

Прогнозирование рыночных цен на арматуру



самолет

Работу выполнили студенты группы ИСП-22:

Кривобокова Ольга

Хусаинов Марат

Салимов Динислам

Преподаватель:

Коновалов Игорь Васильевич



самолет

О кейсе

Постановщик задачи:

Компания «Самолет»

Название кейсового задания:

Прогнозирование рыночных
цен на арматуру

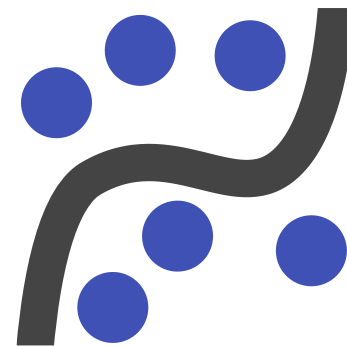
Цель:

Создание модели прогнозирования рыночных цен на арматуру для рекомендации лучшего времени для выгодной закупки арматуры.



Стек технологий

GO



statsmodels

Gin



Flask



самолет



Этапы работы

1

Предварительный
анализ данных

2

Подготовка данных

3

Построение моделей

каждый участник

4

Изучение
и сравнение
полученных решений

5

Объединение
лучших подходов в
одно решение

6

Обучение, настройка
и оптимизация
итоговой модели

7

Получение
и интерпретация
итогового
результата

8

Создание
презентации



самолет

Обработка и подготовка данных



Приведение данных к корректным форматам и типам



Удаление слабо заполненных данных



Слияние данных по датам



Заполнение пропусков методом backfill



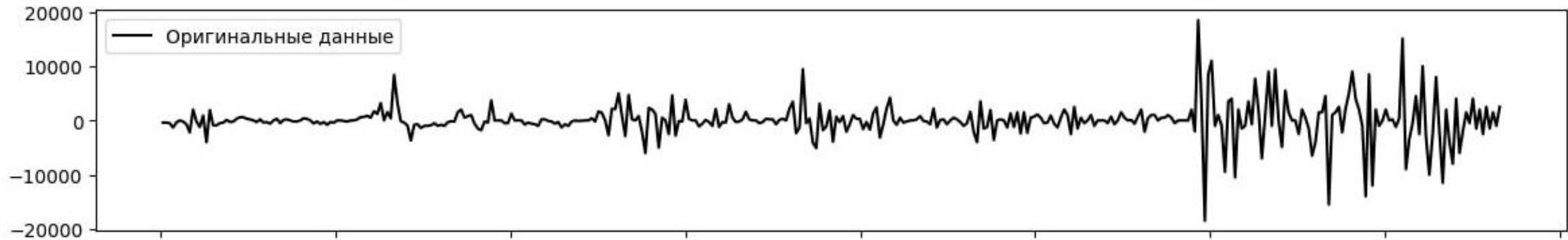
Демонстрация методов обработки показывает, что данные подготовлены
для дальнейшего анализа

Использование методов
`pd.merge` для объединения
DataFrame

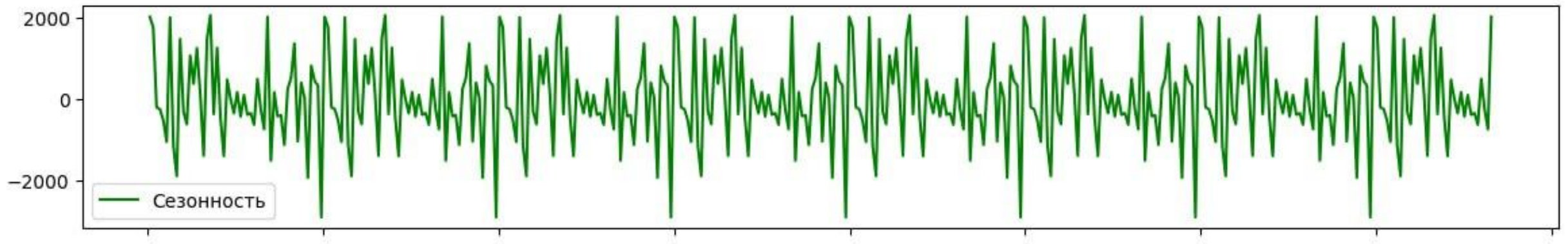
Приведение типов
временных данных с
помощью методов
`pd.to_datetime` для
обработки данных

