

Определение целевого сегмента клиентов для таргетированной рекламы

Кому лучше посылать письма



Наименование задачи:
Задача №2 ПЭК



Представлено командой:
Мы от дяди Баира



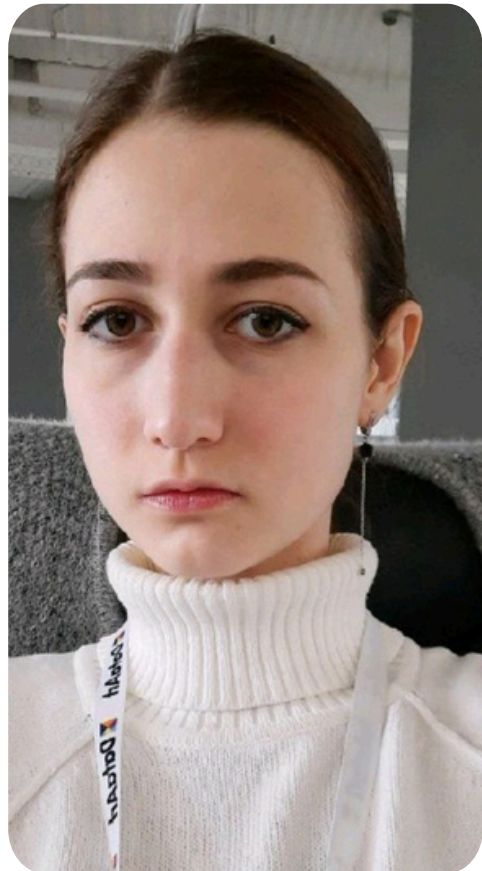
Заказчик:
ООО "ПЭК"



● ● ● Наша команда



**Дмитрий
Зайцев**



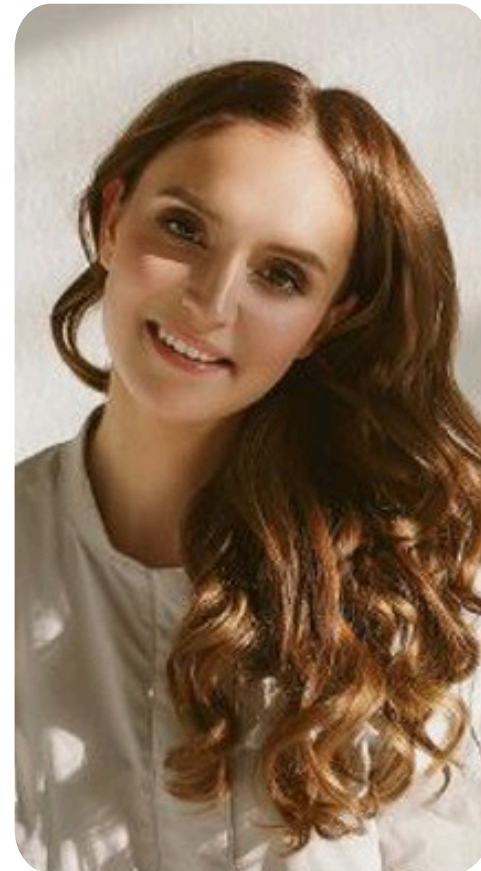
**Карина
Акчурина**



**Елена
Алмаева**



**Михаил
Кузьменков**



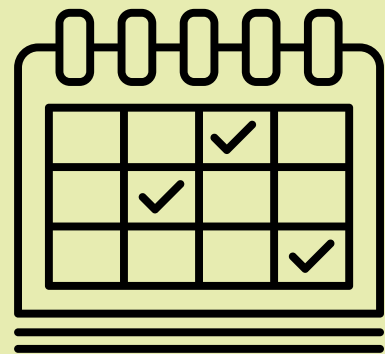
**Ольга
Редченко**



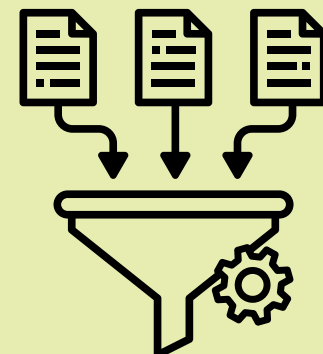
**Баир
Вамбуев**

Таймлайн проекта

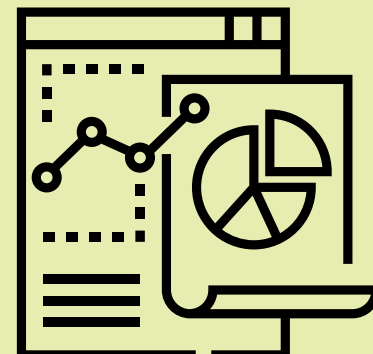
Длительность реализации - 20 дней



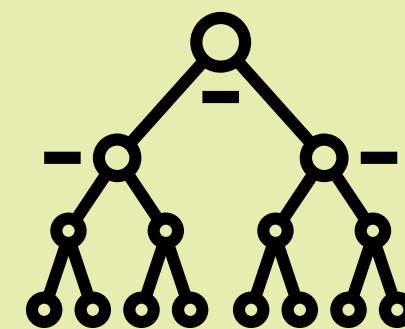
Планирование



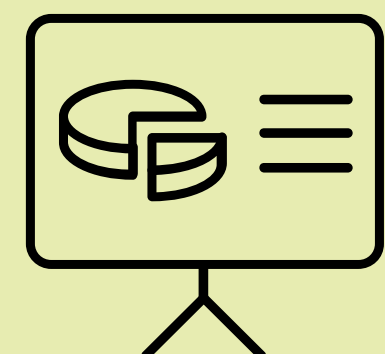
Предобработка
данных



Анализ данных



Обучение
моделей



Подведение
итогов

Предобработка данных



Преобразовано

1



Столбец UnitType

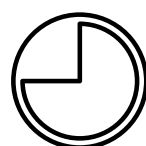


Столбец Amount

Добавлено

2

Признаки:



кварталы



регион



тип услуг



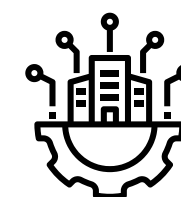
тип поддержки



тип перевозки

Обнаружено

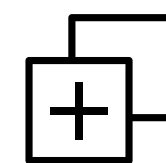
3



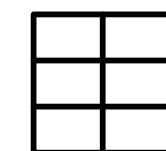
пропуски в ОКВЭД



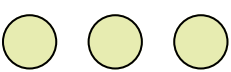
отсутствие
аномалий



отсутствие
дубликатов



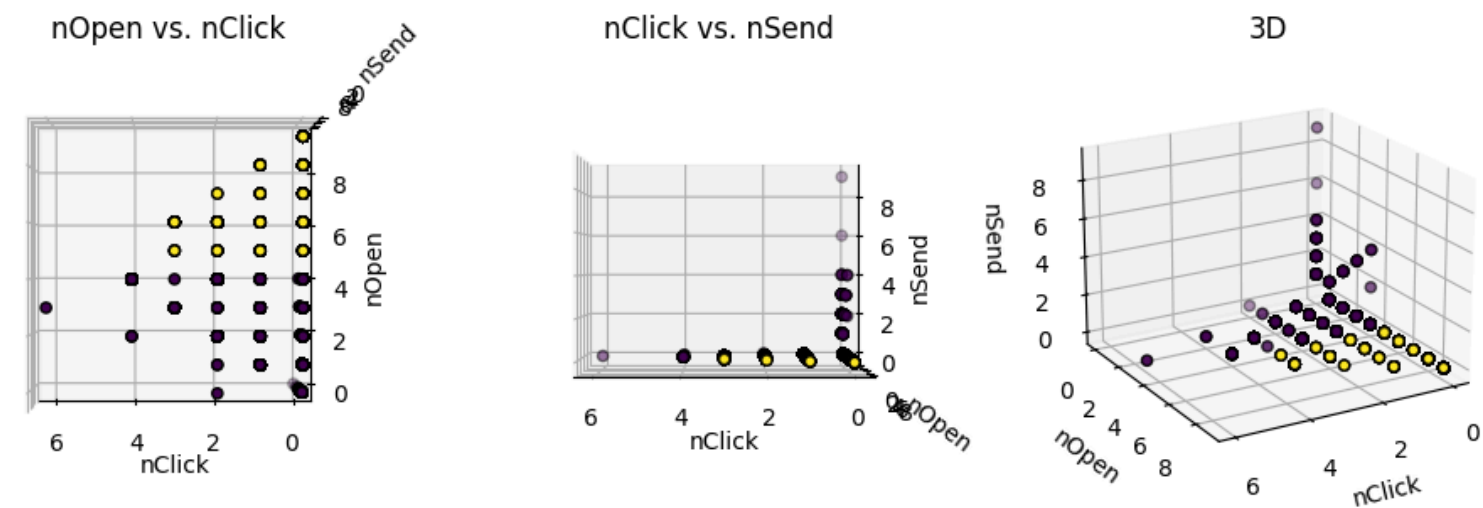
лишние столбцы
после обработки



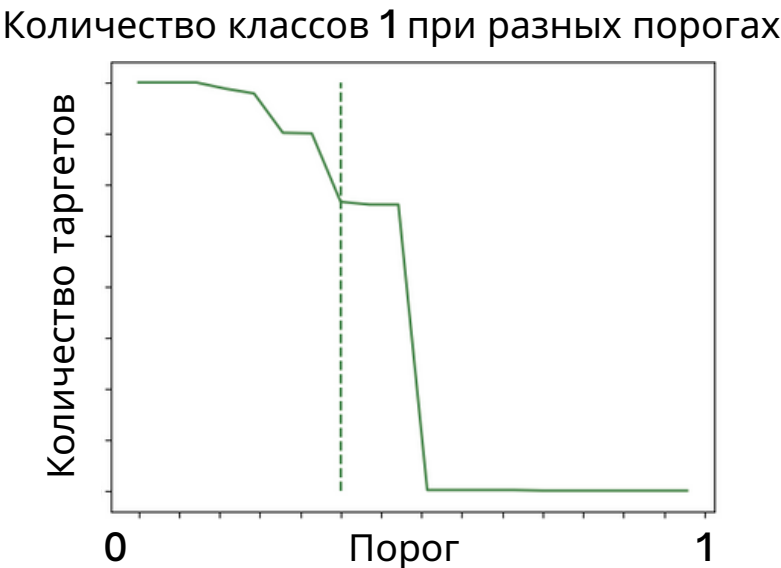
Анализ данных



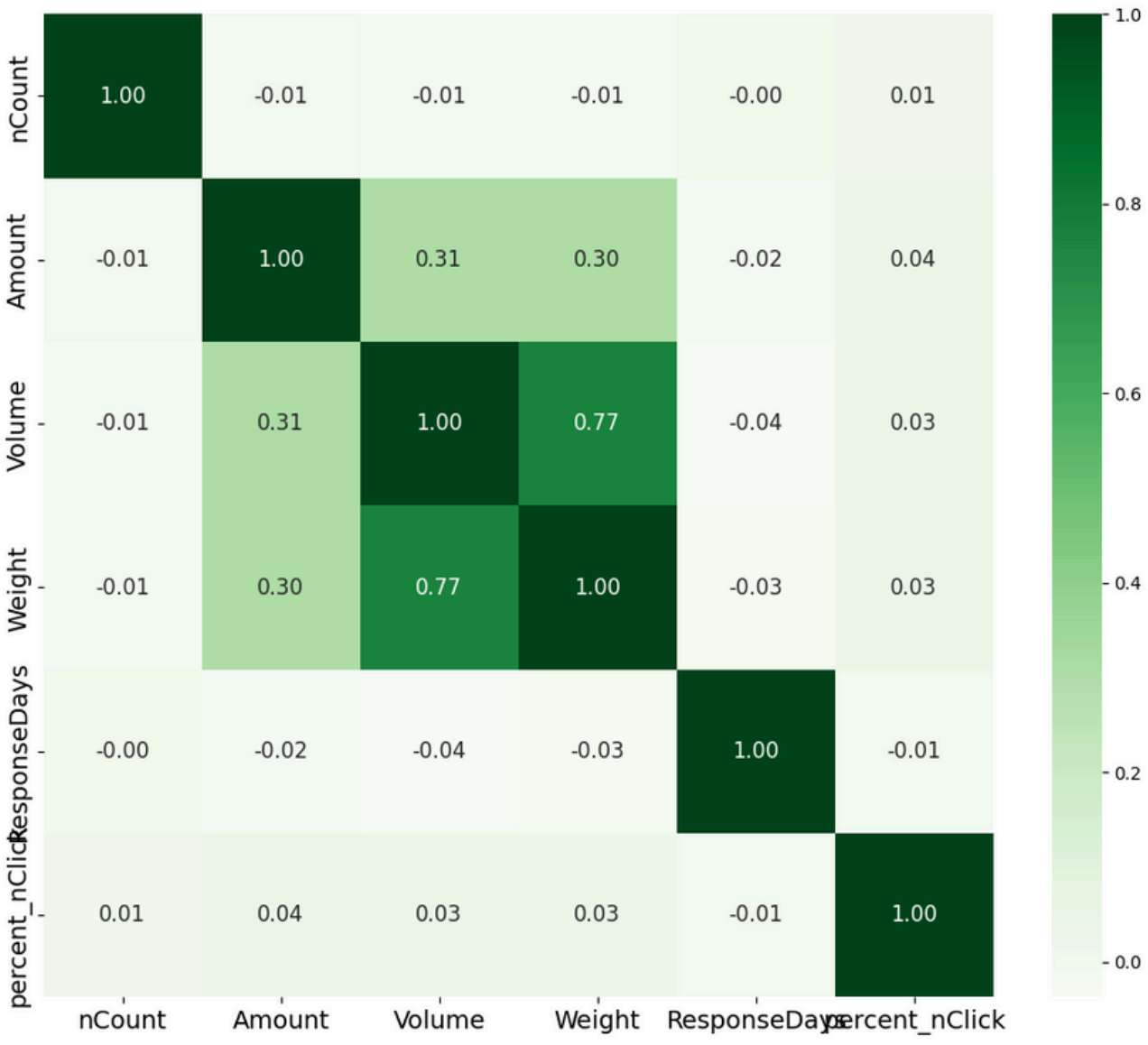
Кластеризация показывает случайное разбиение на 2 класса



