

Прогнозирование продаж

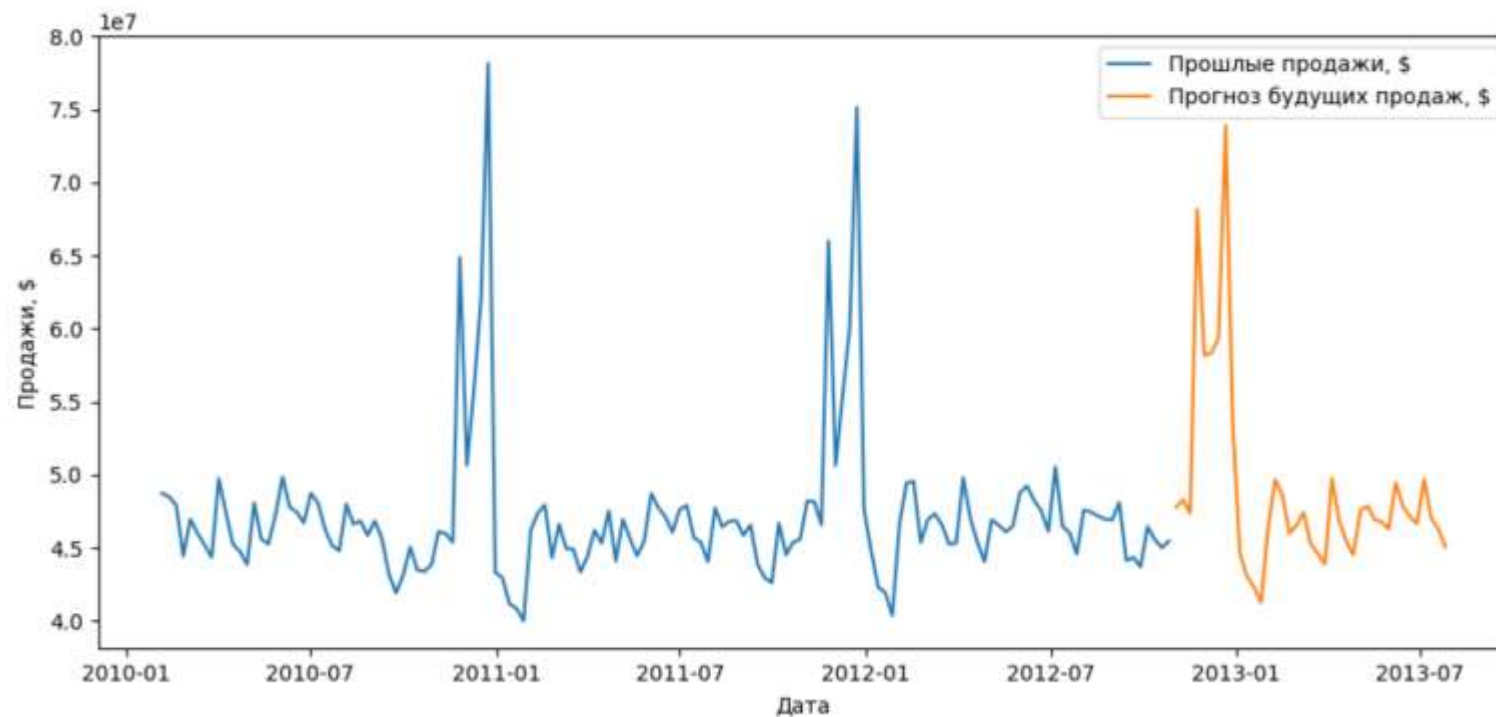
Прогнозирование продаж — ключевая задача для бизнеса, поскольку оно помогает более точно организовать производство, управлять запасами, планировать маркетинговые кампании и распределять финансовые ресурсы.

Авторы: Захаров Иван, Лукьянчик Ян, Савинов Сергей, Шорохова Юлия.



Цель проекта

Разработать модель для точного прогнозирования будущих продаж на основе данных о прошлых продажах, сезонных трендах и других факторах.



Ознакомление с данными

train.csv содержит данные о еженедельных продажах:

Store: Идентификатор магазина.

Dept: Идентификатор отдела.

Date: Дата продаж (в формате YYYY-MM-DD).

Weekly_Sales: Объем продаж за неделю.

IsHoliday: Указатель на праздничную неделю.



Ознакомление с данными

features.csv: Дополнительные экономические данные и информация о промо-акциях.