Использование функции `seasonal_decompose` с периодом в 52 недели

ВЫСОКАЯ ВАЛАТИЛЬНОСТЬ

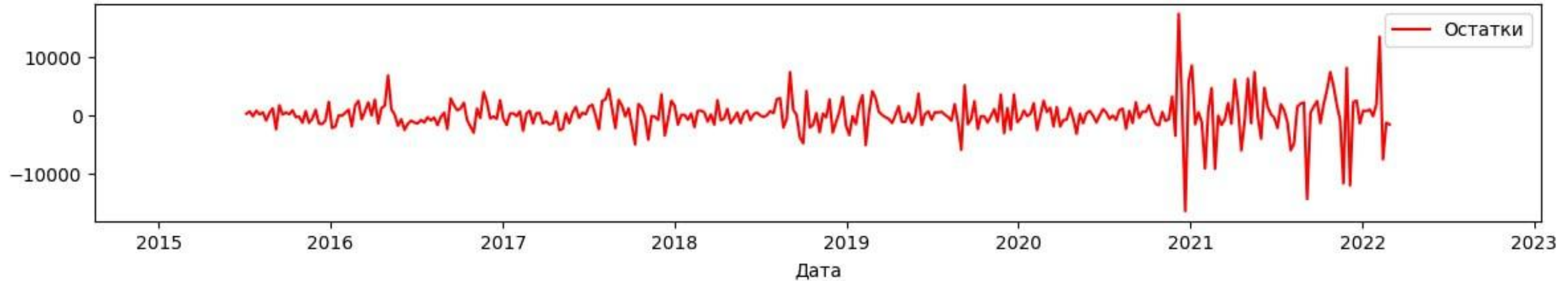


ВЫРАЖЕННАЯ ГОДОВАЯ СЕЗОННОСТЬ



самолет

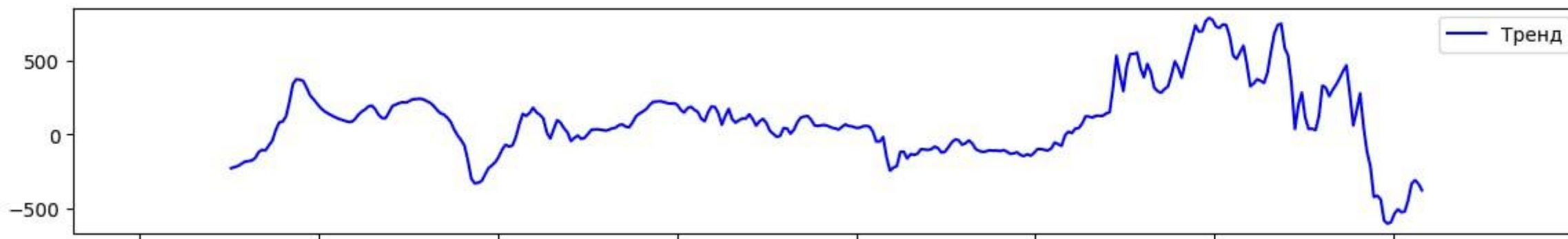
Вначале дисперсия остатков относительно стабильна, но начиная с 2020 года разброс резко увеличивается



Это может указывать на структурные изменения в данных
(например, экономические шоки, изменения в рыночных условиях)



самолет



Предположение, что тренд имеется, начали проверять на стационарность

Применяется **тест Дики-Фуллера**:

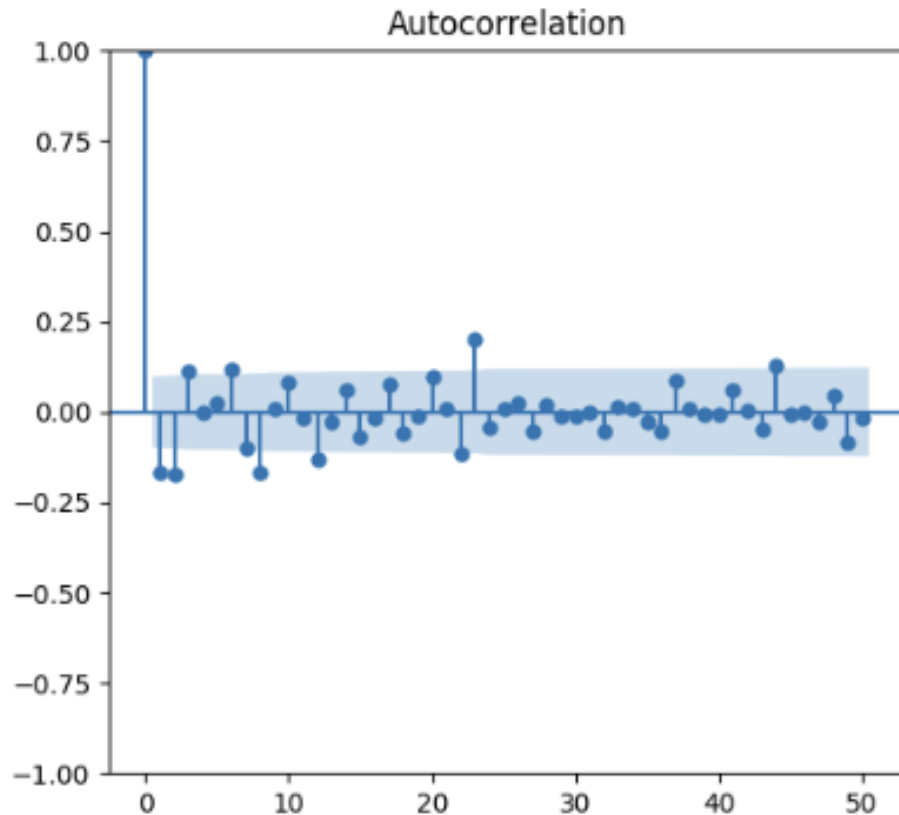
ADF Statistic: -1.5484468381220473

p-value: 0.5094239275287091

Ряд нестационарный, требуется
дифференцирование

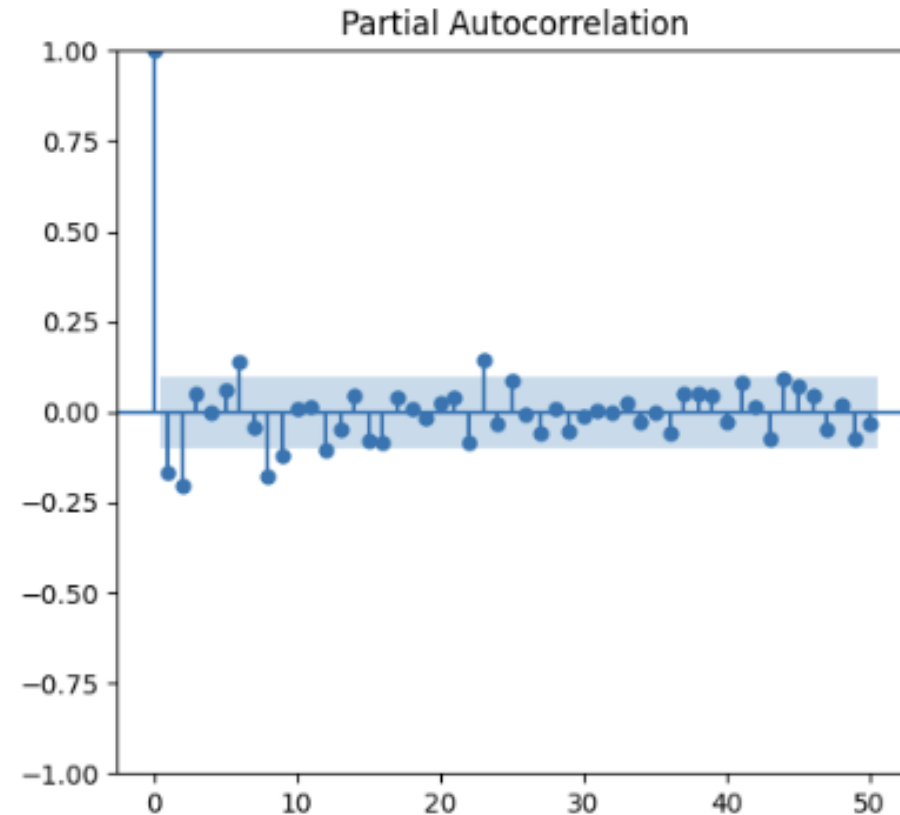
ACF:

- Медленно убывает (не резко обрывается)
- признак не стационарности ряда