Определение бинарного признака на пороге в 0,35



Есть линейная зависимость между объемом и весом



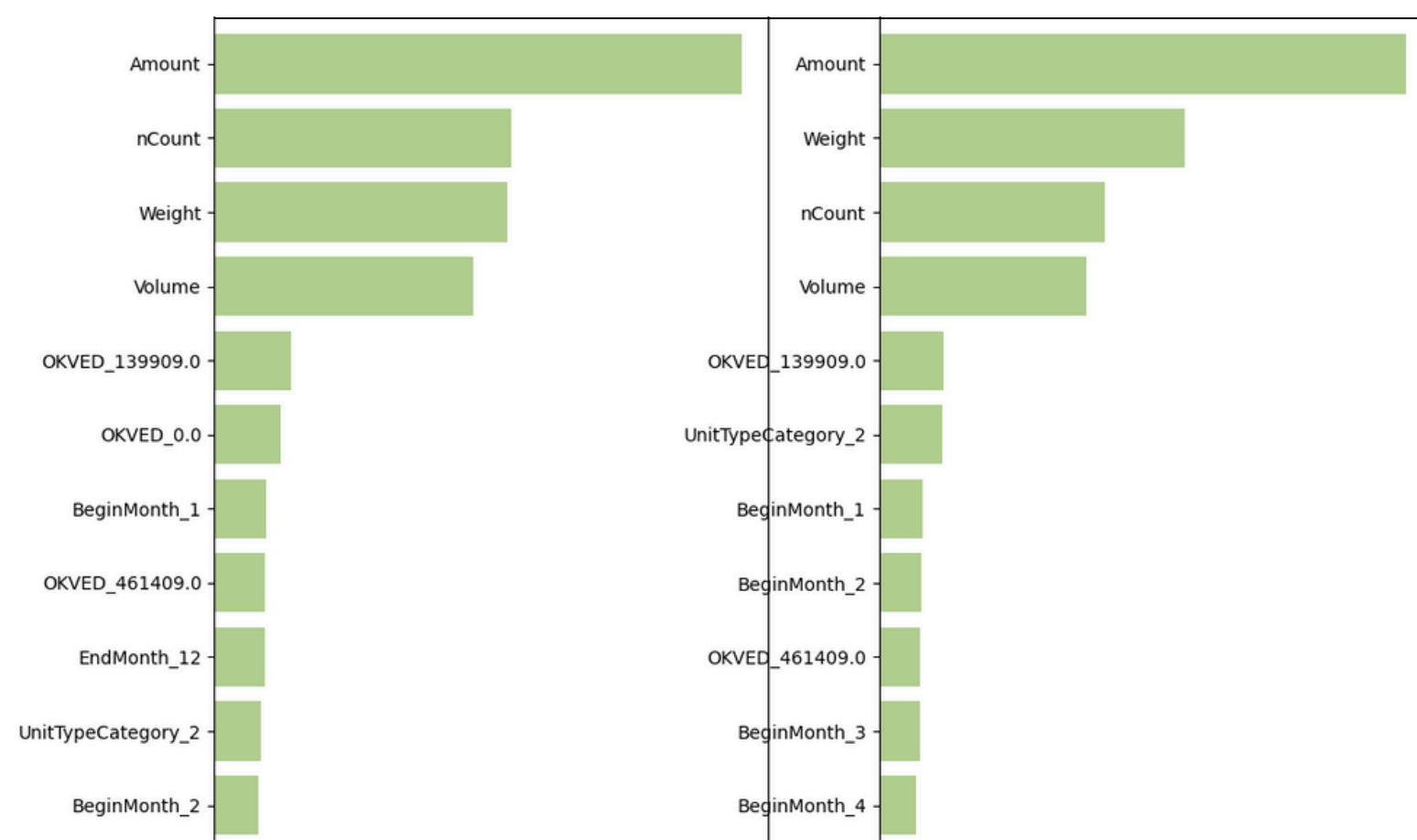
Отбор признаков



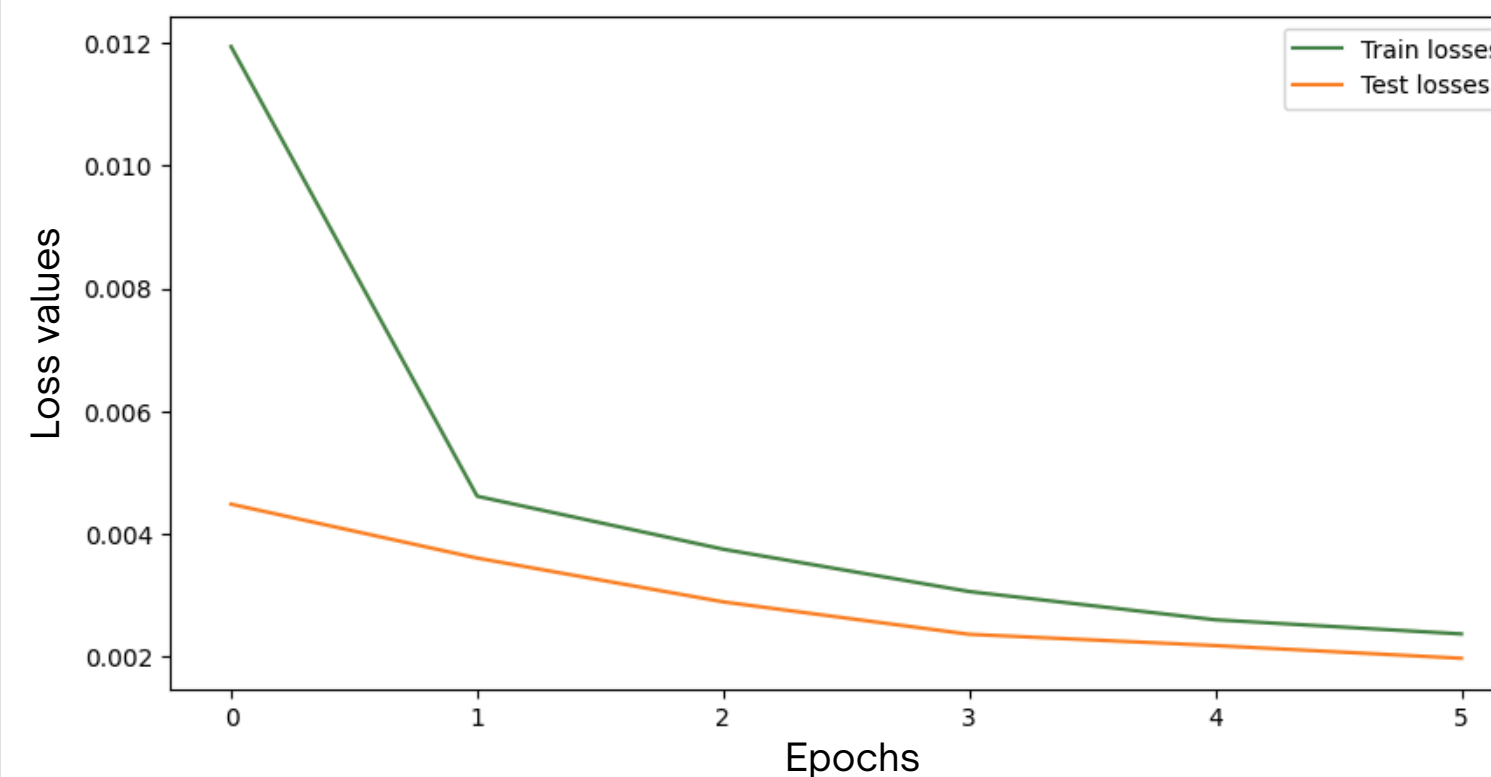
Отбор из 1155 признаков 198 необходимых
для обучения по критерию Gini

