



ИНСТИТУТ
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
УНИВЕРСИТЕТА ИННОПОЛИС



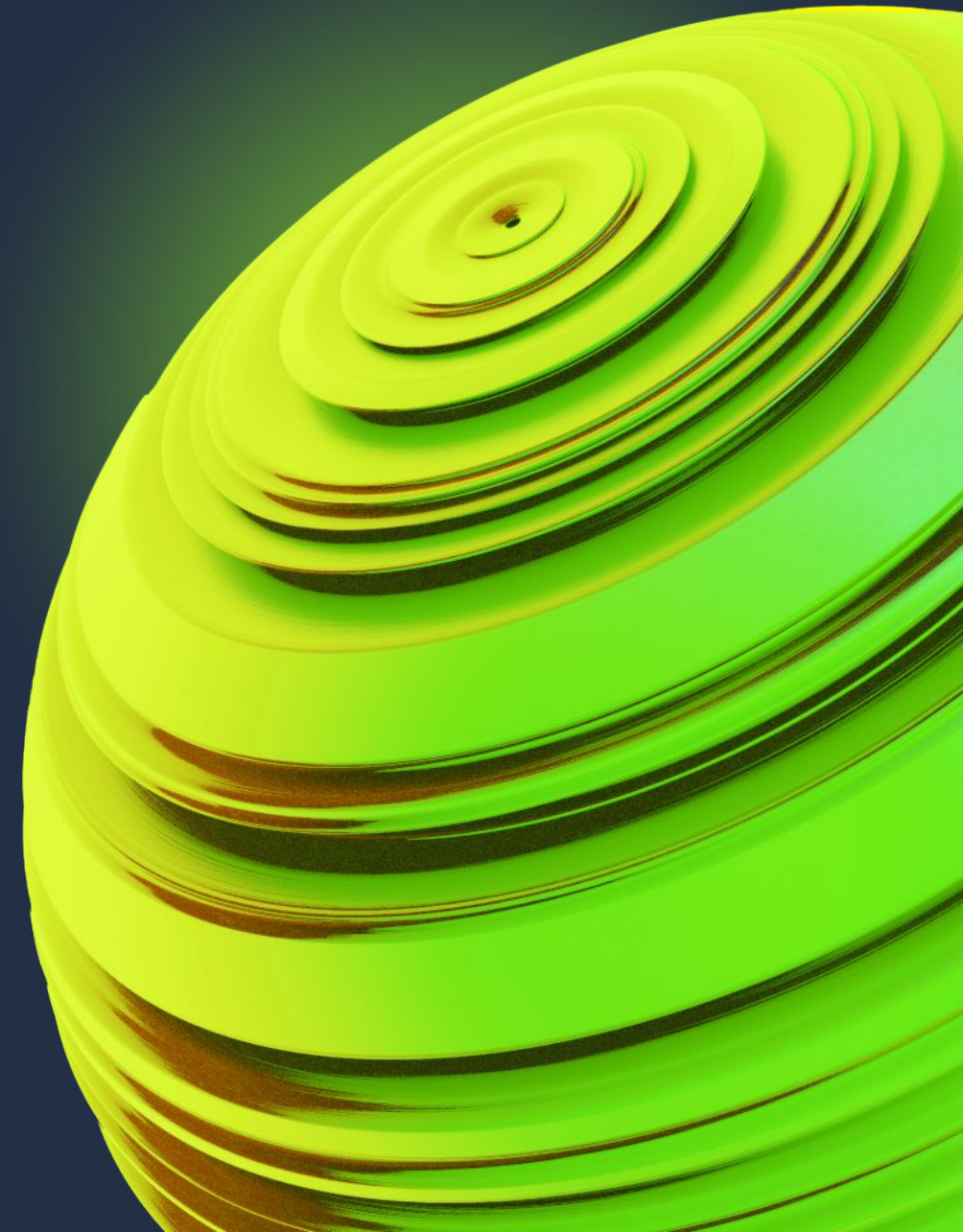
УНИВЕРСИТЕТ
ИННОПОЛИС

Прогнозирование оттока клиентов в E-commerce

Автор: Кошелев Савелий

Курс: Архитектор в области искусственного интеллекта

Дата: Декабрь 2025 год





ИНСТИТУТ
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
УНИВЕРСИТЕТА ИННОПОЛИС