- * Store: Идентификатор магазина.
- * Date: Дата (в формате YYYY-MM-DD).
- * Temperature: Средняя температура за неделю.
- * Fuel_Price: Цена на топливо.
- * Markdown(n): Скидка (n).
- * CPI: Индекс потребительских цен.
- * Unemployment: Уровень безработицы.
- * IsHoliday: Указатель на праздничную неделю.



Ознакомление с данными

stores.csv: Информация о магазинах.

- * Store: Идентификатор магазина.
- * Type: Тип магазина.
- * Size: Размер магазина (в квадратных футах).



Этапы

- Коррекция данных.

(обработка пропущенных значений, исправление ошибок и выявление выбросов)

- Проведение исследовательского анализа данных (EDA) и Feature engineering.

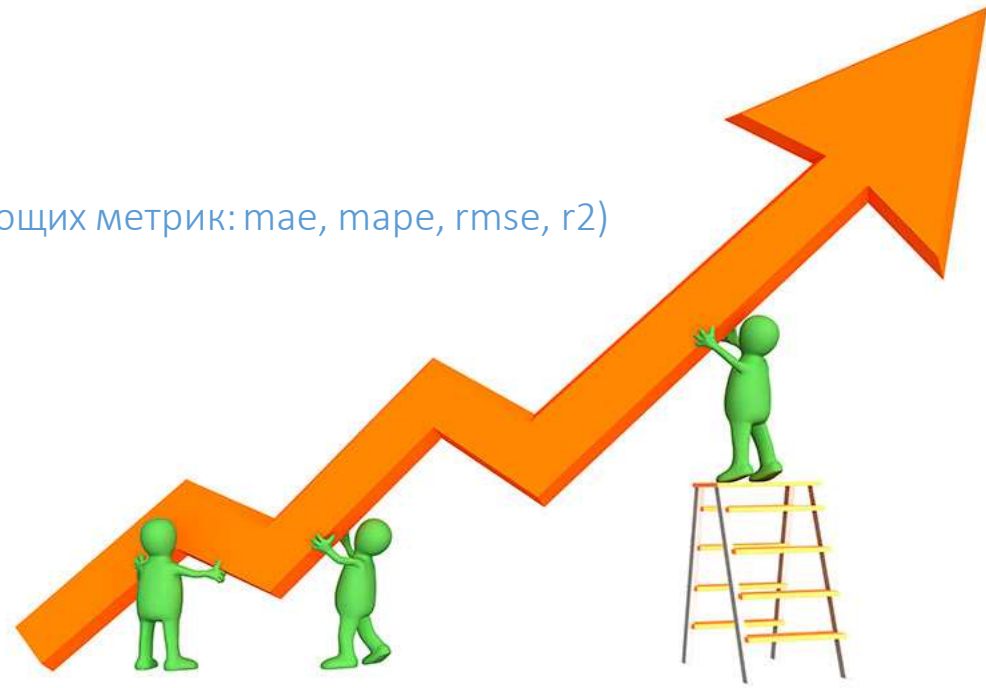
(выявление важных данных и создание новых)

- Выбор рабочей модели.

(выбор среди предложенных моделей лучшую, с использованием следующих метрик: mae, mape, rmse, r2)