для классификации

для регрессии

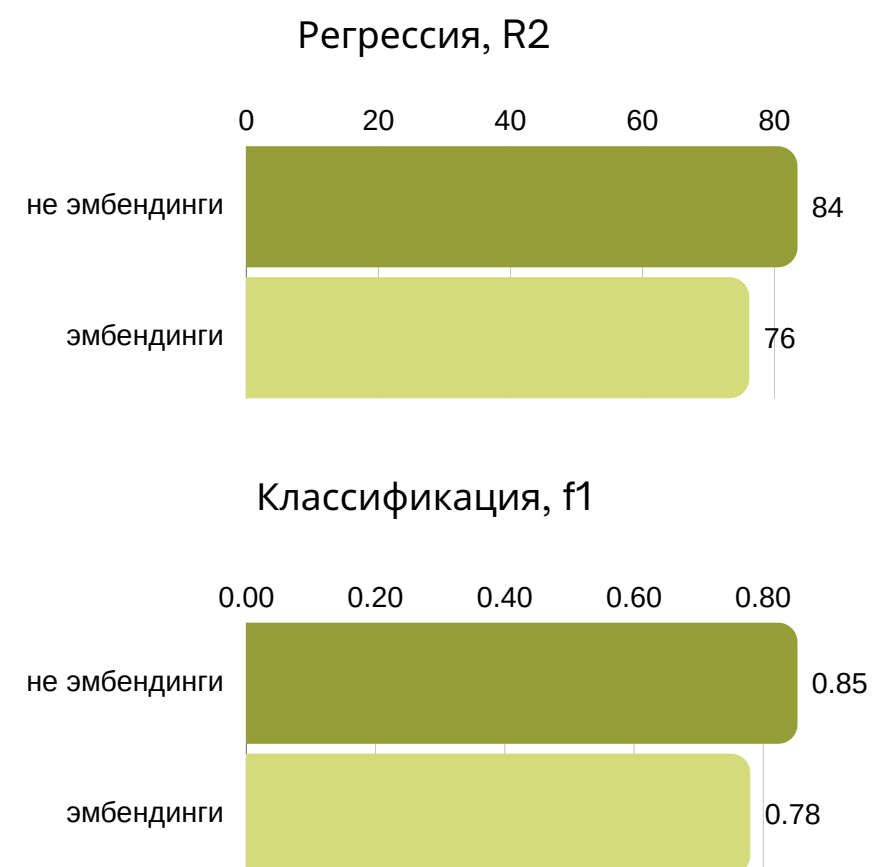


Создан эмбендинг признаков
Потери у сети снижаются, что означает
восстановление данных из скрытого
представления