УНИВЕРСИТЕТ
ИННОПОЛИС

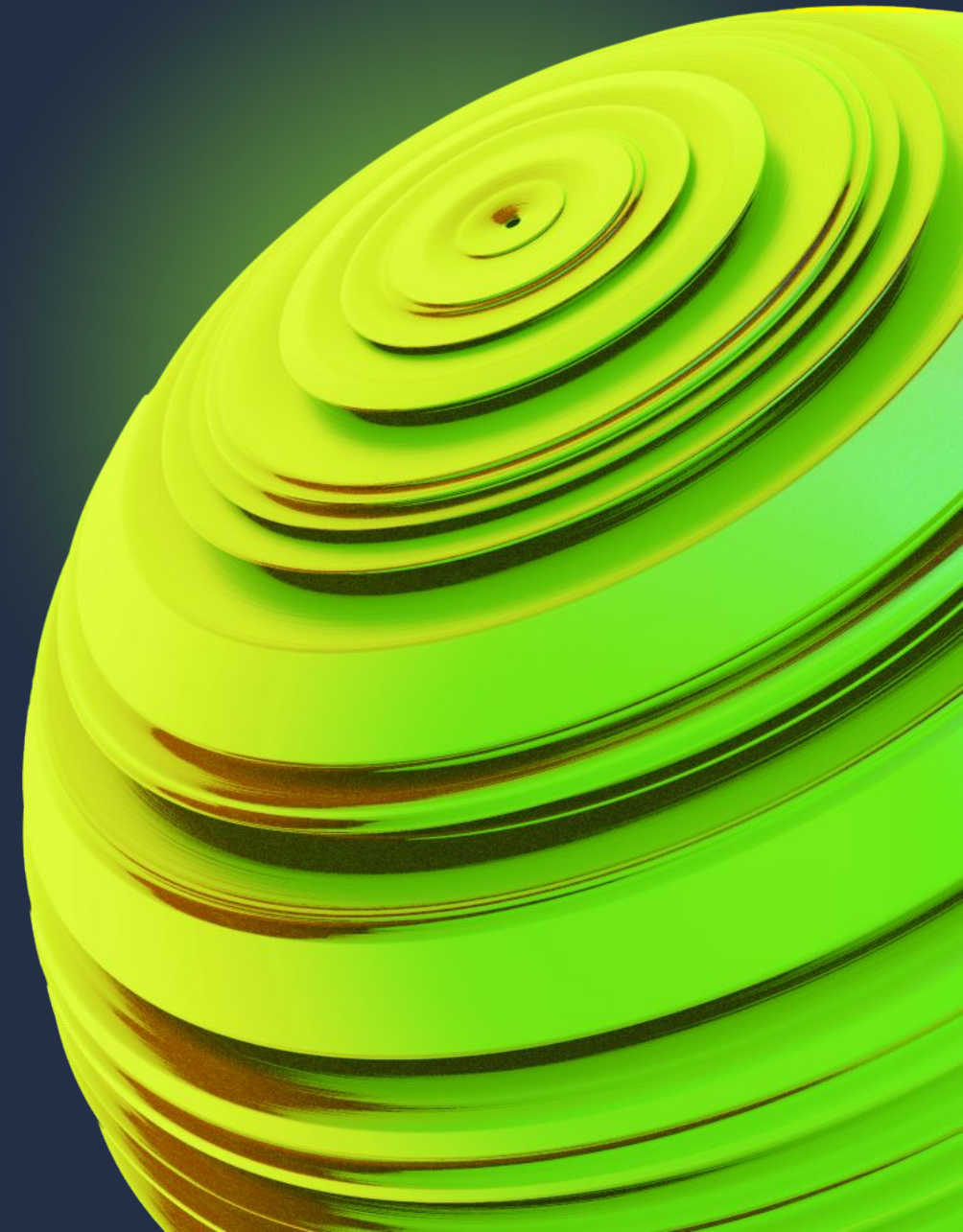


Задача:

