

Интенсив №3

Прогнозирование рыночных
цен на арматуру



самолет

Работу выполнили:

01

Костюк Алексей

02

Грушковский Захар

03

Матвеева Алина



самолет

Постановка задач



самолет

Цель

Помочь категорийному менеджеру, закупающему арматуру, сэкономить. В репозитории файлах `train` и `test` есть понедельная история рыночных цен на арматуру в промежутке 2015-2023 гг. Также в репозитории представлены дополнительные данные, которые можно использовать для прогнозирования цены. Любые другие внешние данные использовать запрещено.



Этапы работы

Предобработка
данных



Выбор наиболее
точной модели



Выбор подходящих
библиотек



Разработка
интерфейса



самолет

В ходе работы, было принято решение по поиску оптимальной модели для нашего проекта. Для достижения поставленной цели, мы работали с разными библиотеками и моделями



Варианты решения



шамолет

LSTM (Long Short-Term Memory) — это тип рекуррентной нейронной сети (RNN), специально разработанный для работы с последовательностями и долгосрочными зависимостями.



Плюсы:

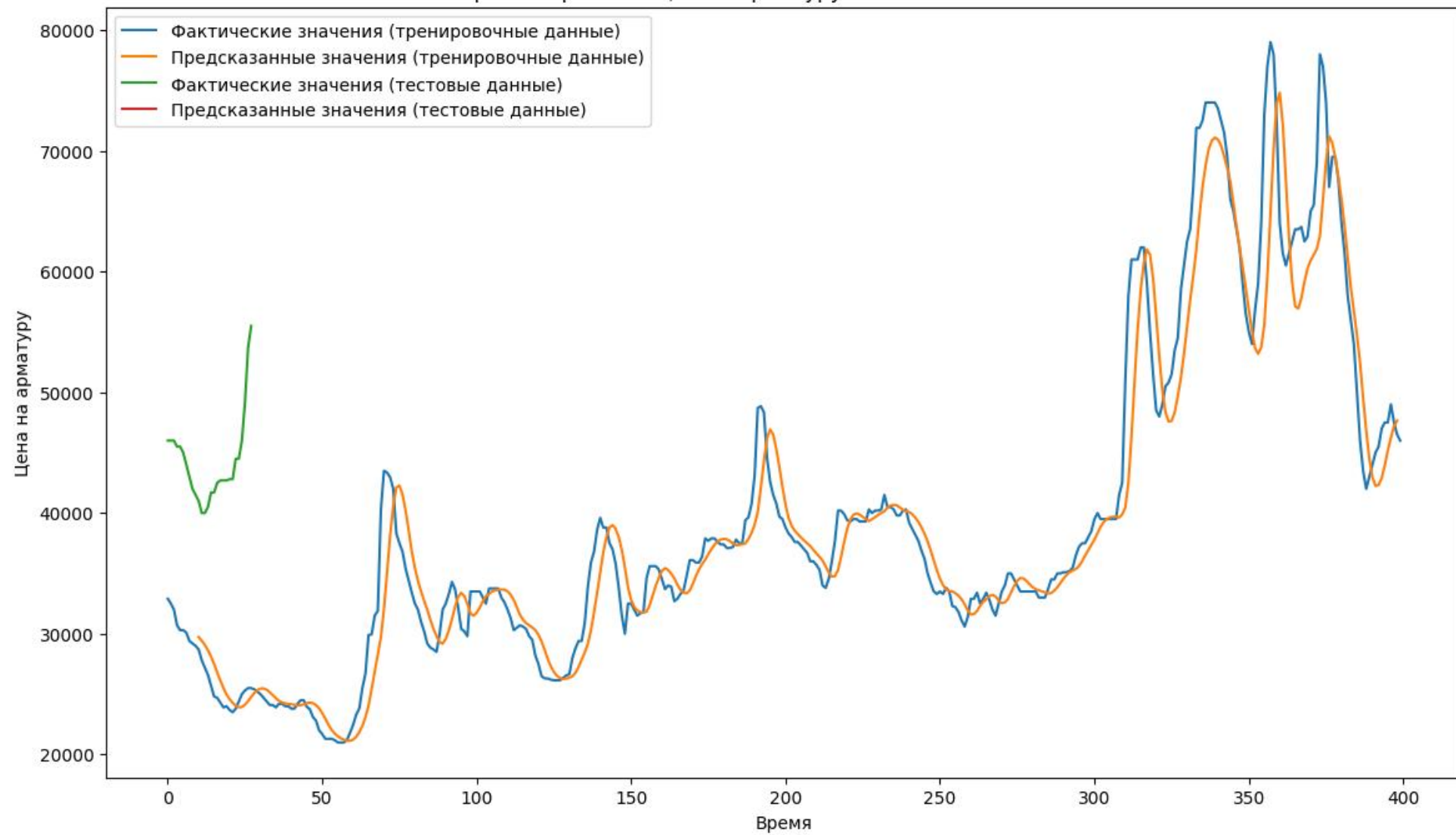
- Умеет работать с долгосрочными зависимостями
- Хорошо подходит для временных рядов и последовательностей
- Гибкость в архитектуре
- Работает с разномасштабными данными
- Устойчивость к шуму