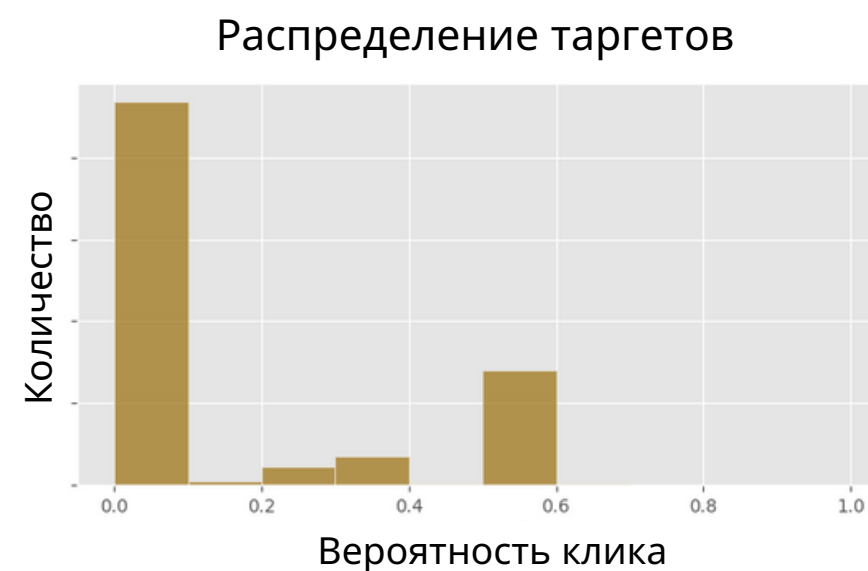


Обучение моделей

Обучение на эмбендингах признаков показало более низкие результаты

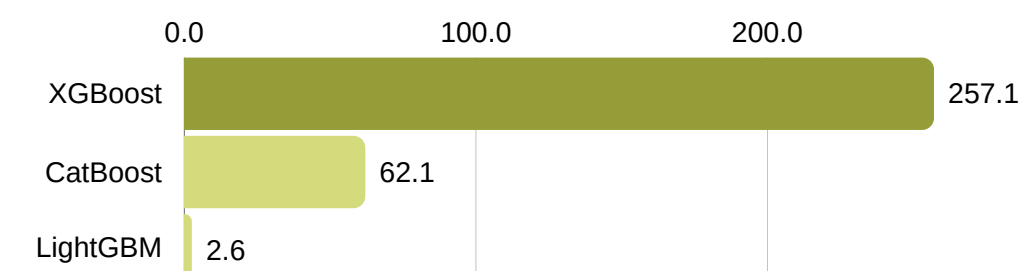


Обучение моделей регрессии

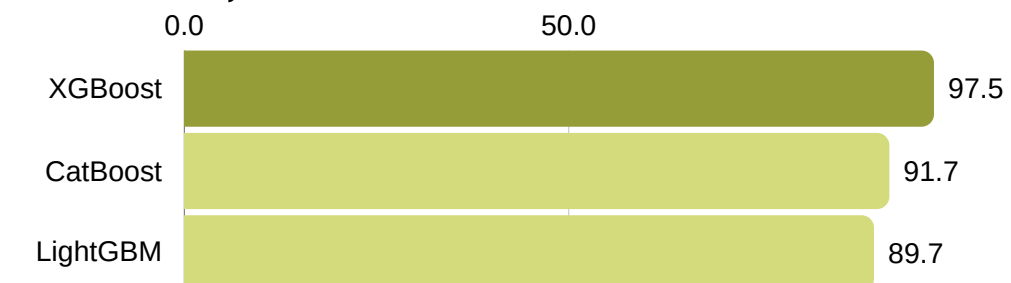


Для обучения моделей использовалась библиотека Optuna для подбора параметров