Прогнозирование оттока клиентов (Customer Churn Prediction)

Описание:

Построить систему, которая предсказывает вероятность того, что клиент перестанет совершать покупки в течение следующих 3 месяцев.

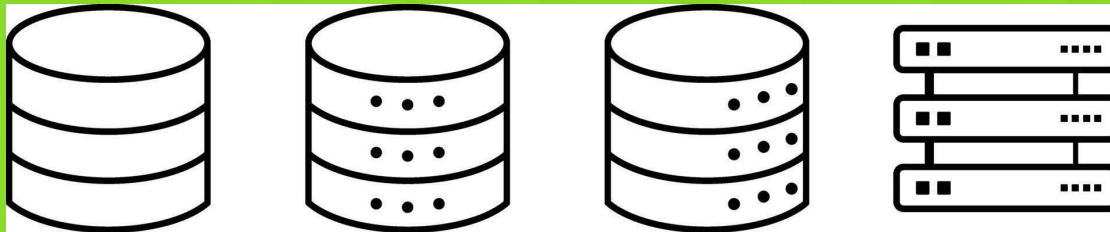




ИНСТИТУТ
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
УНИВЕРСИТЕТА ИННОПОЛИС



УНИВЕРСИТЕТ
ИННОПОЛИС



Датасет:

E-commerce Customer Churn Dataset

- 5630 клиентов
- 20 признаков



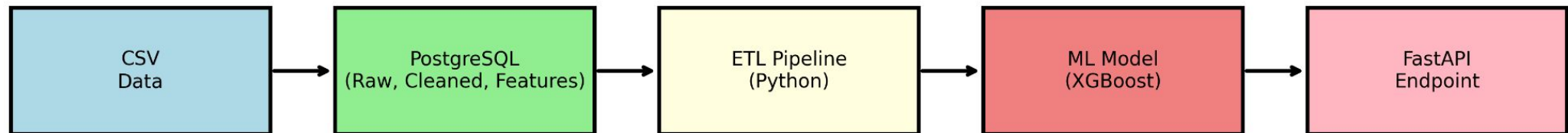
Бизнес-требования:

- $\text{Precision} \geq 0.70$
- $\text{Recall} \geq 0.65$
- Время инференса < 100 мс





Архитектура системы прогнозирования оттока клиентов




Анализ данных (EDA)



 5630 клиентов, 20 признаков

 16.5% ушли (дисбаланс)

 Пропуски: 4–6% (обработаны)

 Важные признаки:

- Complain
- CashbackAmount
- Tenure



Предобработка

Предобработка данных

Заполнение пропусков (медиана)

Кодирование категориальных

Новый признак: CustomerLoyalty

$$\text{CustomerLoyalty} = \text{Tenure} \times \text{OrderCount}$$

Где:

Tenure — сколько месяцев клиент с нами (стаж)

