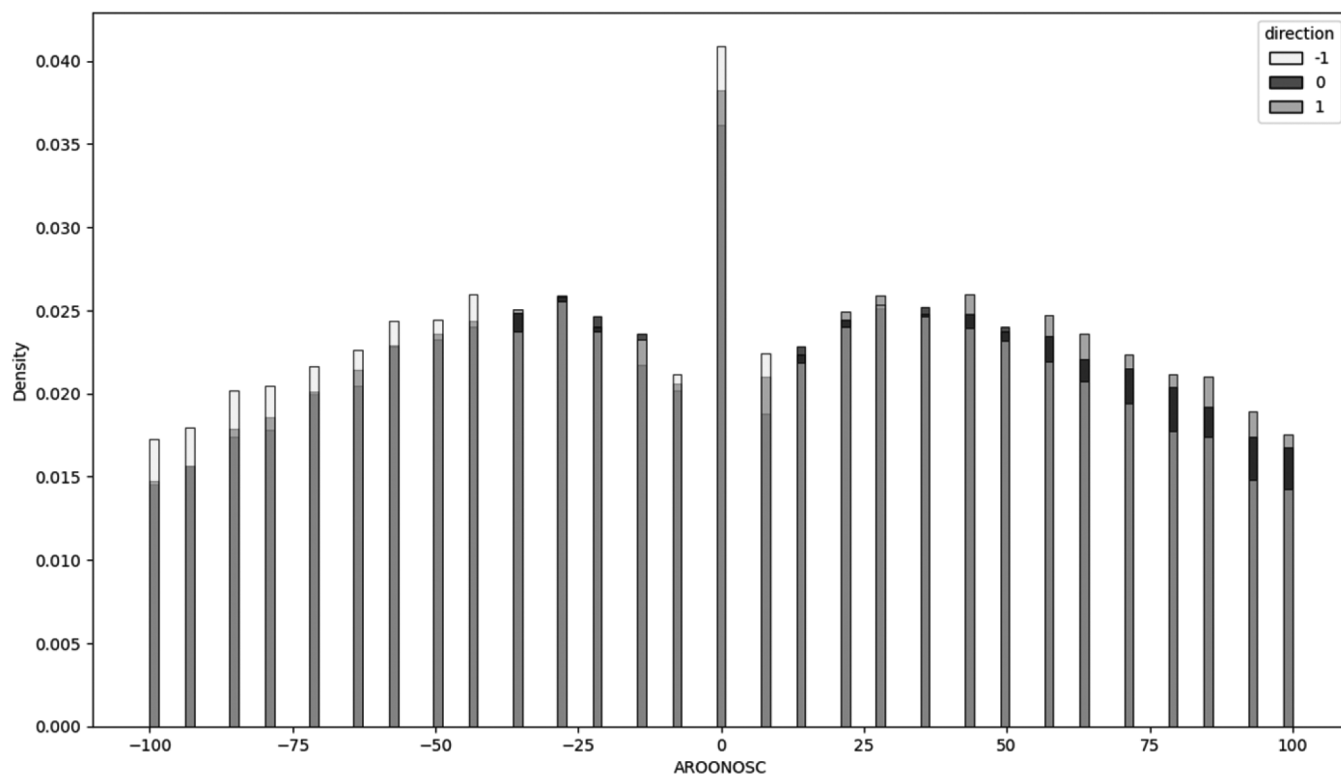


Рис. 2. Гистограмма индикатора aroon oscillator

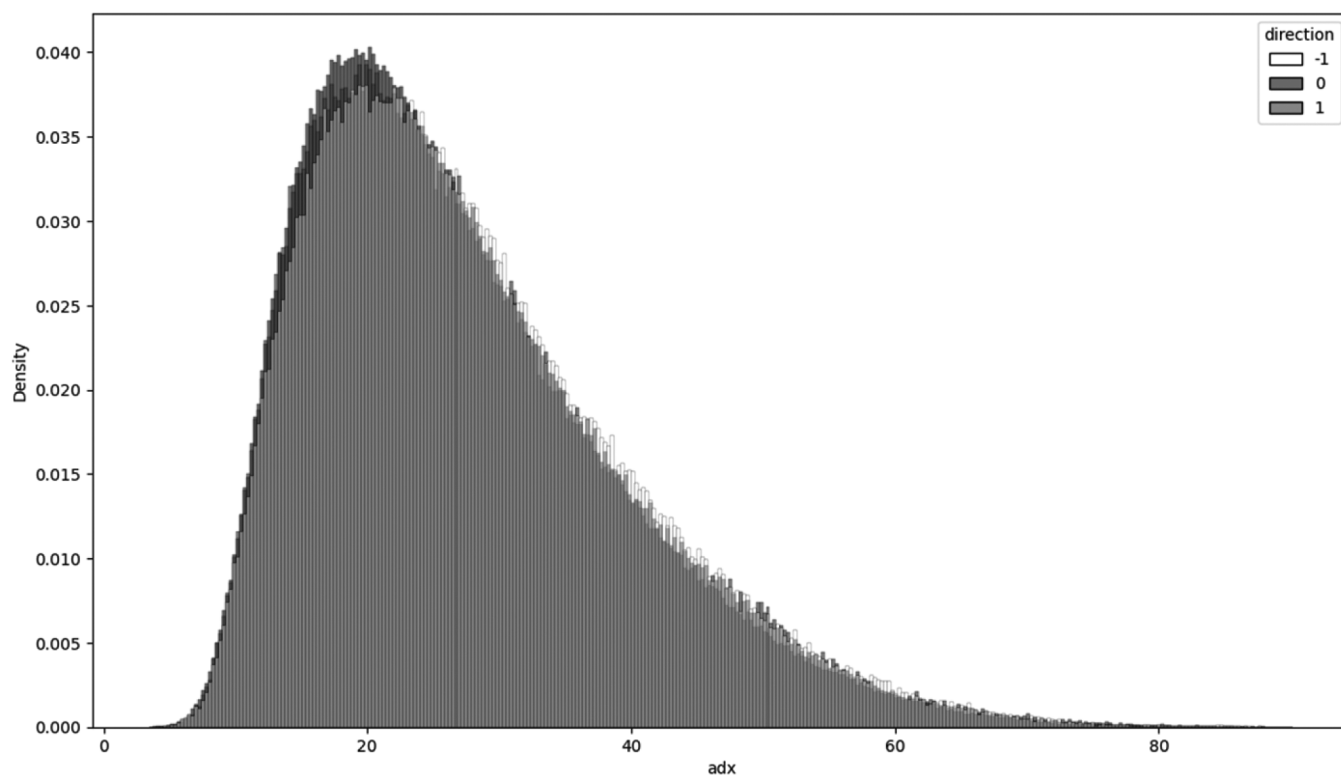


Рис. 3. Гистограмма индикатора adx

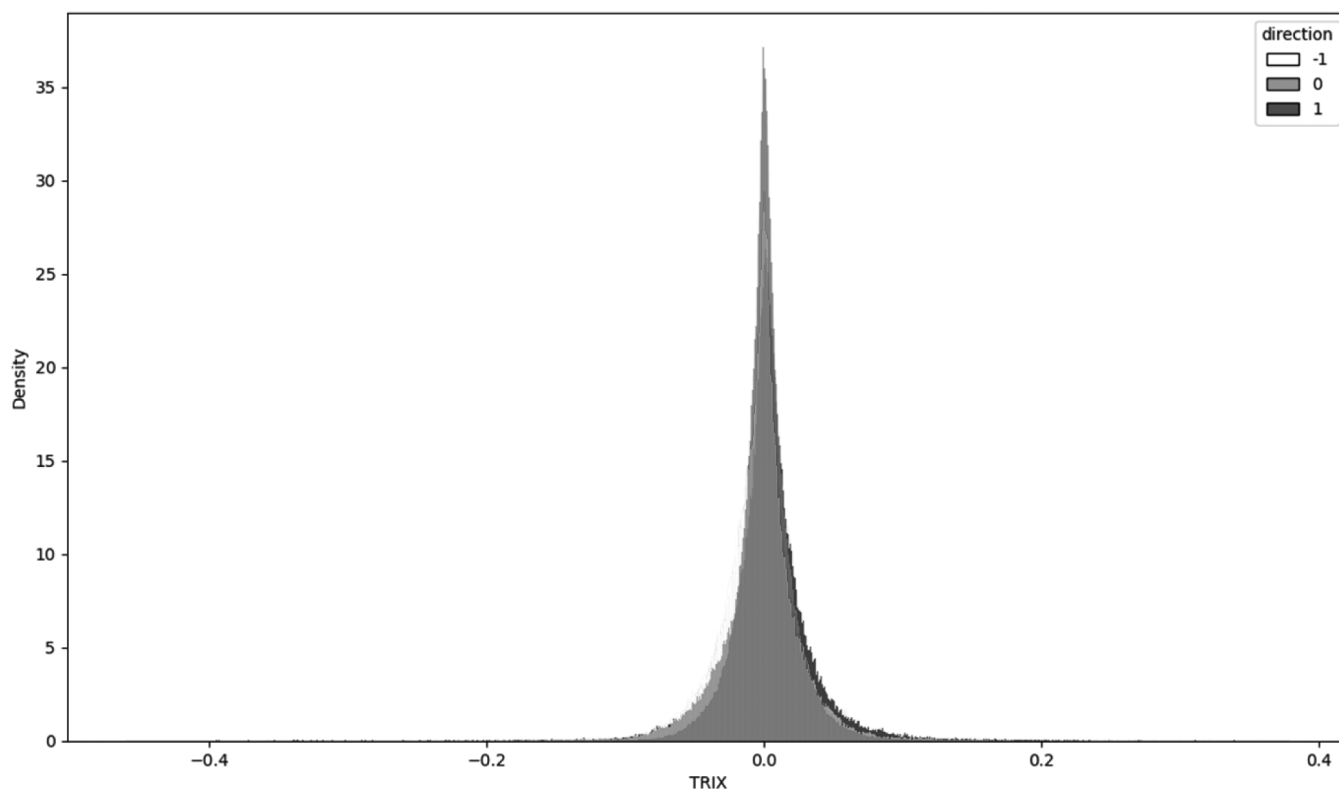


Рис. 4. Гистограмма индикатора TRIX

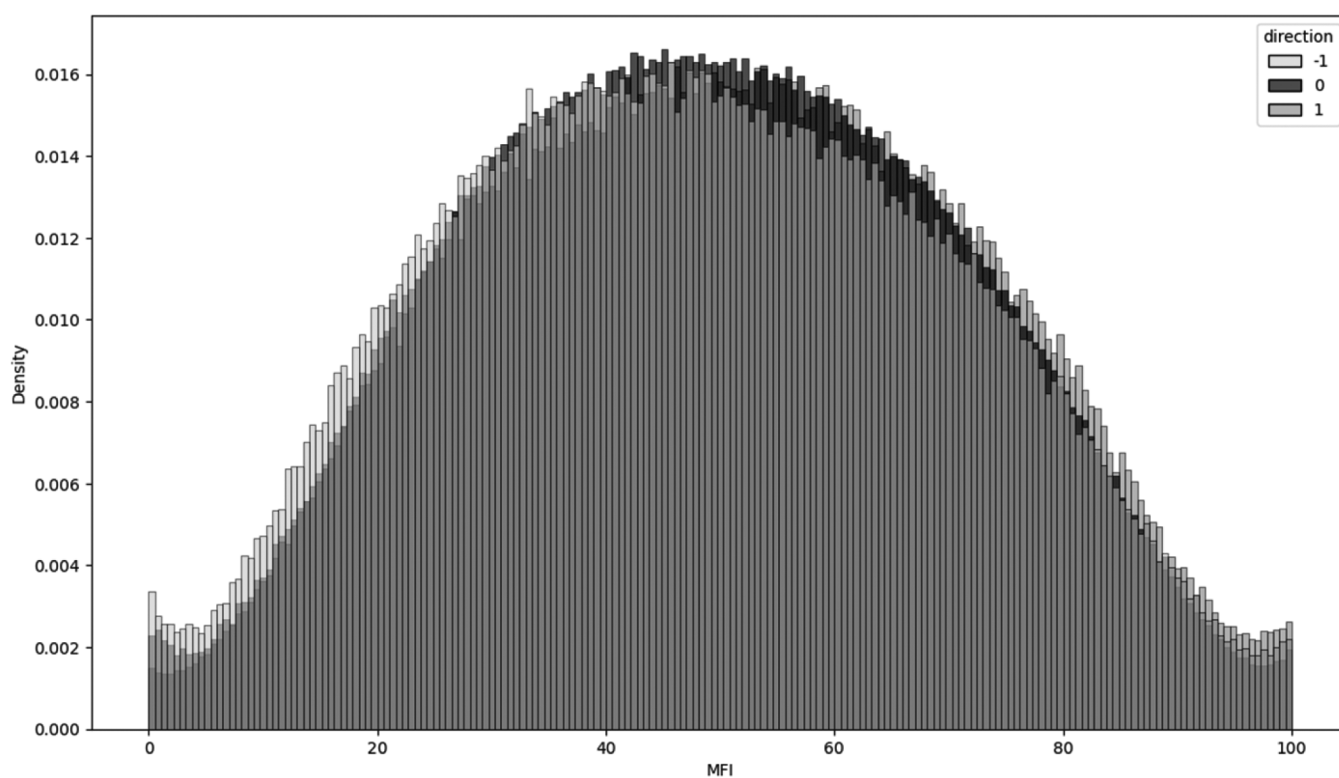


Рис. 5. Гистограмма индикатора MFI

- Индикатор TRIX (Triple Exponential Average) — это технический индикатор, используемый в анализе финансовых рынков для определения трендов и потенциальных точек разворота цены [8]. Гистограмма изображена на рисунке 4.