Минусы:

- Высокая вычислительная сложность
- Требуется большого объема данных
- Сложность настройки гиперпараметров
- Интерпретируемость
- Чувствительность к препроцессингу



Прогнозирование цен на арматуру с использованием LSTM



Random Forest Regression (RFR) — это ансамблевая модель, основанная на множестве решающих деревьев, которая применяется для задач регрессии (прогнозирования числовых значений). Она сочетает в себе простоту интерпретации и высокую устойчивость к переобучению.



Плюсы:

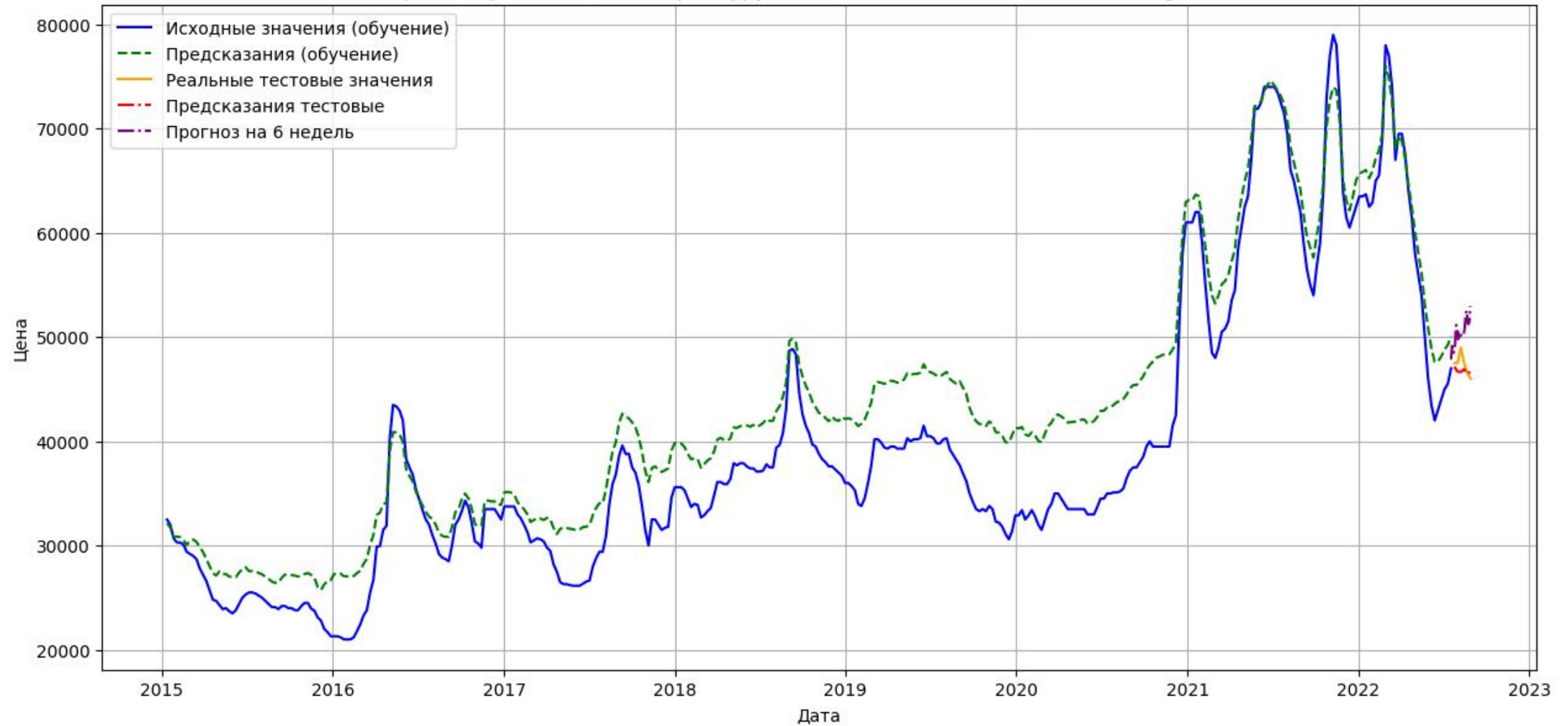
- Устойчивость к переобучению
- Работает с разнотипными данными
- Автоматический отбор важных признаков
- Хорошая точность даже без тонкой настройки
- Интерпретируемость (частичная)
- Поддержка нелинейных зависимостей