OrderCount — сколько всего заказов сделал клиент



Методология

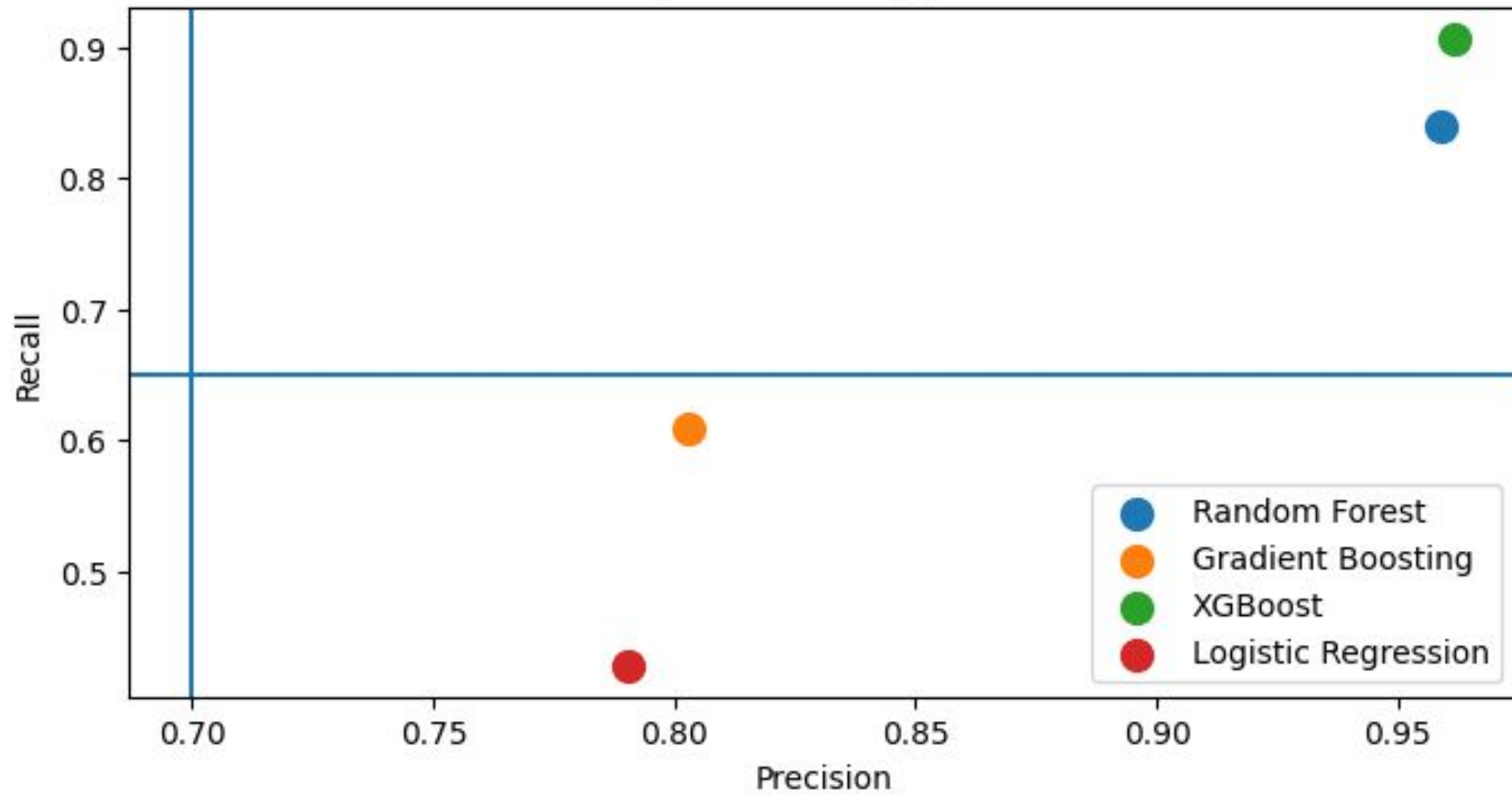


1. Baseline: Logistic Regression
2. Сравнение 4 моделей
3. Оптимизация XGBoost
4. Финальная оценка





Сравнение моделей

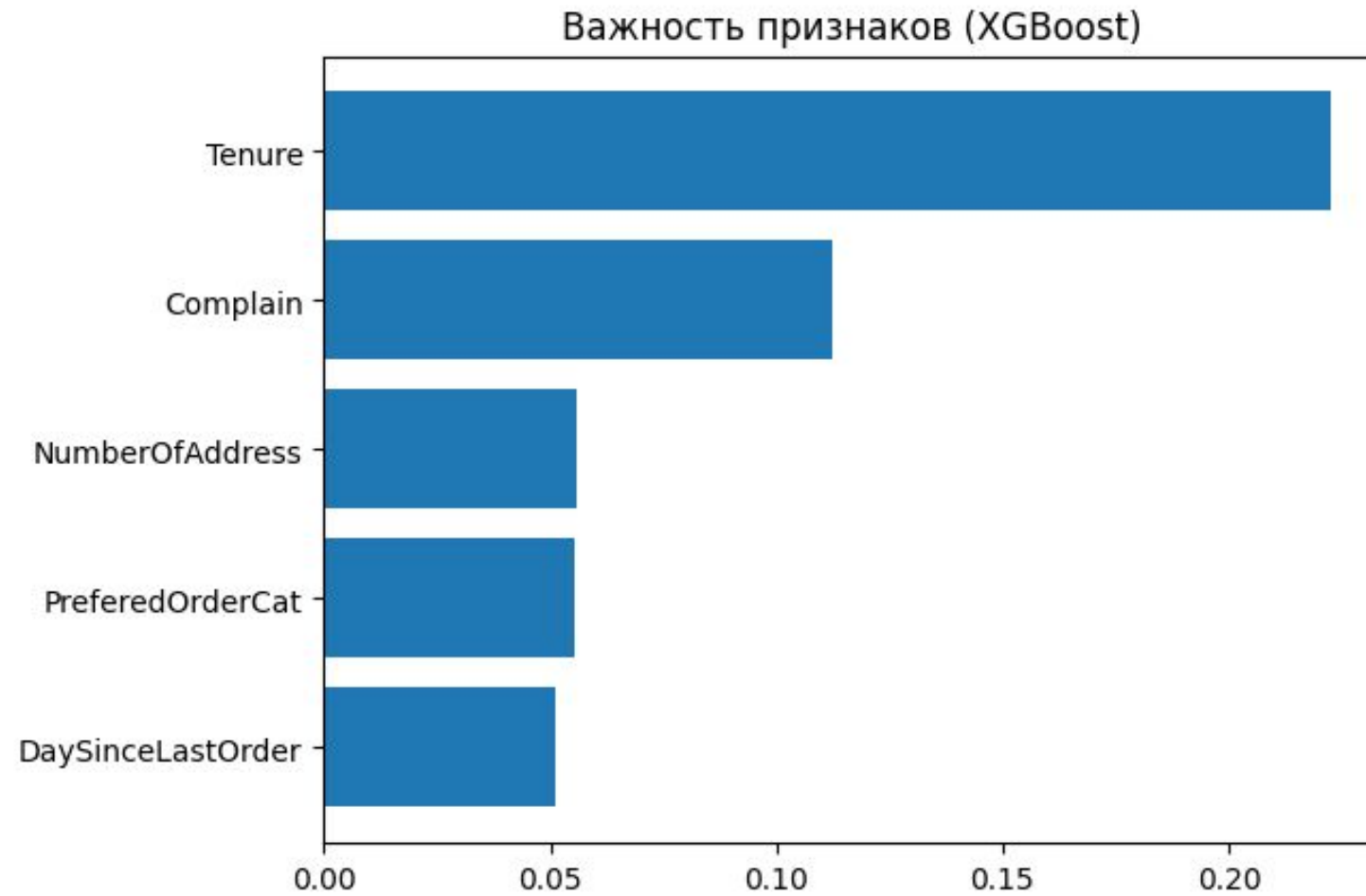


XGBoost

0.962

0.907



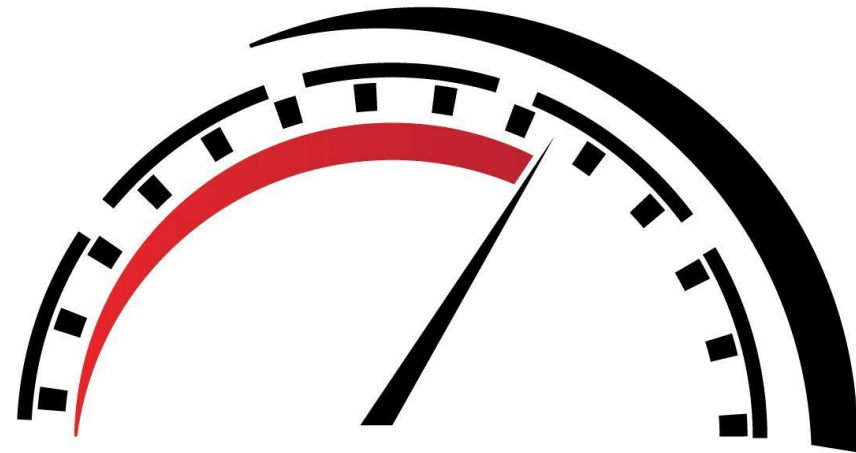




Оптимизация

Лучшие параметры:

- `n_estimators: 200`
- `max_depth: 7`
- `learning_rate: 0.1`
- `scale_pos_weight: 3`





Финальные метрики

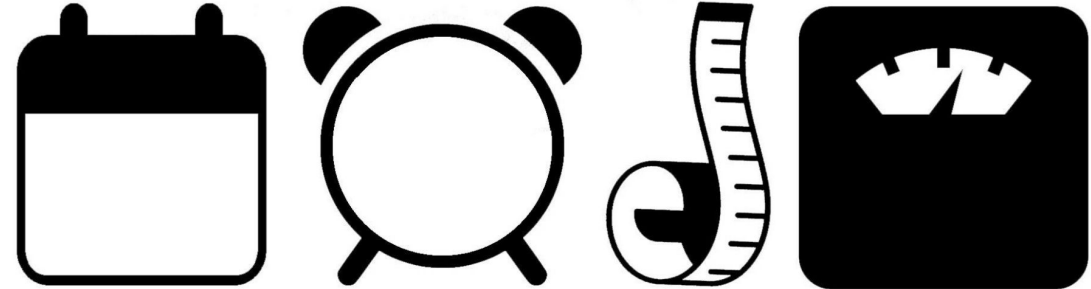
Precision: 0.930

Recall: 0.921

F1-score: 0.925