- Высокие значения на первых лагах (сильная корреляция с соседними точками)



PACF:

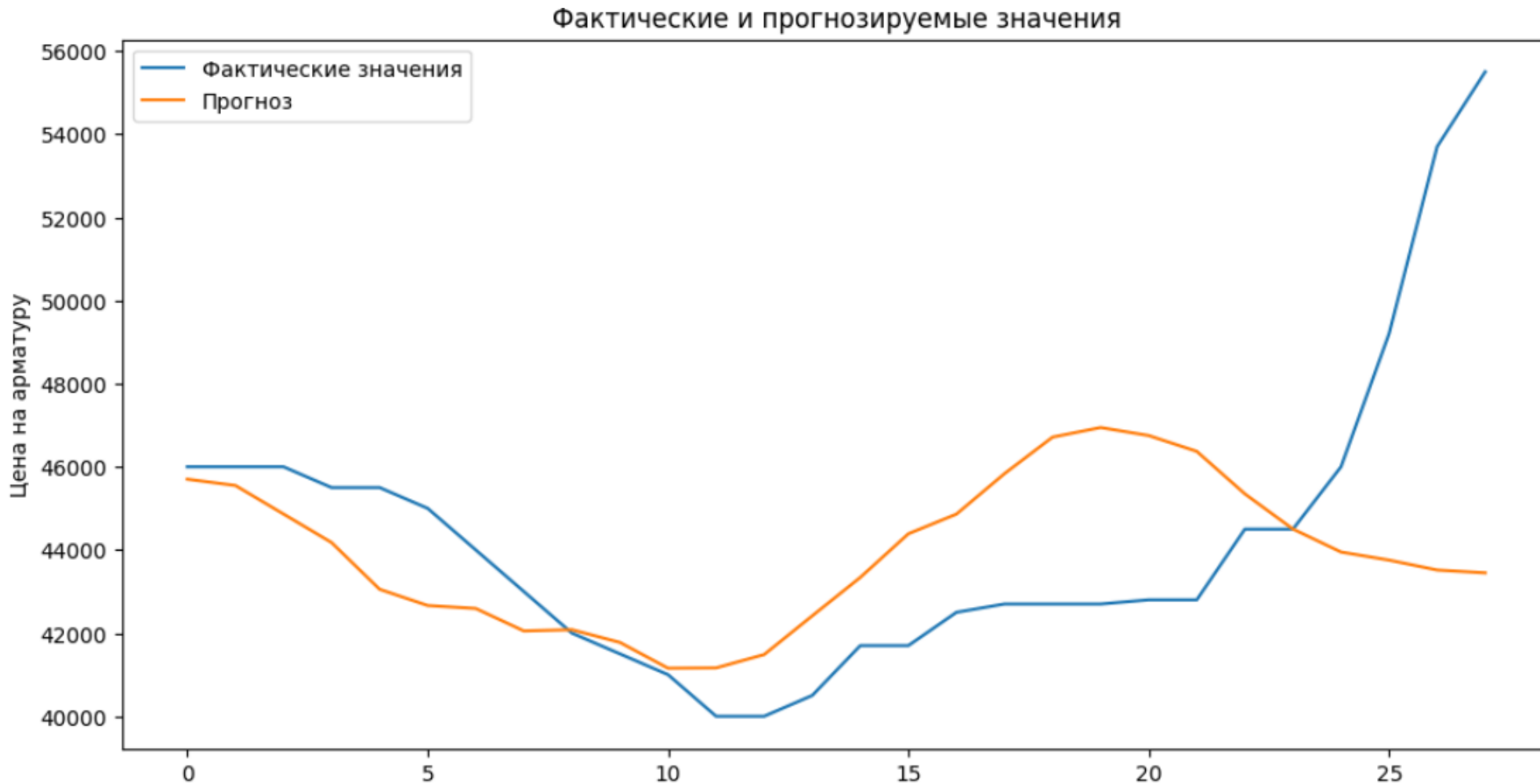
- Резко обрывается после 1-2 лага - это характерно для моделей AR (авто регрессионных)
- Значимый пик на лаге 1, затем быстро затухает