Минусы:

- Менее эффективен на экстраполяции
- Требуется больше вычислительных ресурсов
- Случайность может влиять на результат
- Не идеален для временных рядов
- Не дает гладких предсказаний



Прогнозирование цен на арматуру с использованием Random Forest Regression



CatBoostRegressor

CatBoostRegressor — это мощный алгоритм машинного обучения, особенно эффективный для работы с категориальными данными и временными рядами.

Причины выбора:

- Автоматическая обработка категориальных признаков.
- Устойчивость к переобучению.
- Хорошая интерпретируемость (важность признаков).



Плюсы:

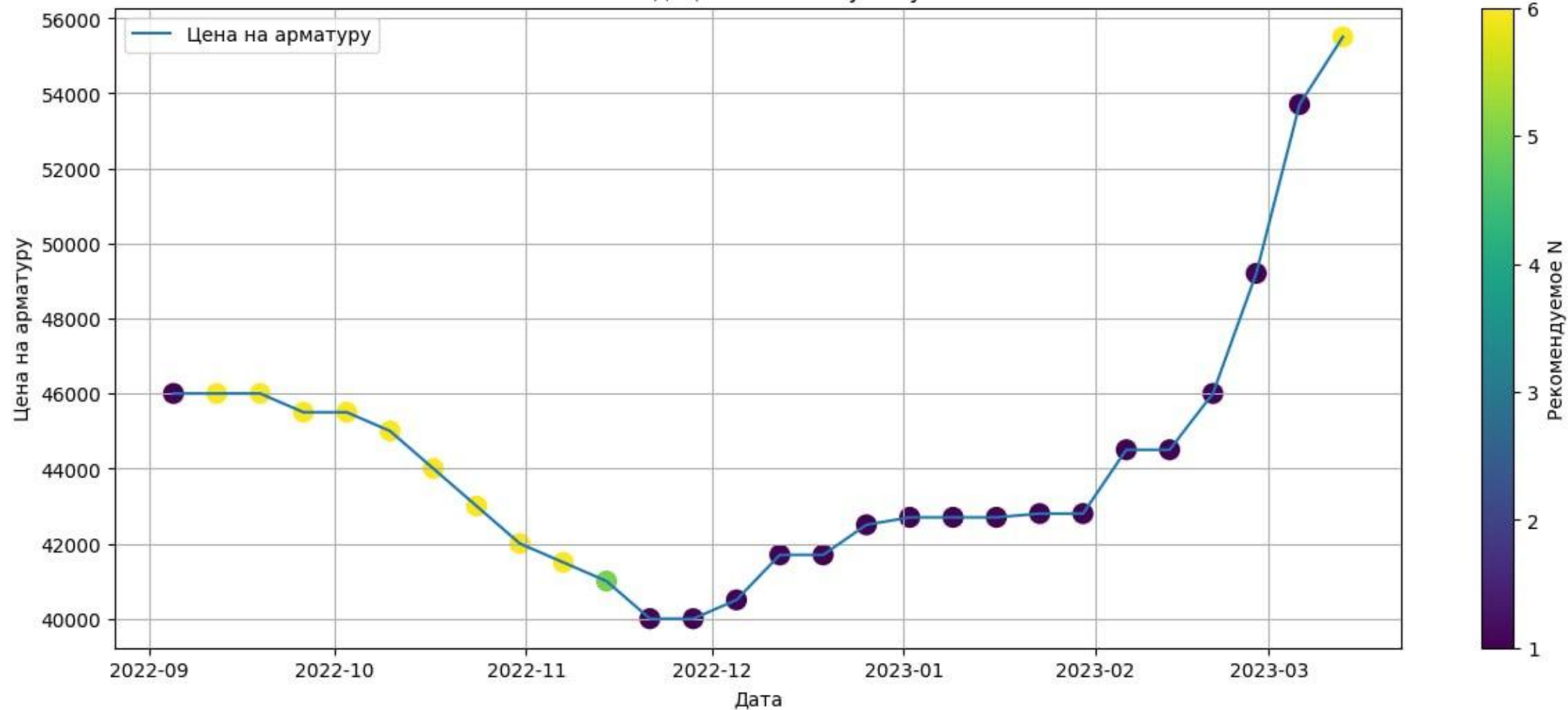
- Точность
- Автоматизация
- Интерпретируемость
- Устойчивость

Минусы:

- Вычислительная сложность
- Чувствительность к шумам
- Ограниченность временных окон



Рекомендации по объему закупки



самолет

Prophet — это библиотека прогнозирования временных рядов с открытым исходным кодом, разработанная командой Facebook.