ROC-AUC: 0.983

Время: < 15 мс




Все требования выполнены

Реализация



Реализация проекта

-  IPYNB (4):
 - 01_data_preparation
 - 02_baseline_model
 - 03_model_experiments
 - 04_final_model



SQL-файлы (3 слоя)
Сохранённая модель (.pkl)

Репозиторий



- ▼ data/processed
 - ecommerce_churn_processed.csv
- ▼ docs
 - architecture.png
 - report.pdf
 - requirements.md
- ▼ models
 - metrics.json
 - xgb_final_model.pkl
- ▼ notebooks
 - 01_data_preparation.ipynb
 - 02_baseline_model.ipynb
 - 03_model_experiment.ipynb
 - 04_final_model.ipynb
- ▼ sql
 - layers.sql
 - queries.sql
 - schema.sql
- > src
 - README.md
 - architecture.png
 - requirements.txt



Выводы

Проект демонстрирует полный цикл архитектора ИИ:

Сбор требований – бизнес-метрики Precision/Recall

Проектирование – 3-слойная архитектура данных

Разработка – 4 модели, оптимизация XGBoost

Тестирование – валидация, статистические тесты

Результат: модель с Precision 0.930, Recall 0.921

Все бизнес-требования выполнены





ИНСТИТУТ
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
УНИВЕРСИТЕТА ИННОПОЛИС

Спасибо за внимание!

Контакты

✉ savelii.k28@gmail.com

