

# ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ

## ПРИМЕНЕНИЕ ГРАДИЕНТНОГО БУСТИНГА ДЛЯ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ

**И. В. Барышева**

*baryshevaiolanta@gmail.com;*

*Научный руководитель – К. В. Василевский, кандидат физико-математических наук,  
доцент*

Рассматривается метод градиентного бустинга для прогнозирования временного ряда, состоящего из цен на криптовалюту Ethereum. Выявляются достоинства и недостатки применения этого метода для решения данной задачи. Проводится сравнительный анализ результатов прогнозирования, полученных с помощью регрессора градиентного бустинга, и результатов прогнозирования модели линейной регрессии. В результате вычислительного эксперимента с обучением моделей градиентного бустинга и линейной регрессии были получены значения метрик, демонстрирующих качество полученных предсказаний.

**Ключевые слова:** прогнозирование временных рядов; машинное обучение; градиентный бустинг; линейная регрессия; язык программирования Python.

### **ВВЕДЕНИЕ**

Прогнозирование временных рядов – это специфическая задача для машинного обучения, так как набор данных временных рядов принципиально отличается от типичной концепции обучающих данных для моделей машинного обучения. Время для такой задачи представляет собой новое измерение, выступающее в роли дополнительной информации для обучения. Поэтому, имея данные, демонстрирующие прошлые и настоящие тенденции, мы получаем возможность предсказывать и будущие тенденции. Таким образом, алгоритмы, использующиеся для предсказания временных рядов, должны обладать важным свойством экстраполяции шаблонов за пределы домена обучающих данных.

В данной работе предлагается изучить новаторский подход к прогнозированию временных рядов – градиентный бустинг. Градиентный бустинг – это продвинутый алгоритм машинного обучения, хорошо зарекомендовавший себя для решения задач классификации и регрессии, однако для задачи прогнозирования временных рядов этот алгоритм стали использовать лишь недавно. Данный метод строит ансамбль из слабых предсказывающих моделей – решающих деревьев. Общая идея алгоритма

заключается в том, что каждая последующая модель стремится свести ошибку предыдущей к минимуму.

Результаты прогнозирования, полученные моделью градиентного бустинга, будут сравниваться с результатами, полученными с помощью линейного регрессора. Для данной работы использовались регрессор градиентного бустинга из библиотеки XGBoost и линейный регрессор из библиотеки Sklearn.

## ОПИСАНИЕ ДАННЫХ ДЛЯ ОБУЧЕНИЯ

Задача заключается в том, чтобы предсказать временной ряд цен на криптовалюту Ethereum по курсу ETH/USDT, поэтому в этой работе использовались данные с финансовой платформы Binance за 5 лет с однодневным интервалом. Платформа Binance предоставляет финансовые данные в виде OHLCV списка данных, который, в свою очередь, представляет собой список из пяти наиболее важных для финансового анализа типов информации – цену открытия, максимум движения цены, минимум движения цены, цену закрытия и объем торгов за выбранный период.

В данной работе предсказывались значения цены закрытия, а в качестве обучающей выборки данных использовались финансовые индикаторы КАМА (адаптивная скользящая средняя Кауфмана) и VIDYA (скользящая средняя с динамическим периодом усреднения) для цен открытия, максимума движения цен, минимума движения цен и объема торгов.

## РЕЗУЛЬТАТЫ

В качестве функции потерь при обучении моделей использовалась метрика MSE (среднеквадратичная ошибка). Такие параметры, как скорость обучения, глубина деревьев и число деревьев, для регрессора градиентного бустинга подбирались автоматически с помощью инструмента GridSearchCV.

Результаты прогнозирования на тестовой выборке представлены в виде следующего графика.