- Индикатор MFI (Money Flow Index, Индекс денежного потока) — это технический индикатор, который измеряет силу притока и оттока денег в актив, основываясь на изменениях цены и объема торгов. MFI учитывает не только цену, но и объем [9]. Гистограмма изображена на рисунке 5.
- Ultimate Oscillator — индикатор, который комбинирует данные за короткий, средний и длинный периоды — это позволяет сгладить ложные сигналы, которые могут возникать при использовании одного таймфрейма. Индикатор учитывает три ключевых компонента: покупательное давление (Buying Pressure), истинный диапазон (True Range) и взвешенные средние этих значений [10]. Гистограмма изображена на рисунке 6.
- Индекс товарного канала (CCI) — технический индикатор, основанный на анализе текущего изменения отклонения цены от её среднего значения за определённый период и среднестатистического абсолютного значения этого параметра [11]. Гистограмма изображена на рисунке 7.

Обучение модели

Для оптимизации модели были выбраны следующие диапазоны гиперпараметров:

- Скорость обучения (learning rate) — от 0.01 до 0.3
- Максимальная глубина дерева (depth) — от 1 до 10
- Штраф от L2-регуляризации (l2 leaf reg) — от 0.0001 до 10
- Количество границ (border count) — от 1 до 255
- Уровень случайности при формировании деревьев (bagging temperature) — от 0 до 1

Обучение производилось с помощью кросс-валидации на 5 различных наборах данных, итоговое значение MAPE = 0.029, значения гиперпараметров для лучшей модели перечислены в таблице 1.

Таблица 1.

Значения гиперпараметров для лучшей модели

Гиперпараметр	Значение
Скорость обучения	0.155
Максимальная глубина дерева	9
Штраф от L2-регуляризации	6.665
Количество границ	75
Уровень случайности при формировании деревьев	0.909

Заключение

В данной работе была предложена реализация градиентного бустинга в качестве метода прогнозирования

