**МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ**  
**УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**   
«**РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА», РУТ (МИИТ)**

АКАДЕМИЯ «ВЫСШАЯ ИНЖЕНЕРНАЯ ШКОЛА»

**КУРСОВАЯ РАБОТА**

по курсу

«Нейронные сети»

на тему

«Прогнозирование финансовых временных рядов»

Выполнил(и):

студент(ы) 4 курса ВИШ РУТ МИИТ

Орлов В.М., Васильев Д.E.

Руководитель:

Н. С. Мартыненко

Cдана на проверку:

19 декабря 2024

Дата защиты и оценка:

19 декабря 2024

## Введение

Финансовые рынки характеризуются высокой динамичностью и сложностью, что делает прогнозирование ценовых движений важной задачей для инвесторов, трейдеров и аналитиков. Прогнозирование временных рядов позволяет предсказывать будущие значения финансовых показателей на основе исторических данных, что способствует принятию обоснованных инвестиционных решений и управлению рисками. В данной работе рассматривается задача прогнозирования финансовых временных рядов с использованием различных методов машинного обучения и статистических моделей. Основной целью является оценка эффективности различных подходов и выбор наилучшей модели для предсказания цен закрытия активов.

## Постановка задачи и идея решения

Задача прогнозирования финансовых временных рядов заключается в предсказании будущих значений финансовых показателей, таких как цена закрытия активов, на основе исторических данных. В данной работе рассматривается прогнозирование цены закрытия активов с использованием ряда моделей, включая традиционные статистические методы (ARIMA, SARIMA, PROPHET), методы машинного обучения (Random Forest, SVM, CatBoost, XGBoost) и нейронные сети (GRU, LSTM, TCN).

Идея решения состоит в сравнении различных моделей по их способности предсказывать временные ряды, оценке их точности с помощью метрик средней квадратичной ошибки (MSE) и средней абсолютной процентной ошибки (MAPE), а также анализе затрат ресурсов (память и время). Это позволит выявить наиболее эффективные подходы для задач финансового прогнозирования.

## Данные

Для исследования использовались финансовые временные ряды, содержащие следующие характеристики:

* Время начала сделки (startTime)
* Цена открытия (openPrice)
* Максимальная цена (highPrice)
* Минимальная цена (lowPrice)
* Цена закрытия (closePrice)
* Объём торгов и другие параметры

Источник данных: Bybit API

Предобработка данных включала очистку от пропусков, удаление выбросов и нормализацию значений для улучшения качества моделей. Данные охватывали период в четыре года, из которых три года использовались для обучения моделей, а последний год — для тестирования. Таким образом, данные были разделены на обучающую, валидационную и тестовую выборки в соотношении 75:15:10. Такой подход обеспечивает достаточное количество данных для обучения моделей и объективную оценку их качества на невидимых данных.

## Модель

В качестве baseline-метода была выбрана модель Support Vector Machine (SVM). Выбор SVM в качестве базовой модели обусловлен её широким применением в задачах регрессии и классификации, а также способностью работать с нелинейными зависимостями благодаря использованию различных ядерных функций. Несмотря на свою популярность и теоретическую эффективность, SVM показала наименее эффективные результаты по MSE среди рассмотренных моделей (MSE = 12 240 194.88), что делает её подходящей для сравнения с более продвинутыми методами.

Для решения задачи прогнозирования были выбраны следующие модели:

1. **ARIMA** (MSE = 189 411 971.75): традиционная статистическая модель, хорошо подходящая для стационарных временных рядов.
2. **SARIMA** (MSE = 262 001 812.72): расширение ARIMA, учитывающее сезонные компоненты.
3. **PROPHET** (MSE = 442 307 079.52): модель от Facebook, предназначенная для прогнозирования временных рядов с сезонными эффектами.
4. **Random Forest** (MSE = 7 342 773.34): ансамблевый метод, способный улавливать нелинейные зависимости.
5. **CatBoost** (MSE = 5 811 496.96): градиентный бустинг, хорошо работающий с категориальными данными.
6. **XGBoost** (MSE = 6 577 975.71): эффективный алгоритм градиентного бустинга с регуляризацией.
7. **GRU** (MSE = 8 705 696.67): рекуррентная нейронная сеть с затвором обновления.
8. **LSTM** (MSE = 8 687 840.55): продвинутая рекуррентная нейронная сеть, способная учитывать долгосрочные зависимости.
9. **TCN** (MSE = 8 163 108.62): Temporal Convolutional Network, использующая сверточные слои для обработки временных данных.