самолет

Модели

SARIMA без доп данных



самолет

Модели

RandomForestRegressor с доп данными

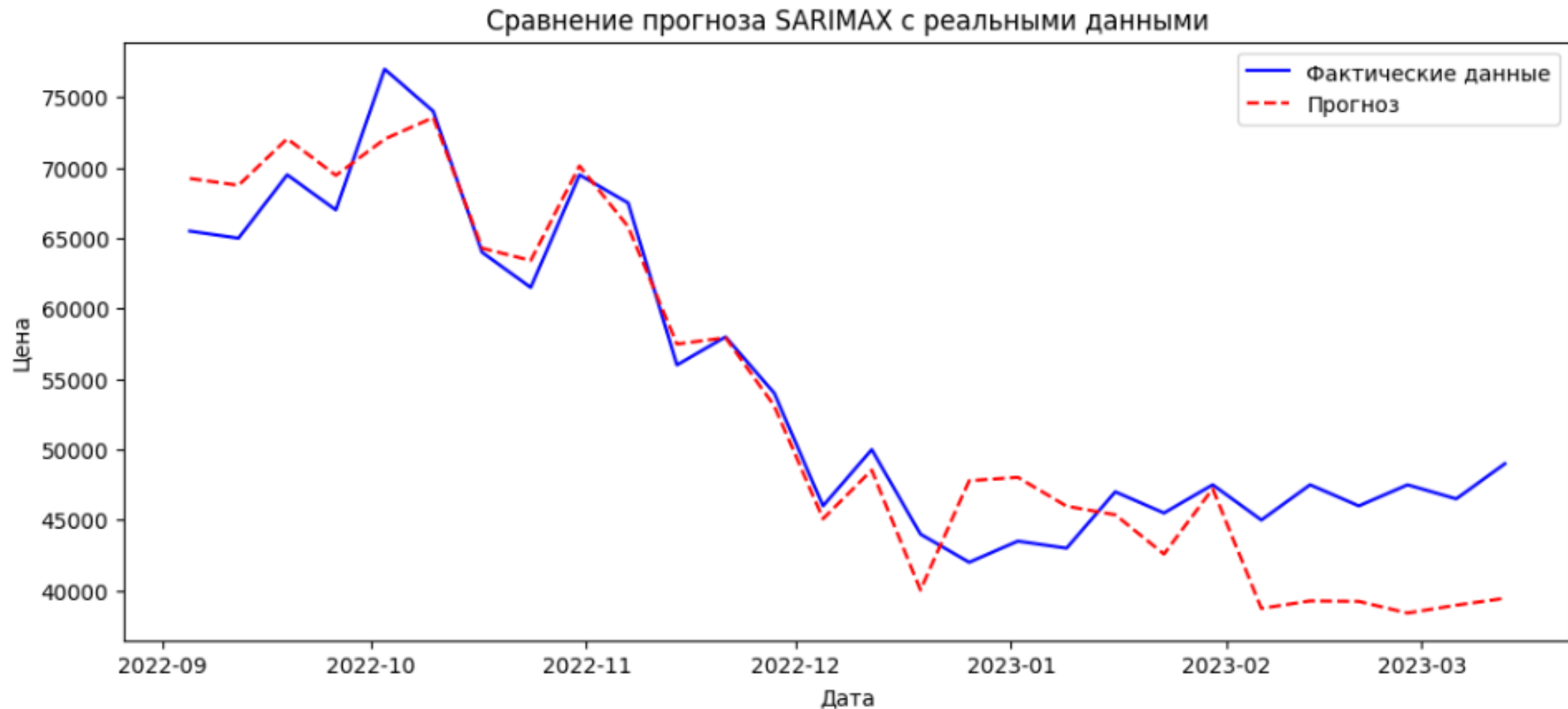
MAE: 4688.16
RMSE: 6073.10
MAE: 4065.52



самолет