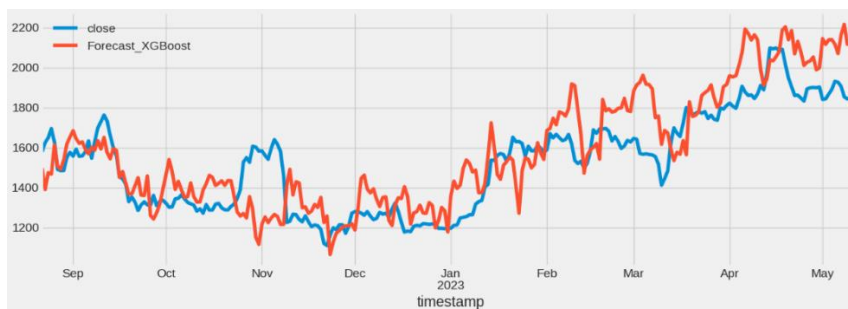


Рис. 1. Результаты прогнозирования алгоритма градиентного бустинга

Таким же образом для дальнейшего сравнительного анализа представлены и результаты прогнозирования модели линейной регрессии.

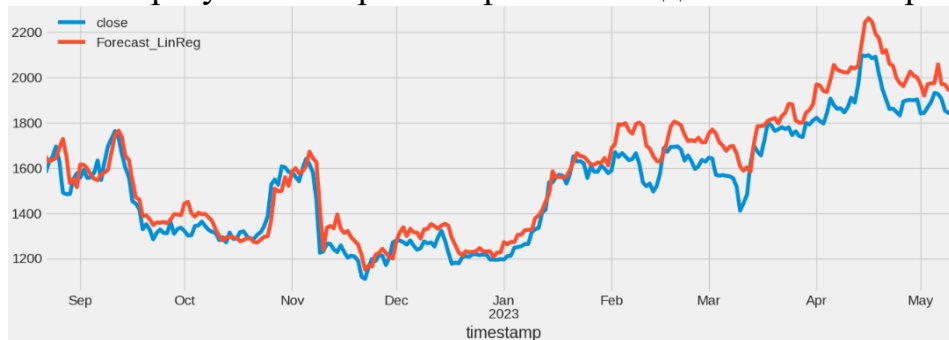


Рис. 2. Результаты прогнозирования модели линейной регрессии

Для оценки качества полученных используемыми моделями результатов использовались ошибки RMSE (корень из среднеквадратичной ошибки) и MAE (средняя абсолютная ошибка).

**Значения ошибок, полученных регрессором градиентного бустинга XGBoost и линейным регрессором**

Ошибка	Регрессор градиентного бустинга XGBoost	Линейный регрессор
RMSE	168,18	90,83
MAE	134,19	74,14

**ЗАКЛЮЧЕНИЕ**

Судя по результатам, наилучшие предсказания временного ряда были получены линейной регрессией. Проведя анализ используемых этими моделями алгоритмов, в этой работе были сделаны выводы, что для решения данной задачи прогнозирования более целесообразным является использование модели линейной регрессии, нежели модели из деревьев решений.

Алгоритм градиентного бустинга хорошо выявляет сложные нелинейные зависимости между входными и выходными данными, а также отлично справляется с теми ситуациями, когда необходимо интерполировать в домене уже известных данных. Однако данная задача временного прогнозирования – совершенно не такой случай. Зависимость между входными и выходными данными в той постановке задачи, которая представлена в этой работе, скорее линейная, чем нелинейная, и, как было отмечено ранее, для алгоритмов, используемых в прогнозировании временных рядов, важно свойство экстраполяции за пределы домена обучающей выборки, а алгоритм градиентного бустинга данным свойством не обладает, следовательно, его не стоит применять в решении

задач прогнозирования нестационарных временных рядов. Именно поэтому в данной задаче алгоритм градиентного бустинга потерпел неудачу.

### **Библиографические ссылки**

1. *Николенко С., Кадурич А., Архангельская Е.* Глубокое обучение. СПб.: Питер, 2020. 480 с.: ил. - (Серия «Библиотека программиста»).
2. *Шакла Нишант.* Машинное обучение и TensorFlow. - СПб.: Питер, 2019. 336 с.: ил. (Серия «Библиотека программиста»).
3. *Нильсен Э.* Практический анализ временных рядов: прогнозирование со статистикой и машинное обучение / Э. Нильсен. – Москва: Вильямс, 2021. 544 с.
4. *Афанасьев В.Н.* Анализ временных рядов и прогнозирование / В.Н. Афанасьев, М.М. Юзбашев. Москва: Финансы и статистика, 2001. 228 с.