Плюсы:

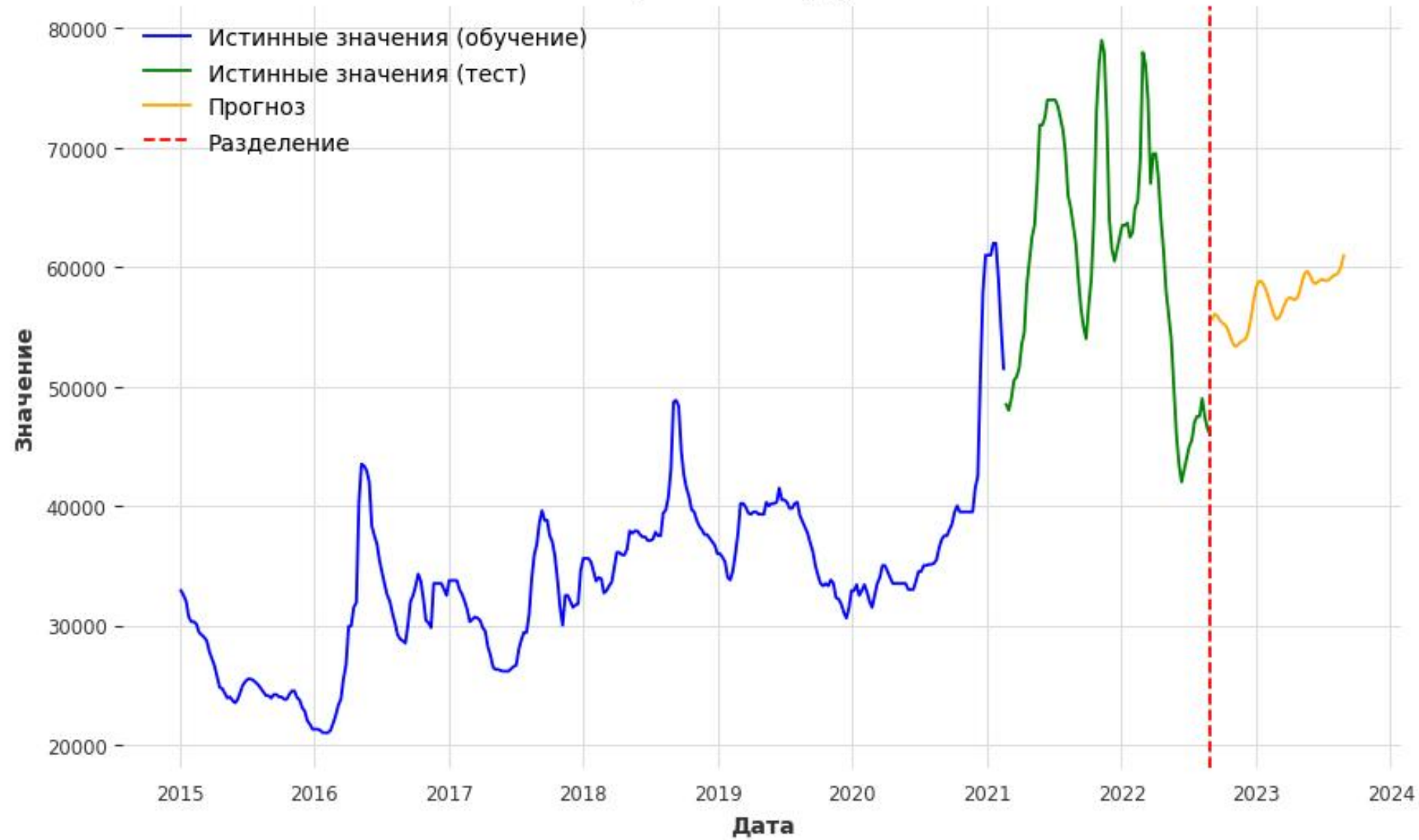
- Простота использования
- Учет сезонности и трендов
- Работа с выбросами и пропусками
- Поддержка пользовательских событий
- Интерпретируемость