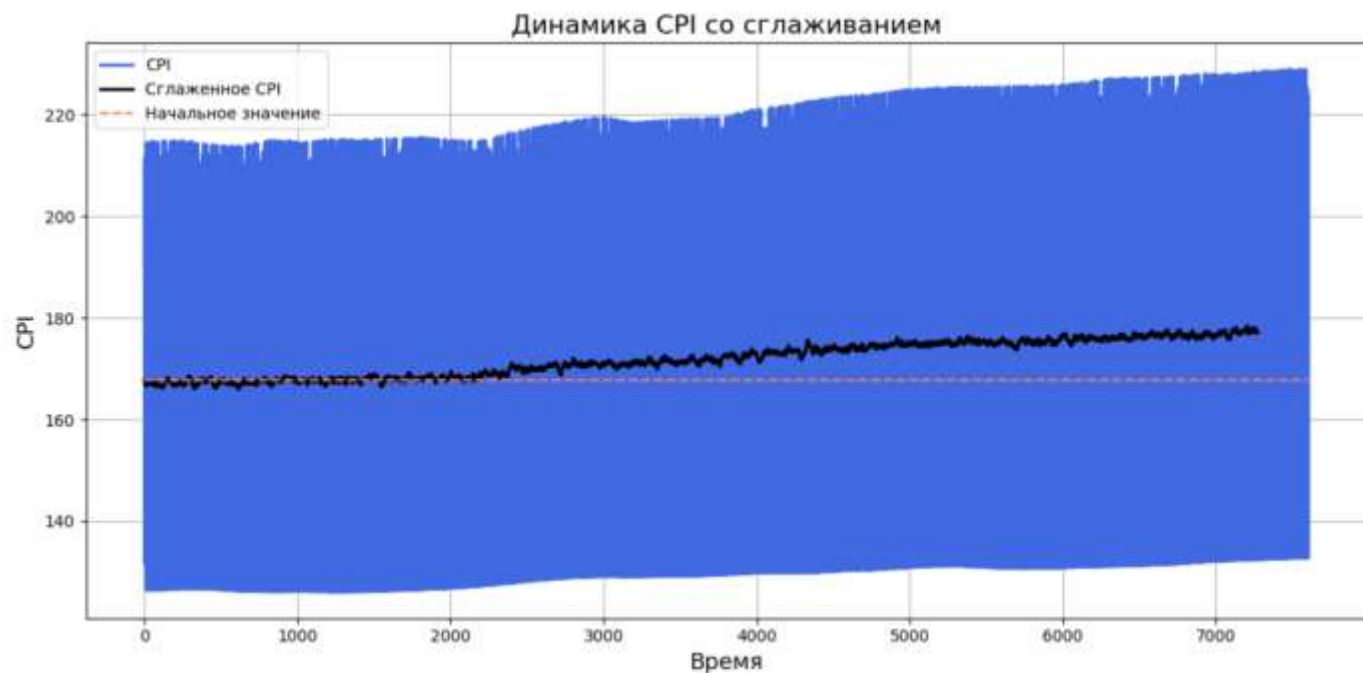
- Создание прототипа приложения.

(рабочий прототип с минимальным интерфейсом)



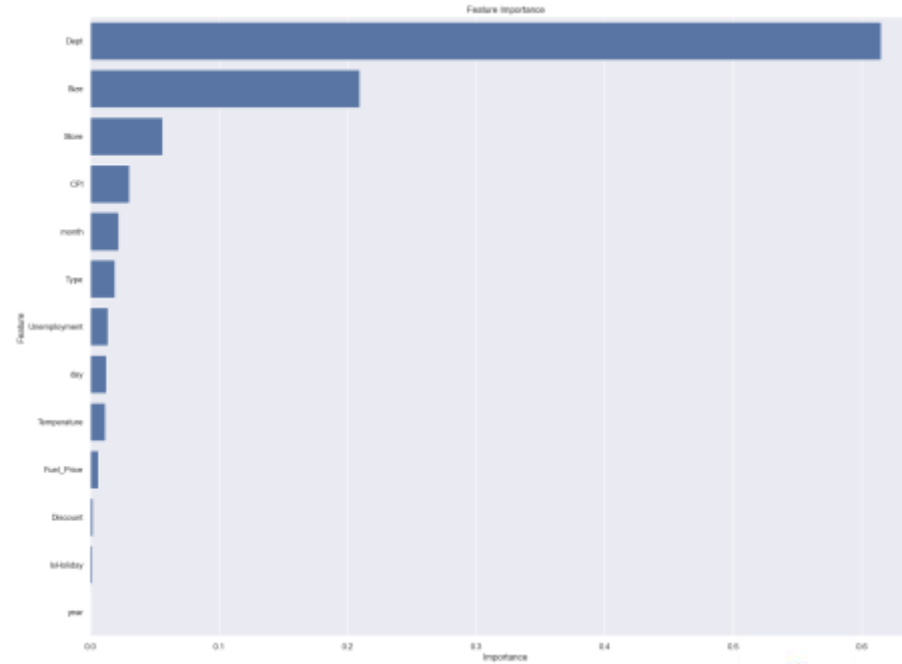
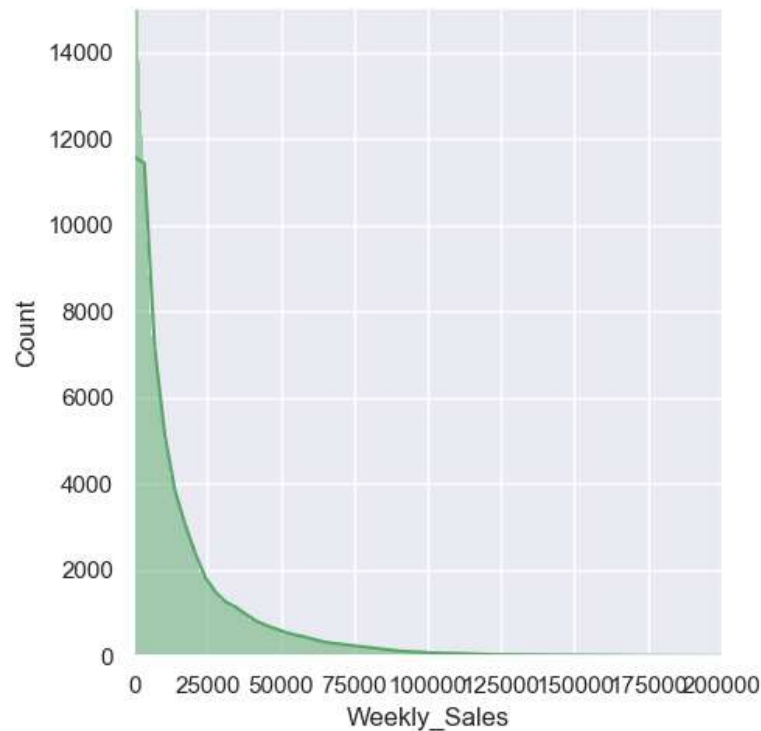
Коррекция данных