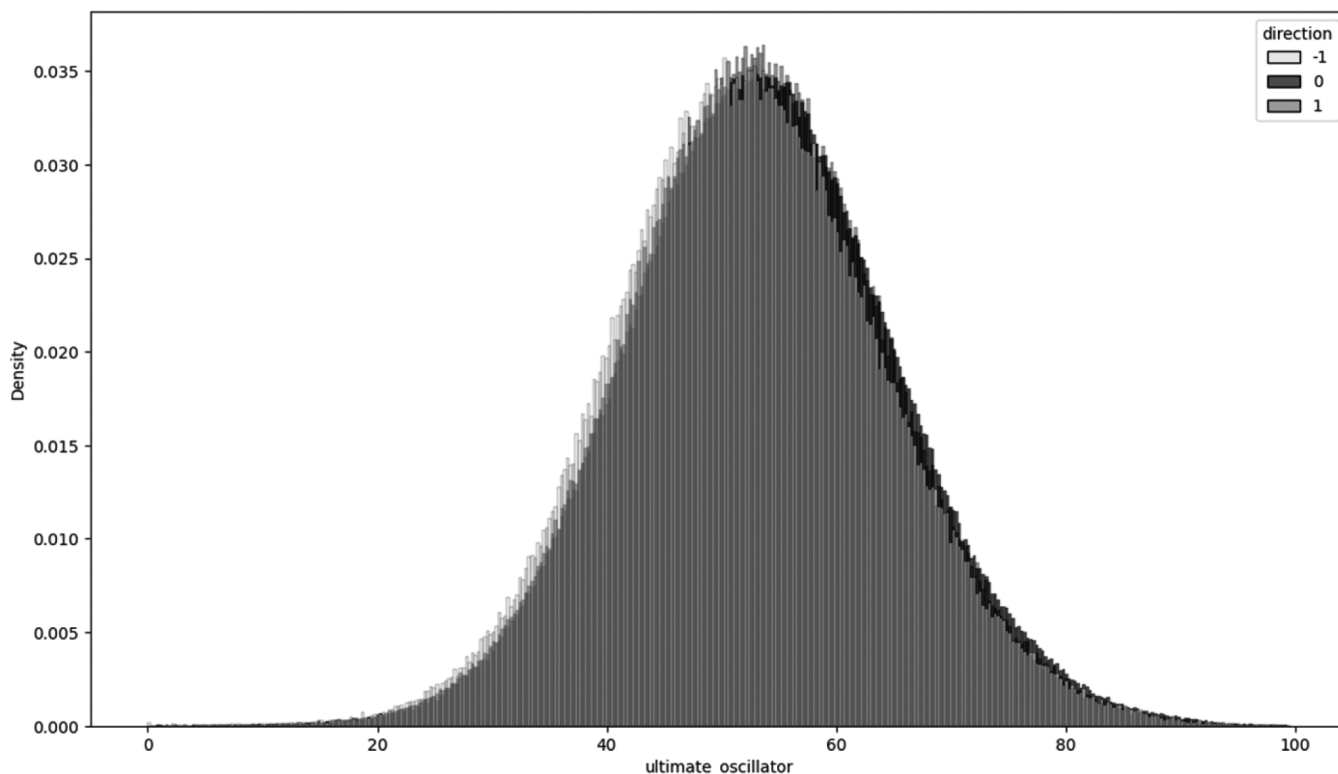


Рис. 6. Гистограмма индикатора ultimate oscillator

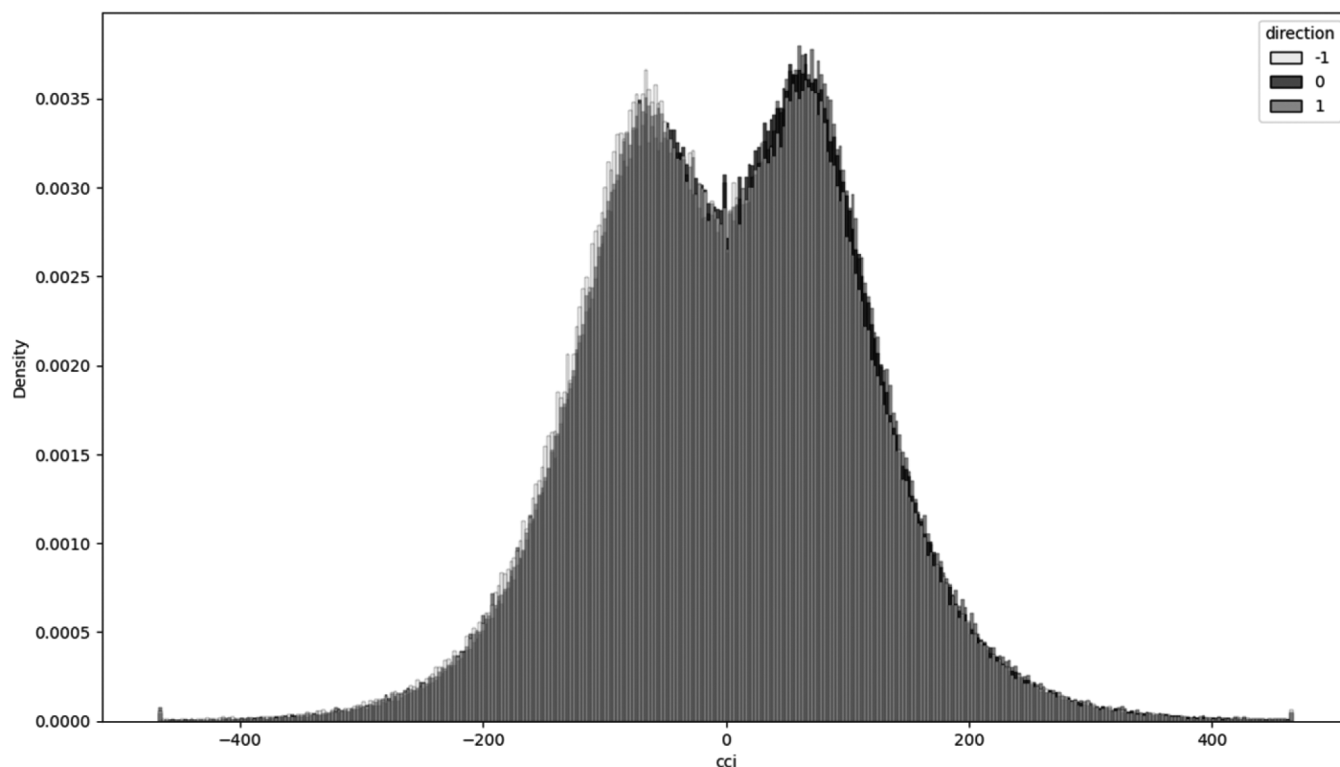


Рис. 7. Гистограмма индикатора csi

стоимости акций на основе признаков, имеющих влияние на итоговую цену, с помощью кросс-валидации была отобрана лучшая модель со значением MAPE = 0.029,

данная точность позволяет использовать эту модель в качестве системы управления инвестиционным портфелем.

ЛИТЕРАТУРА

1. Christopher M. Bishop Pattern Recognition and Machine Learning. — 1е изд. — Singapore: Springer Science+Business Media, LLC, 2008. — 760 с.
2. Исследование методов прогнозирования стоимости ценных бумаг на основе машинного обучения // Российский университет дружбы народов URL: <https://events.rudn.ru/event/107/papers/504/files/857-Kravets.pdf> (дата обращения: 03.07.2025).
3. Trevor Hastie, Robert Tibshirani, Jerome Friedman The Elements of Statistical Learning. — 1-е изд. — Спб.: «Санкт-Петербургский Центр эволюционных исследований сознания человека» (внутренний путь), 2009. — 764 с.