Минусы:

- Ограниченная гибкость
- Проблемы с долгосрочными прогнозами
- Требуется однородных временных рядов
- Не поддерживает мультимодальные данные
- Вычислительные затраты на больших данных



Прогноз на будущее



самолет

Выбор модели

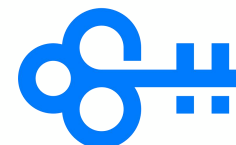
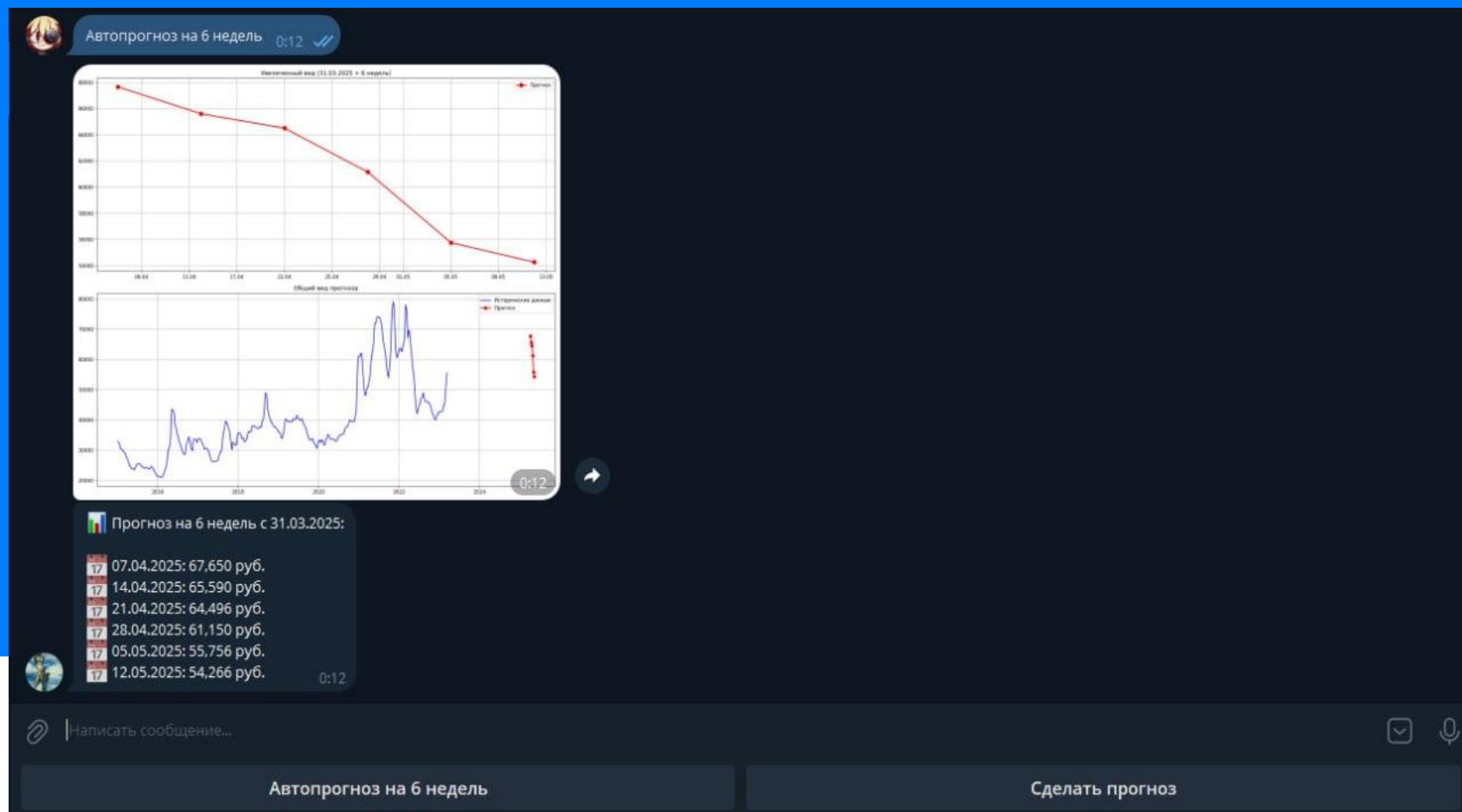


самолет

RandomForestRegressor

Наиболее лучшие результаты
нам показала модель
RandomForestRegressor из
библиотеки Scikit-Learn

Пример прогнозирования нашего приложения



самолет