- Значения `nan` в столбцах Markdown заменены на нули.
 - Значения типа `bool` конвертированы в `int`.
 - Столбец `date` разбит на четыре столбца и конвертирован в `int`.
 - Столбцы `Markdown` просуммированы в столбце `Discount`.
 - ...
- нны.



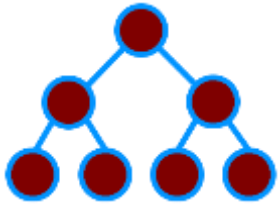
EDA и Feature engineering

- Убраны отрицательные значения продаж.
- Выявлены важные признаки для обучения.
- Созданы новые столбцы в таблице.



Выбор рабочей модели

XGBRegressor



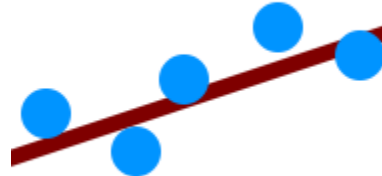
MAE: 3315.37

RandomForest



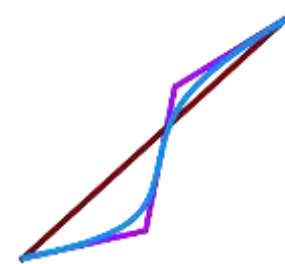
MAE: 1767.68

Линейная регрессия



MAE: 14622.08

ElasticNet

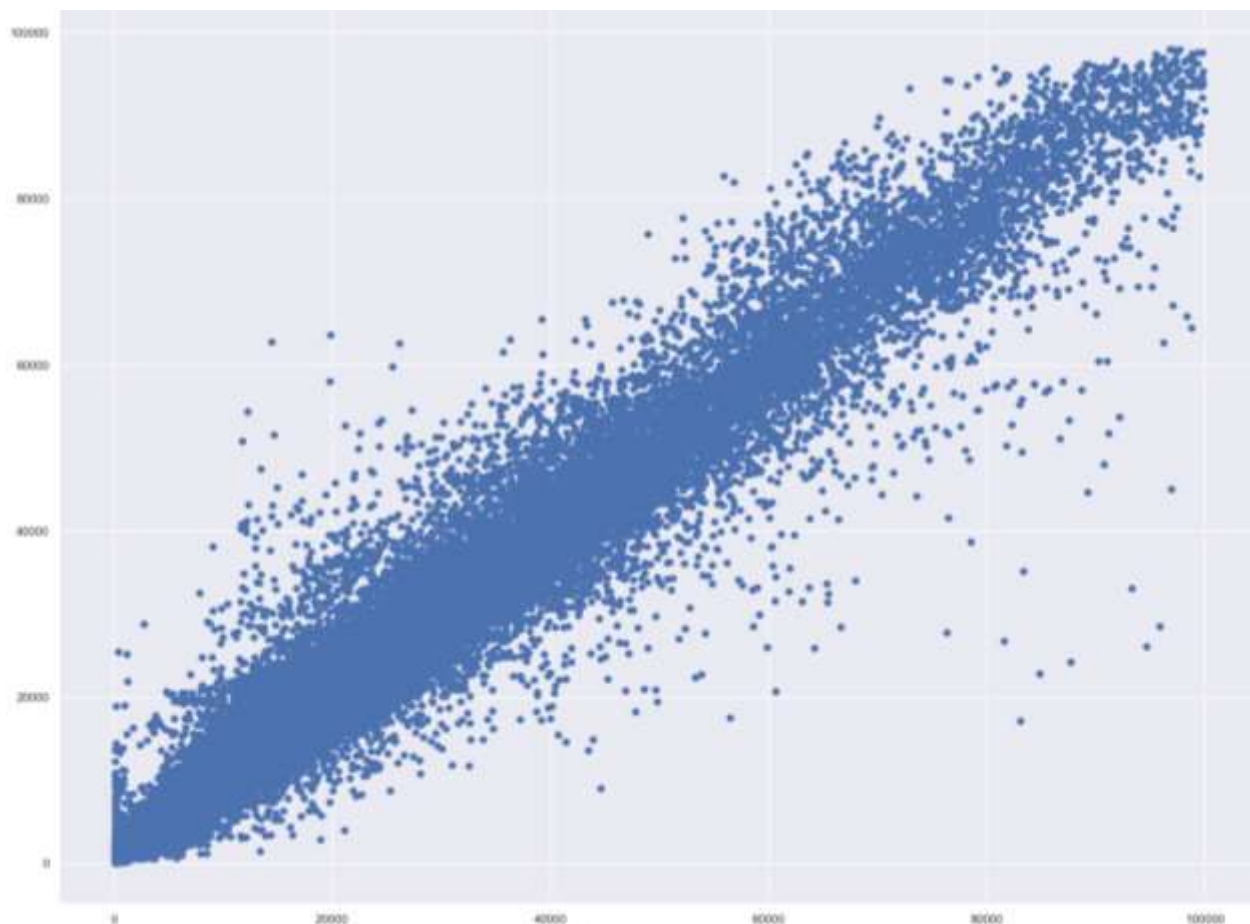


MAE: 14672.45