4. Павлов В.А., Солобуто А.В. Предобработка данных в задаче прогнозирования динамики ценных бумаг // Современные и информационные технологии в социальной сфере. — Чебоксары: Волжский филиал МАДИ, 2024. — С. 42–45.
5. Стивен Б. Акелис Technical Analysis from A to Z. — 2 изд. McGraw Hill, 2013. — 400 с.
6. J. Welles Wilder, Jr. New Concepts in Technical Trading Systems — М.: ПРИЗМА-15, pp 40 — 41
7. Колби Роберт Энциклопедия технических индикаторов рынка. — 2 изд. — М.: Альпина Паблишер, 2004. — 837 с.
8. ЛеБо Ч., Лукас Д.В Компьютерный анализ фьючерсных рынков. — 6 изд. — М.: Альпина Паблишер, 2011. — 304 с.
9. Кауфман Перри Системы и методы биржевой торговли. — 1-е изд. — М.: Альпина PRO, 2024. — 1248 с.
10. Донна Л. МакКормик, Джеффри Оуэн Кац Энциклопедия торговых стратегий. — М.: Альпина Паблишер, 2012. — 394 с.
11. Larry Williams Trading Indicators // Larry Williams Trading Indicators URL: <https://williamspercentr.com/newsletters/ULTI.pdf> (дата обращения: 21.04.2025).

© Солобуто Алексей Викторович (29352138@s.mfua.ru); Павлов Валерий Анатольевич (29359332@s.mfua.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»