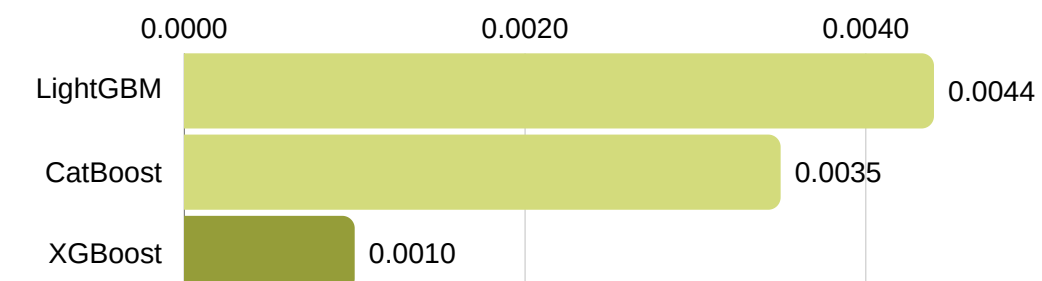
Время обучения, с

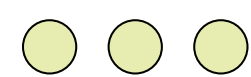


R2-score, %



MSE





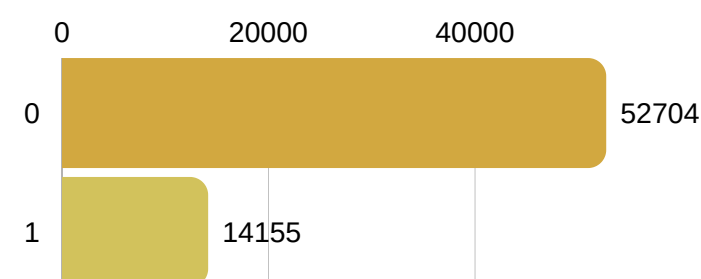
Обучение моделей



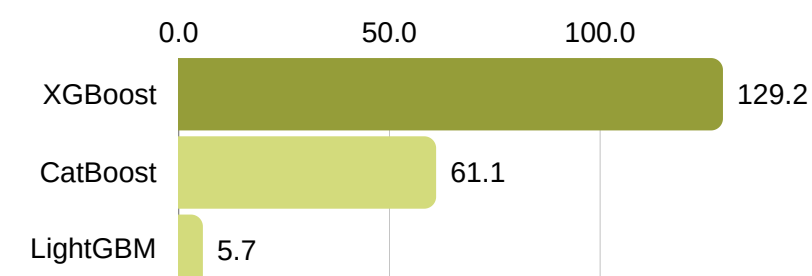
Обучение моделей классификации