Для оценки качества моделей использовались метрики средней квадратичной ошибки (MSE) и средней абсолютной процентной ошибки (MAPE). MSE хорошо отражает рассеяние предсказанных значений относительно истинных, что важно для финансовых приложений, где даже небольшие ошибки могут иметь значительные последствия. MAPE позволяет оценить относительную точность моделей, выраженную в процентах, что упрощает интерпретацию результатов.

Основная модель для детального анализа — CatBoost, выбранная из-за её способности эффективно обрабатывать категориальные данные и автоматически справляться с переобучением. В процессе оптимизации были настроены гиперпараметры, такие как количество деревьев, глубина деревьев и скорость обучения, с использованием кросс-валидации.

## Результаты

|  |  |
| --- | --- |
| Модель | MSE (ошибка среднего квадрата) |
| ARIMA | 189411971.75 |
| CATBOOST | 5811496.96 |
| GRU | 8705696.67 |
| LSTM | 8687840.55 |
| PROPHET | 442307079.52 |
| RANDOM\_FOREST | 7342773.34 |
| SARIMA | 262001812.72 |
| SVM | 12240194.88 |
| TCN | 8163108.62 |
| XGBoost | 6577975.71 |

|  |  |
| --- | --- |
| Модель | MAPE (относительная ошибка) % |
| ARIMA | 3988.71% |
| CATBOOST | 3.99% |
| GRU | 4.93% |
| LSTM | 4.74% |
| PROPHET | 916.81% |
| RANDOM\_FOREST | 3.69% |
| SARIMA | 4705.92% |
| SVM | 8.33% |
| TCN | 3.37% |
| XGBoost | 3.86% |

### Примеры работы модели

## Заключение

В ходе исследования было проведено сравнение различных моделей для прогнозирования финансовых временных рядов. На основе полученных результатов модели градиентного бустинга (CatBoost, XGBoost) показали наилучшую эффективность по метрикам средней квадратичной ошибки (MSE) и средней абсолютной процентной ошибки (MAPE), демонстрируя высокую точность предсказаний. В частности, модель CatBoost показала наименьшее значение MSE (5 811 496.96) и конкурентоспособный MAPE (3.99%), что свидетельствует о её высокой точности и надежности.

Традиционные статистические методы, такие как ARIMA, SARIMA и PROPHET, оказались значительно менее эффективными для данной задачи, что подтверждается их высокими значениями MSE и MAPE. Эти модели плохо подходят для прогнозирования финансовых временных рядов из-за их ограниченной способности улавливать сложные и нелинейные зависимости, характерные для высоковолатильных рынков. Особенно модели, учитывающие сезонные компоненты (SARIMA и PROPHET), не смогли адекватно справиться с нестандартными и резкими изменениями цен, что часто наблюдается на финансовых рынках.

Нейронные сети (GRU, LSTM, TCN) также показали хорошие результаты, уступая лишь моделям градиентного бустинга. Однако они требуют значительных вычислительных ресурсов и тщательной настройки гиперпараметров, что может усложнять их применение на практике.

Сравнение с baseline-моделью (SVM) подтвердило её наихудшую эффективность по обоим метрикам (MSE = 12 240 194.88, MAPE = 8.33%). Это указывает на необходимость использования более сложных и современных методов для задач прогнозирования финансовых временных рядов.

В целом, предложенные методы показали успешность в сравнении с baseline, подтвердив гипотезу о наличии значимой связи между входными и выходными данными для задачи прогнозирования финансовых временных рядов. Наиболее эффективными оказались модели градиентного бустинга, способные улавливать сложные зависимости и обеспечивать высокую точность предсказаний при относительно низких вычислительных затратах.

## Приложение

Рабочая ссылка на каталог с данными и кодом: https://github.com/K0RTEK/financial\_time\_series