Лучшая модель

SARIMAX с доп данными



самолет

Выбор модели и автоматический подбор параметров

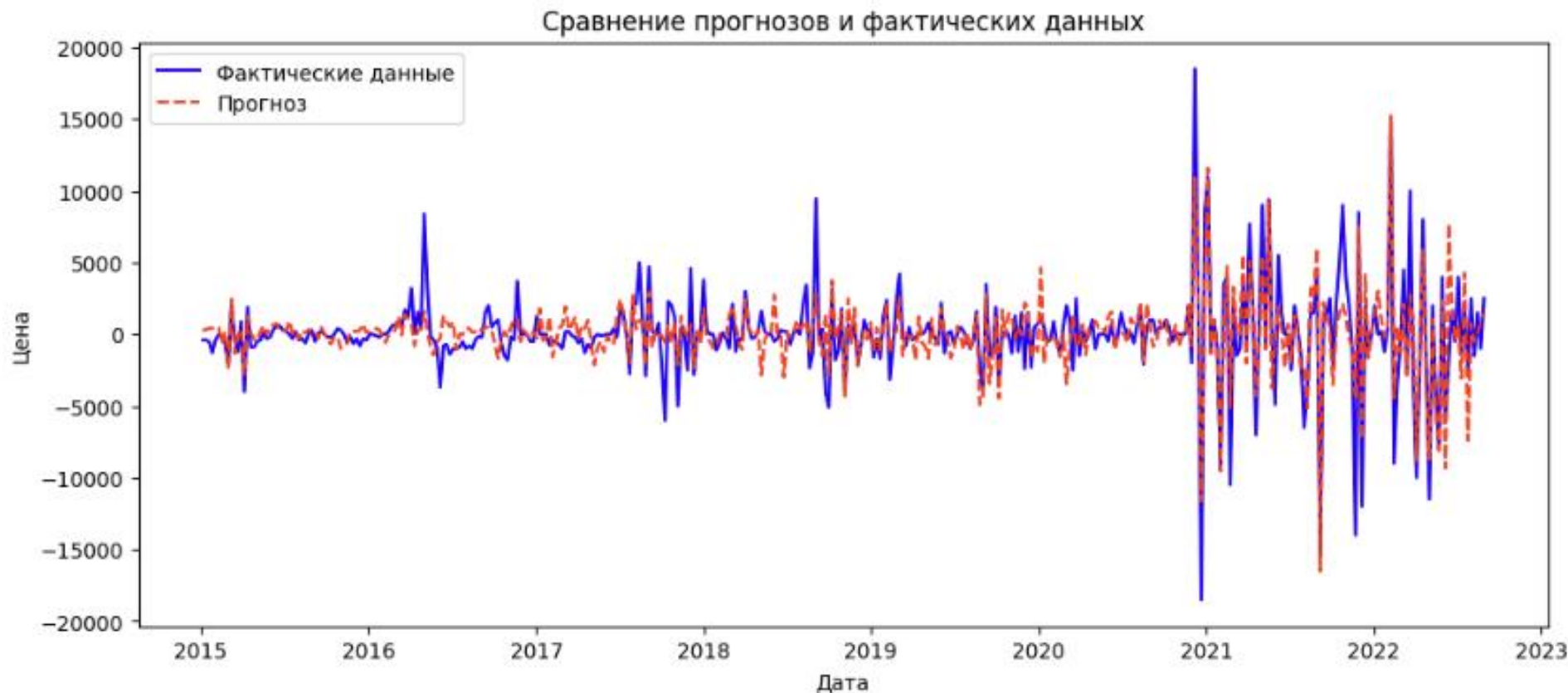
Выбранная модель: **ARIMA(2,0,0)(0,0,1)[52]**