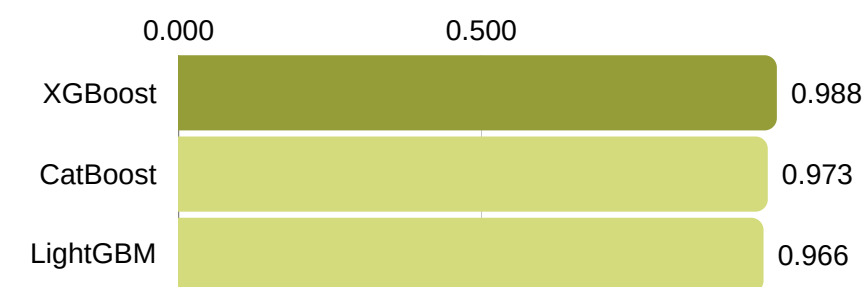
Баланс классов



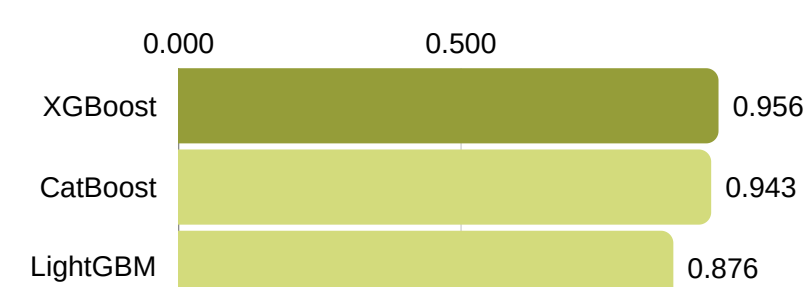
Время обучения, с



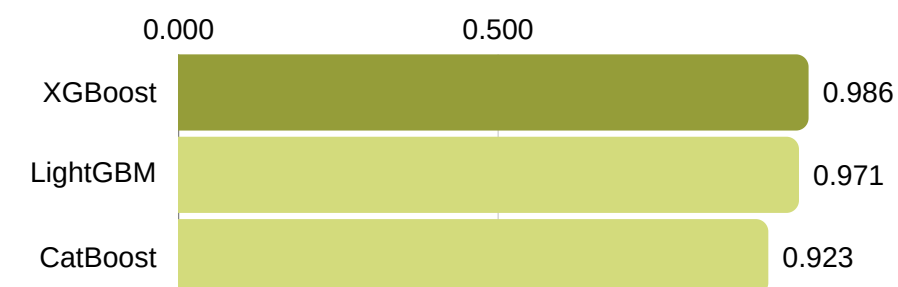
Accuracy



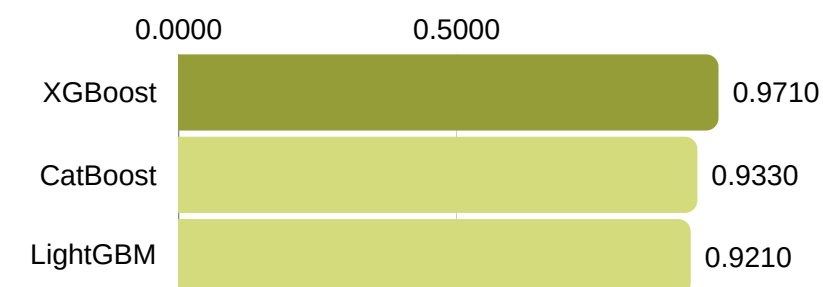
Precision



Recall



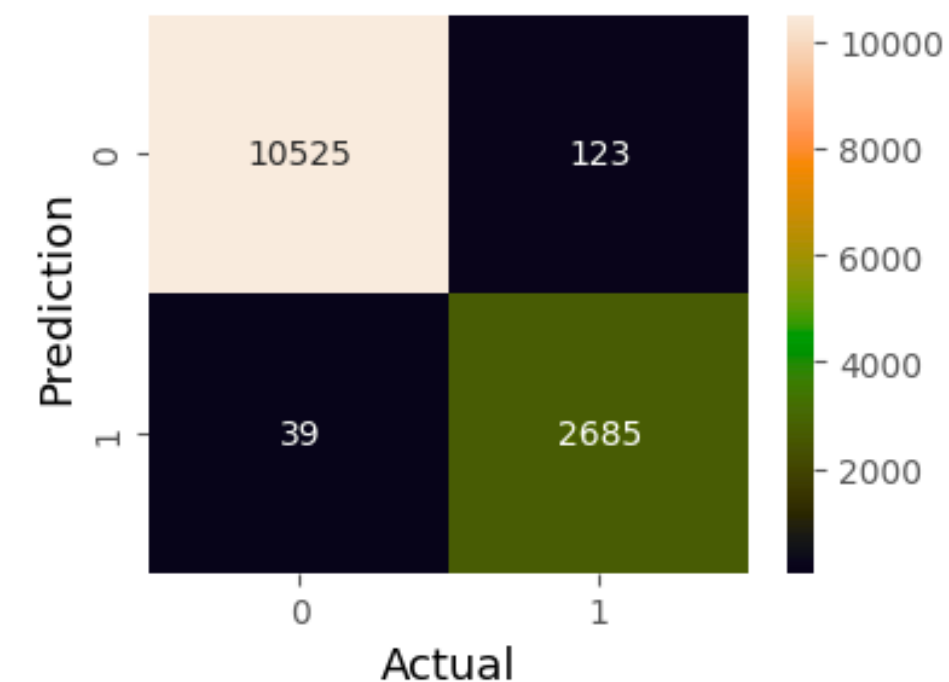