Лучшая модель

RandomForestRegressor



mean absolute error: 1321.8543718935089

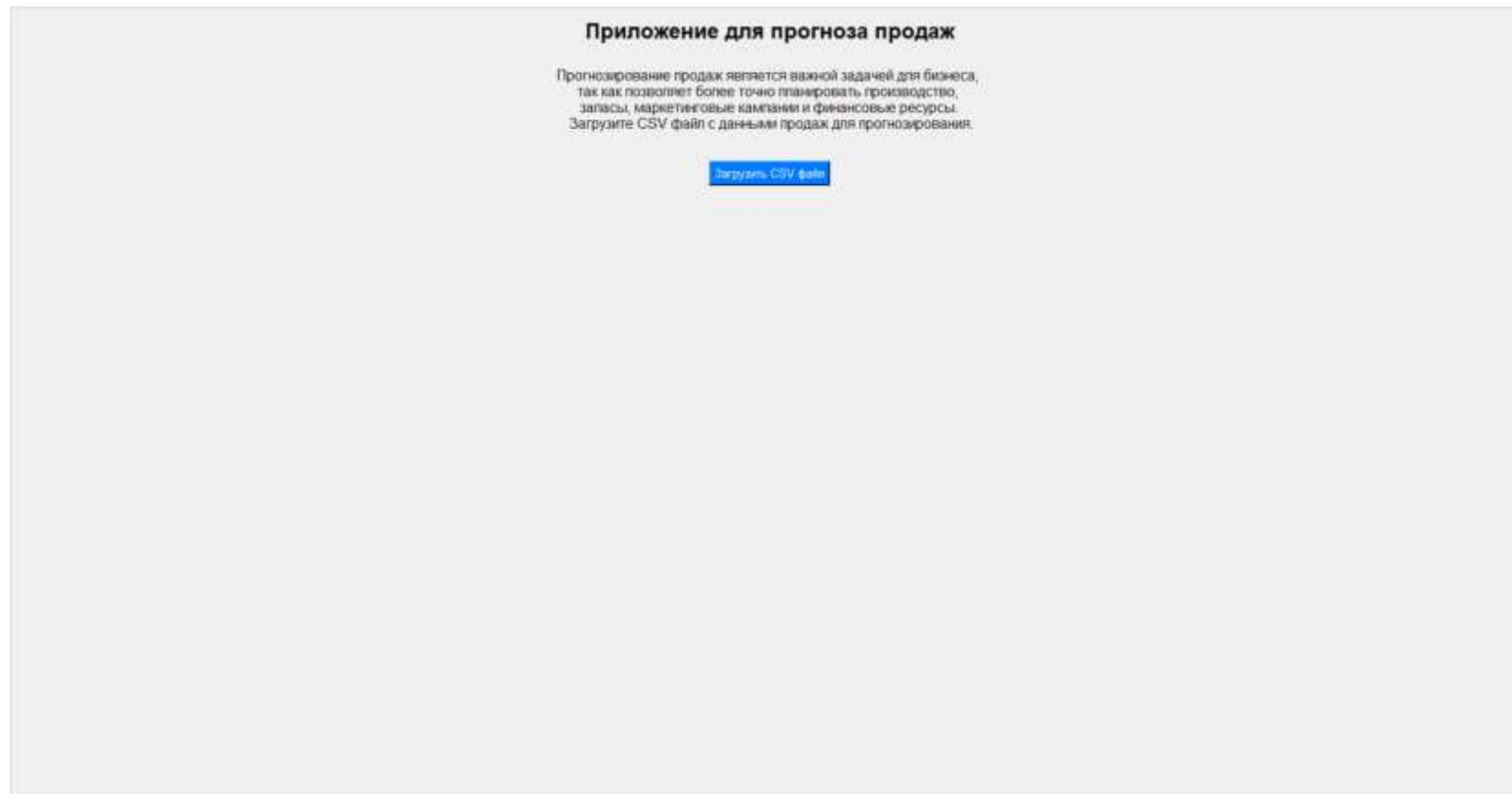
mape 0.6863329126815438

mse: 10149278.273816774

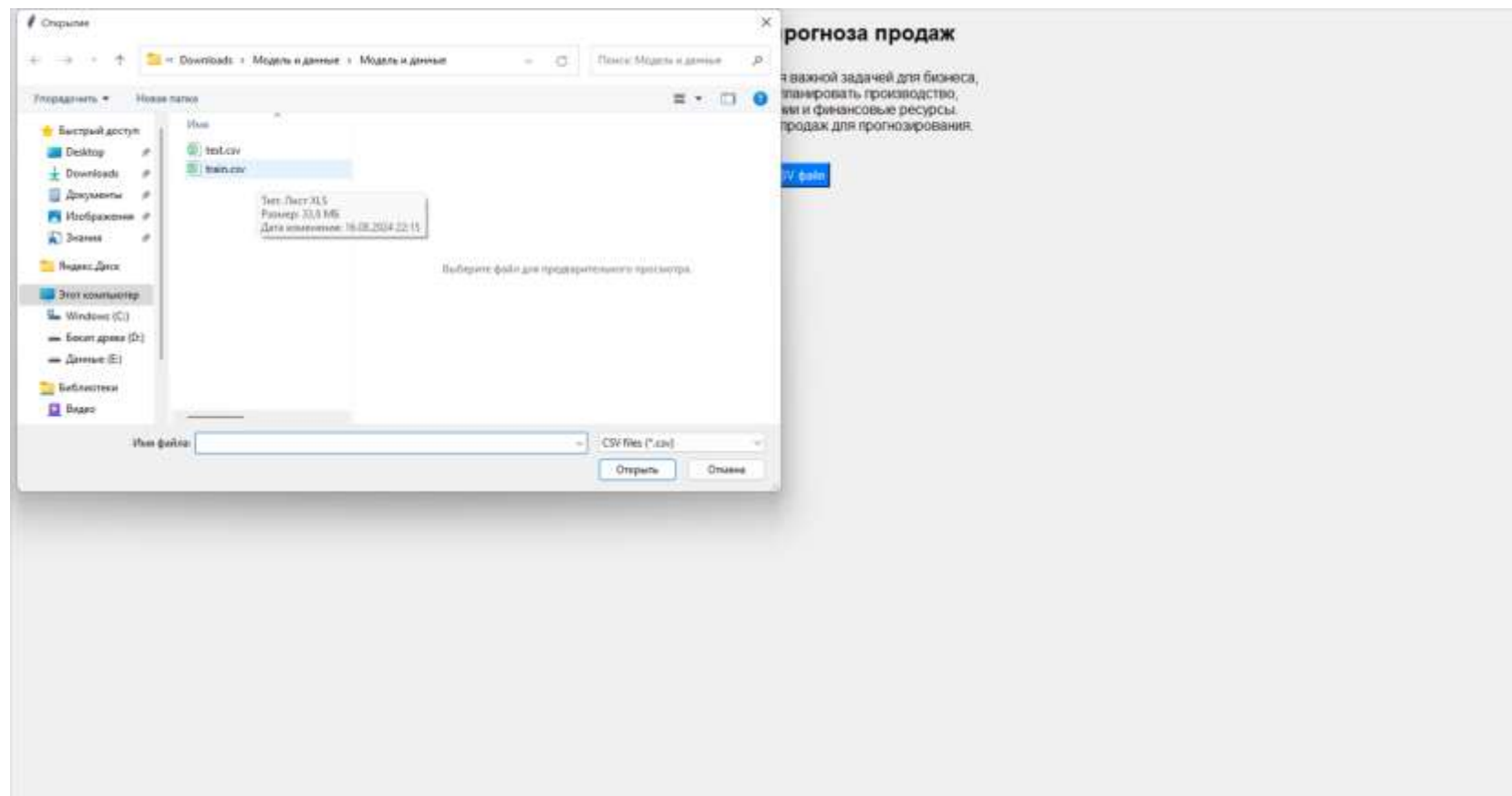
r2_score: 0.9801980131207815