- **(2,0,0)** – авторегрессия ($AR=2$), без дифференцирования ($d=0$), без МА-компонента ($MA=0$)

- **(0,0,1)[52]** – сезонная часть с $MA=1$ и периодом **52 недели** (годовой цикл)

Оценка качества модели

- $R^2 = 0.827$ → модель объясняет **82,7%** изменчивости цен
- **RMSE = 4442.37** → средняя ошибка ~**4442 руб**
- **MAE = 3459.11** → типичная ошибка ~**3459 руб**



самолет

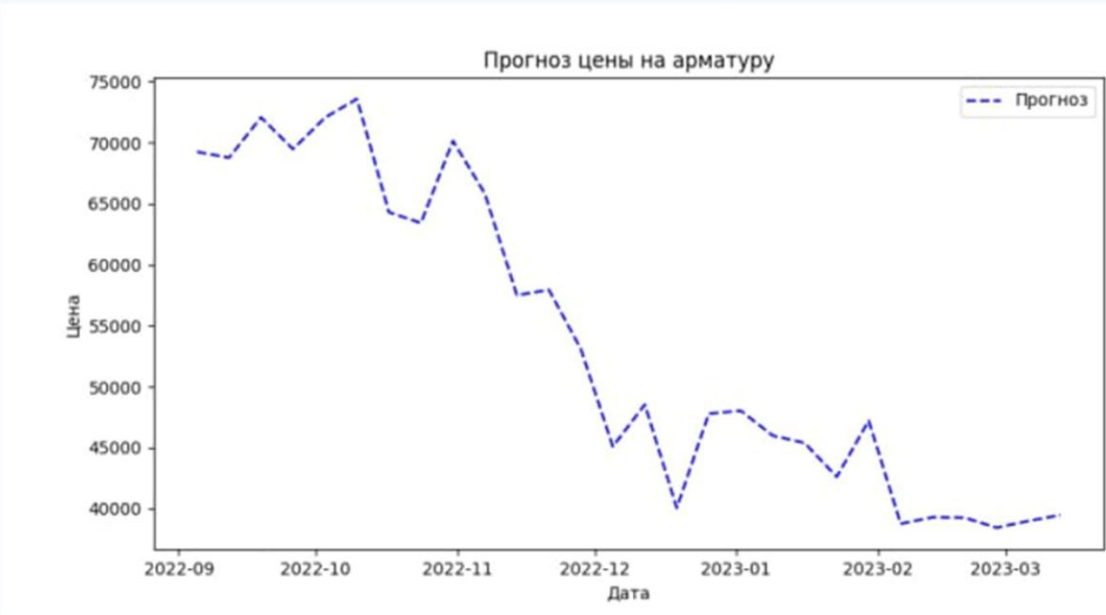
Интеграция:

Результаты предсказания



Изменение цены: Прогнозируется снижение цен. Закупайтесь на 1 неделю (X тонн), не стоит увеличивать объем.