F1-score



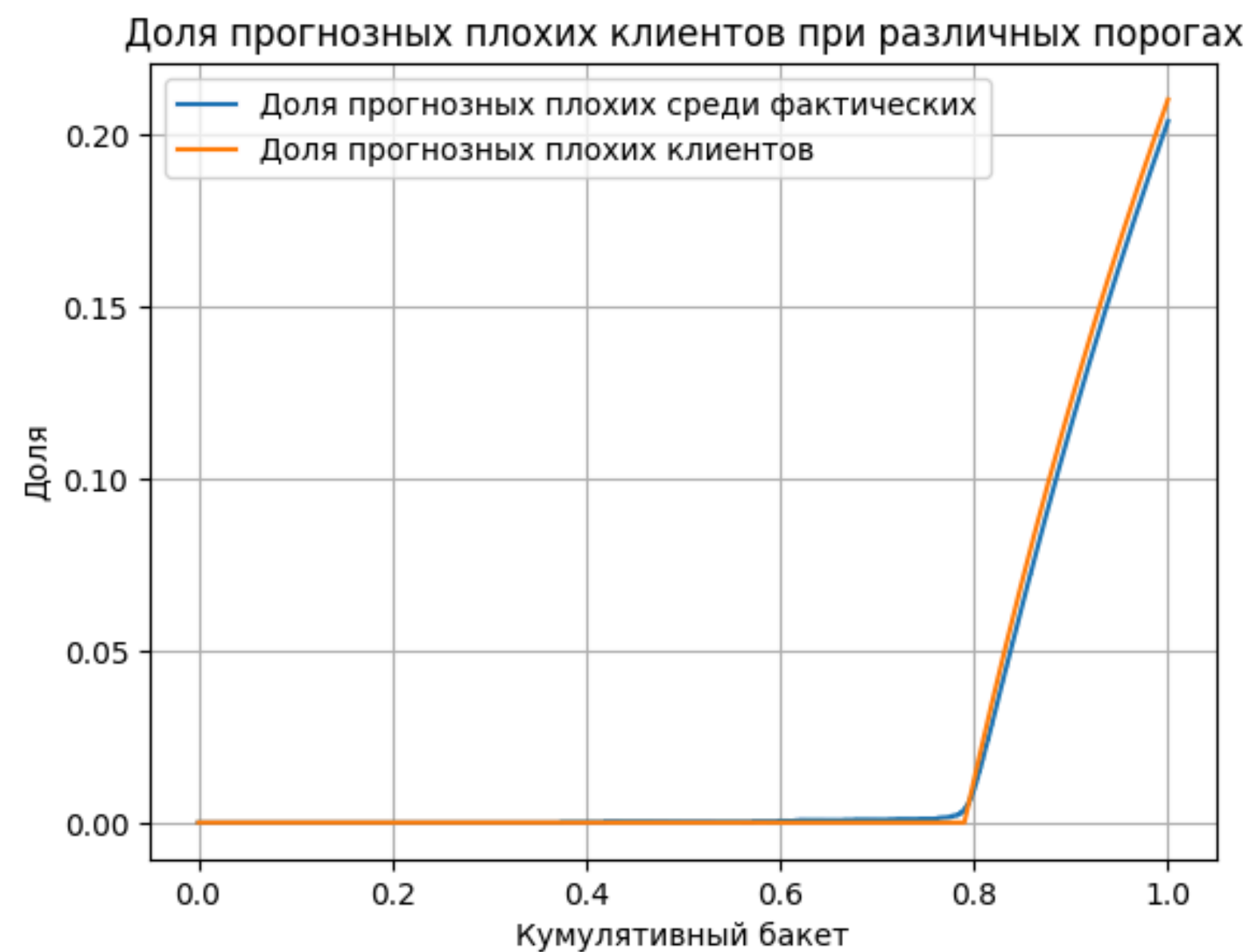
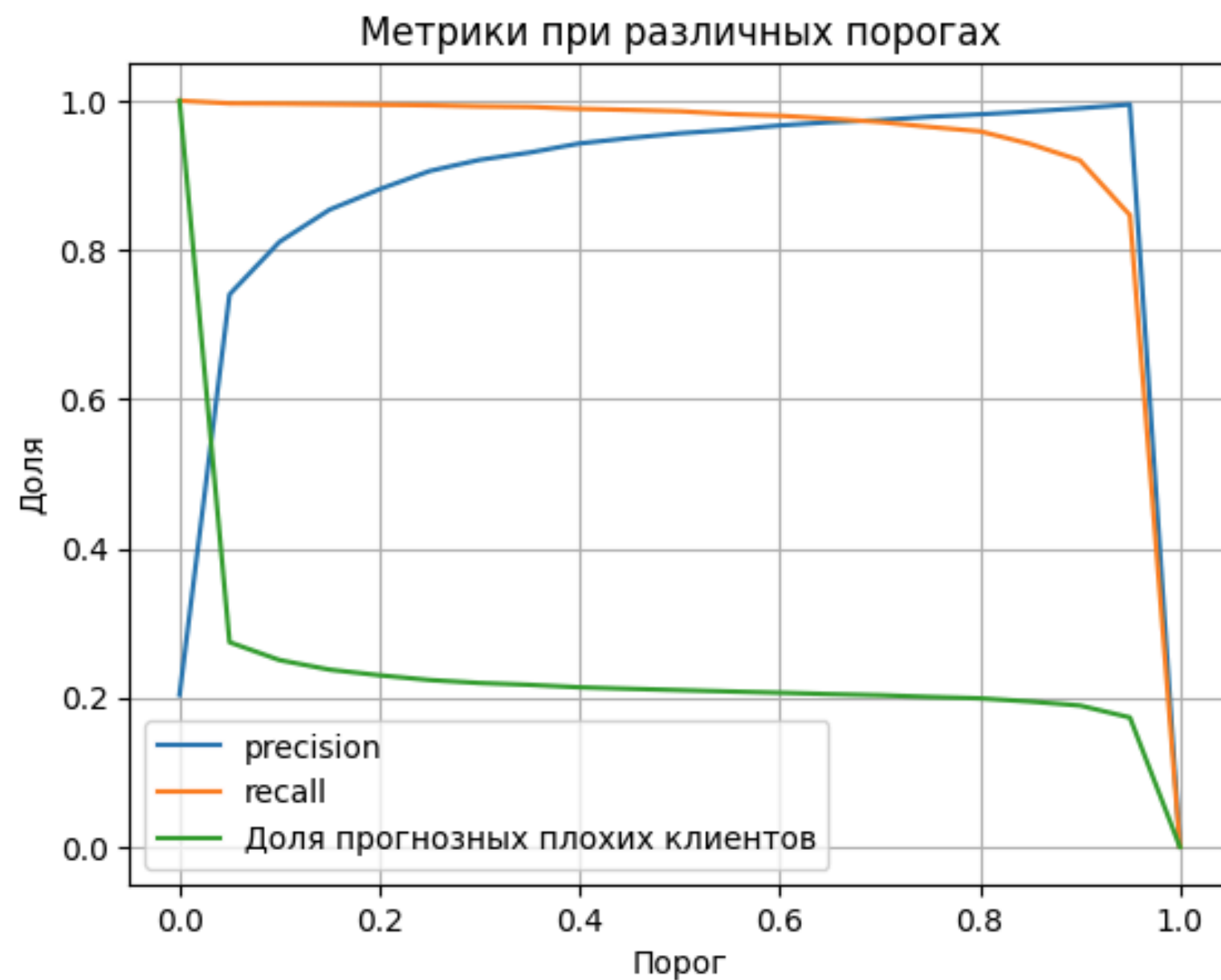
Согласно матрице на модели **XGBoost**, ошибок первого рода меньше, чем ошибок второго рода.

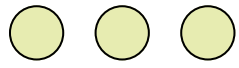


Confusion Matrix XGB



Оценка метрик





Спасибо за внимание