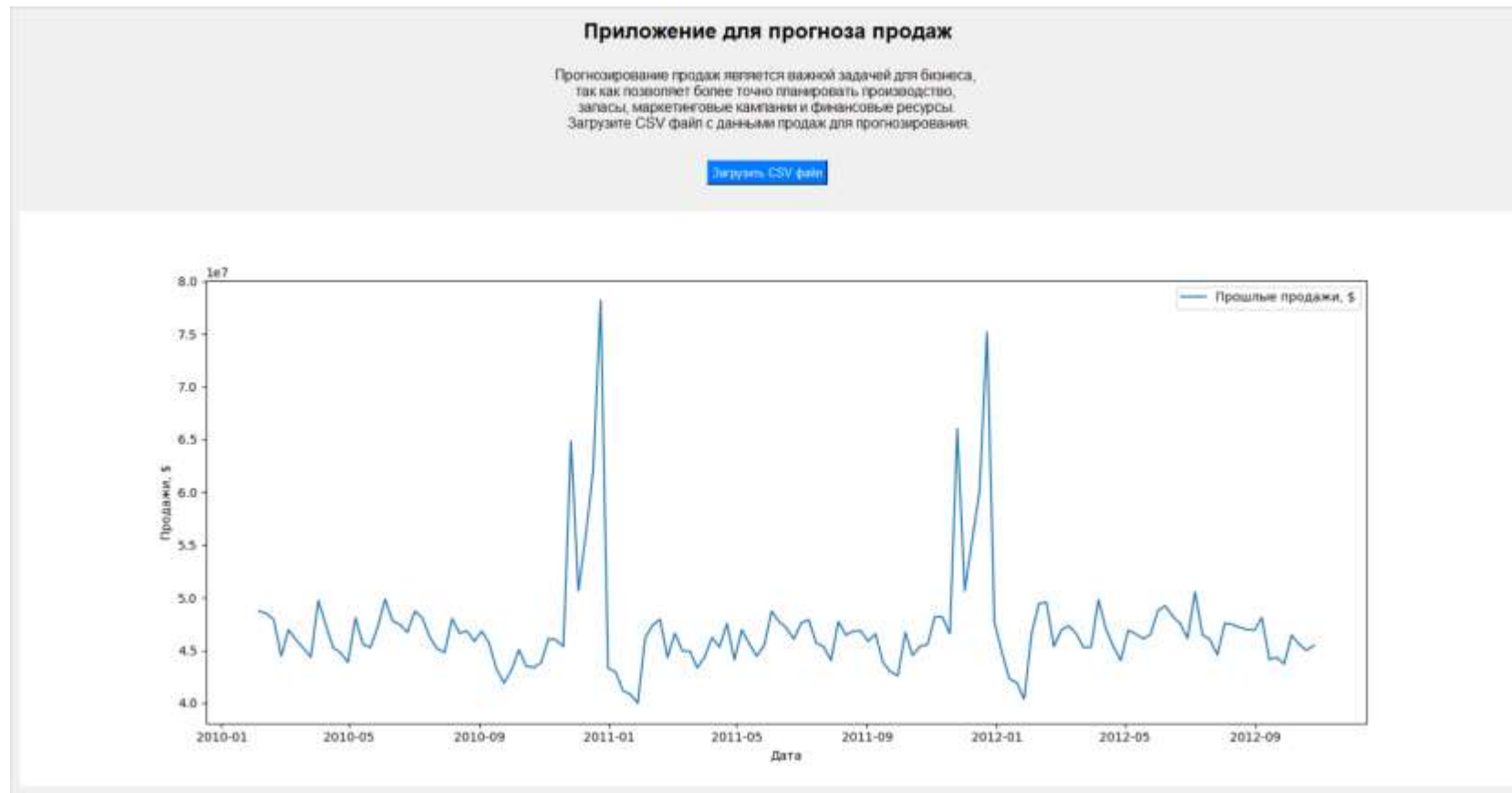
Создание прототипа приложения



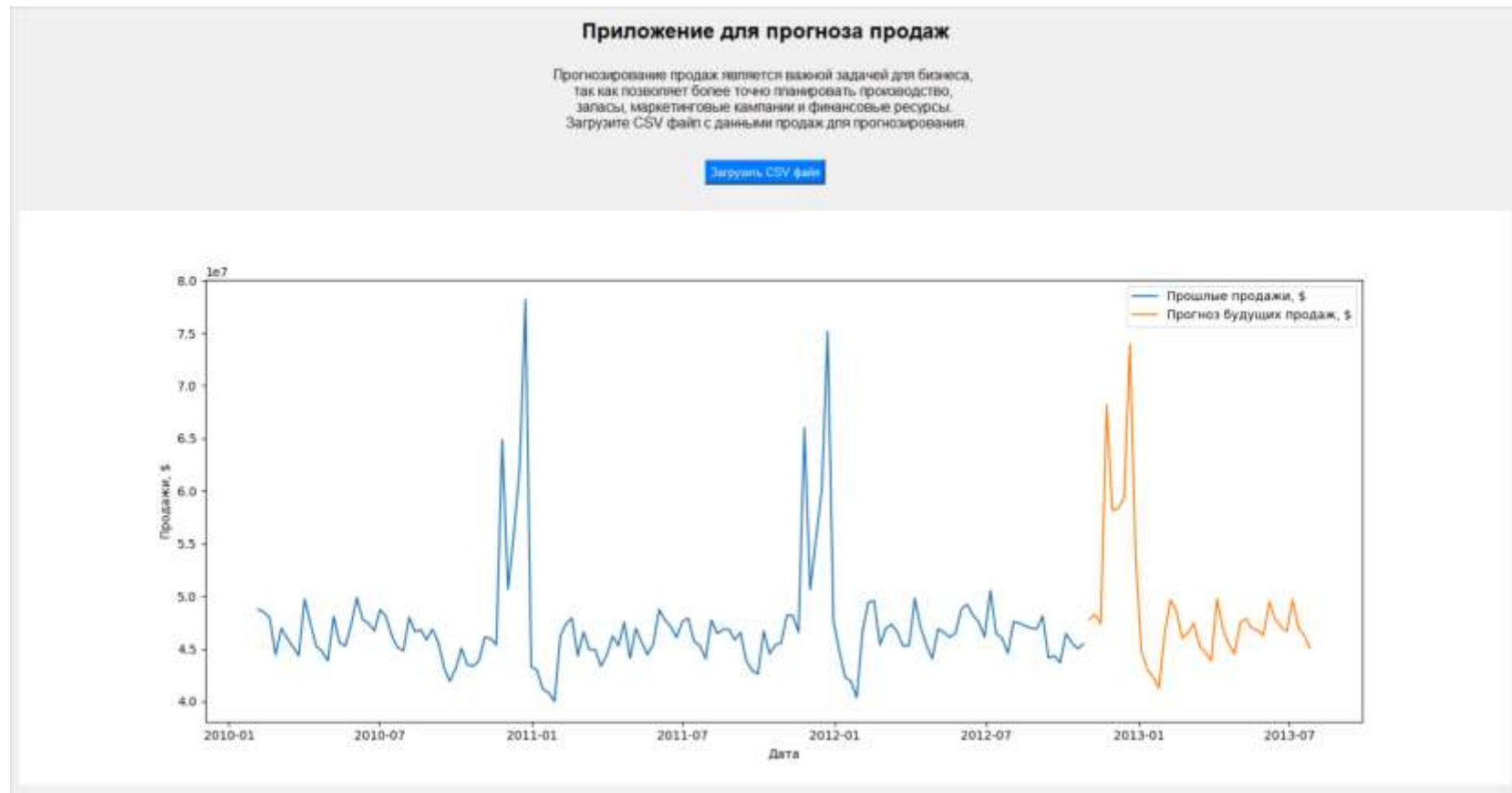
Создание прототипа приложения



Создание прототипа приложения



Создание прототипа приложения



Спасибо за внимание.