График предсказаний:



Числовые данные:

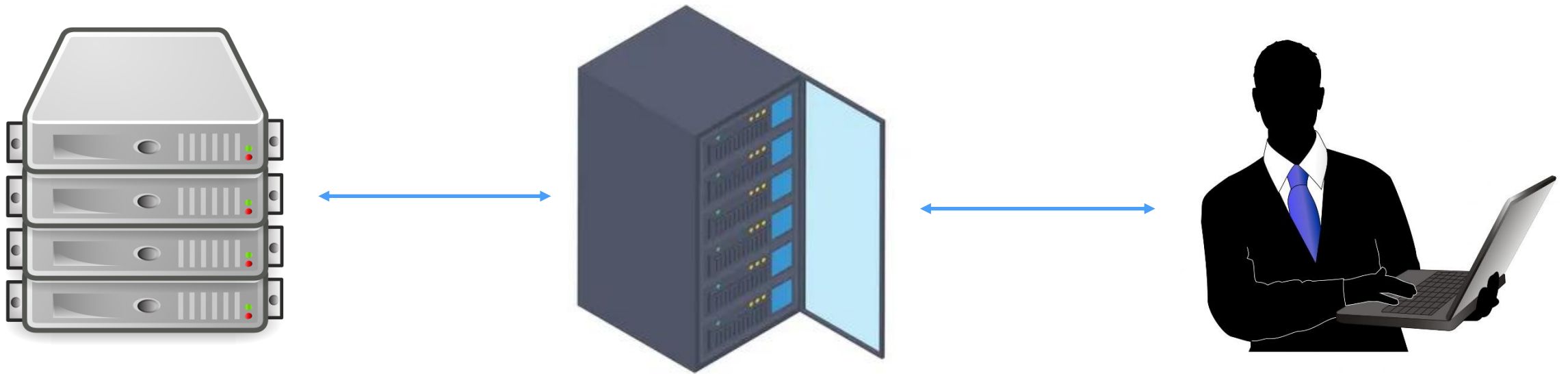
Дата	Предсказанная цена
2022-09-05	69235.16035513811
2022-09-12	68752.02222646668
2022-09-19	72060.33951657797
2022-09-26	69458.0976499989
2022-10-03	72033.44280487433
2022-10-10	73568.16163878934
2022-10-17	64289.535036973466
2022-10-24	63412.73836248913
2022-10-31	70135.7612205564
2022-11-07	65810.8538180039
2022-11-14	57487.76961564558
2022-11-21	57941.0379590933
2022-11-28	53088.99276246144
2022-12-05	45083.711067050994
2022-12-12	48539.16003861715
2022-12-19	40030.196806653636
2022-12-26	47788.59072442148
2023-01-02	48027.82045957172
2023-01-09	45981.680604235546
2023-01-16	45371.11228953819
2023-01-23	42584.55125854277

Перейдите в раздел "Параметры".



самолет

Интеграция: архитектура



Flask сервер с
моделью

Gin-сервер

Пользователь



самолет

Вывод по проделанной работе

- Удалось собрать качественный датасет для прогнозирования цен на арматуру.
- Разработаны три модели с разной сложностью и точностью.
- Улучшена предсказательная способность за счет дополнительных признаков.
- Настроена интеграция с веб-интерфейсом через API.
- Основные проблемы связаны с обработкой данных и оптимизацией моделей, но были успешно решены.

