Презентация итоговой работы

“Модель оценки кредитного риска - выход клиента в дефолт по кредиту”

Выполнил: Валиев Р.И.
специализация Machine Learning Engineer

Цели и задачи

Цель

- Создание модели оценки кредитного риска – предсказание выхода клиента в дефолт по кредиту (ROC-AUC 0,75)

Задачи

- Изучение датасета
- Подготовка данных, разработка фич
- Обучение моделей и выбор лучшей
- Внедрение модели в пайплайн

Изучение датасета

train_data.zip

- Данные о заемщиках и их кредитной истории
- Одна запись = один конкретный кредитный продукт, выданный конкретному заёмщику
- Данные бинаризованы

train_target.csv

- Одна запись = один конкретный заёмщик
- «flag» - Целевая переменная, 1 – факт ухода в дефолт.

Далее, в целях обучения модели с “учителем” по атрибуту **id** объединяем исследуемые таблицы

Подготовка данных

Этапы

- Удаление неинформативных столбцов
- OHE-кодирование
- Агрегация по ID
- Удаление дубликатов
- Полное удаление дубликатов с разной целевой переменной

Создание фич

- Создание новых фич затруднено, данные бинаризованы случайным образом

Далее, данные в целях обучения разделены на тренировочную и тестовую выборки в соотношении 80/20 со стратификацией по целевой переменной

Обучение моделей

Gradient Boosting Classifier

- ROC-AUC 0,743
- Recall 0,00

MLPClassifier

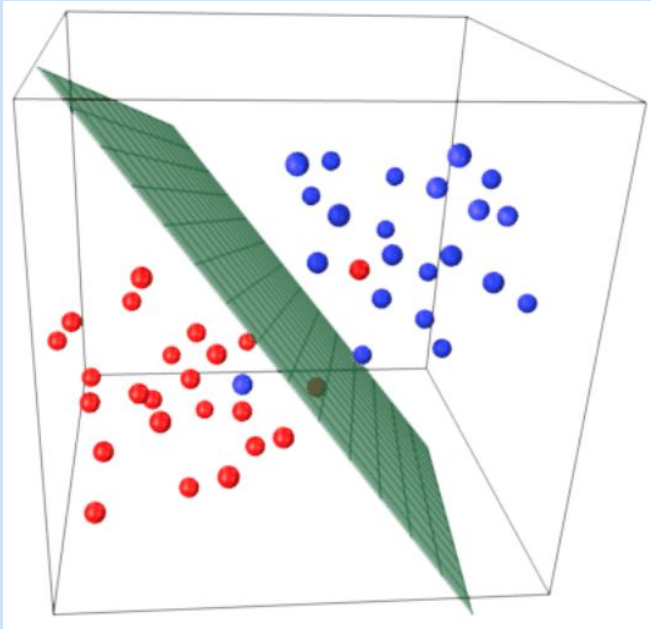
- ROC-AUC 0,741
- Recall 0,00

CatBoostClassifier

- ROC-AUC 0,754
- Recall 0,698

LGBMClassifier

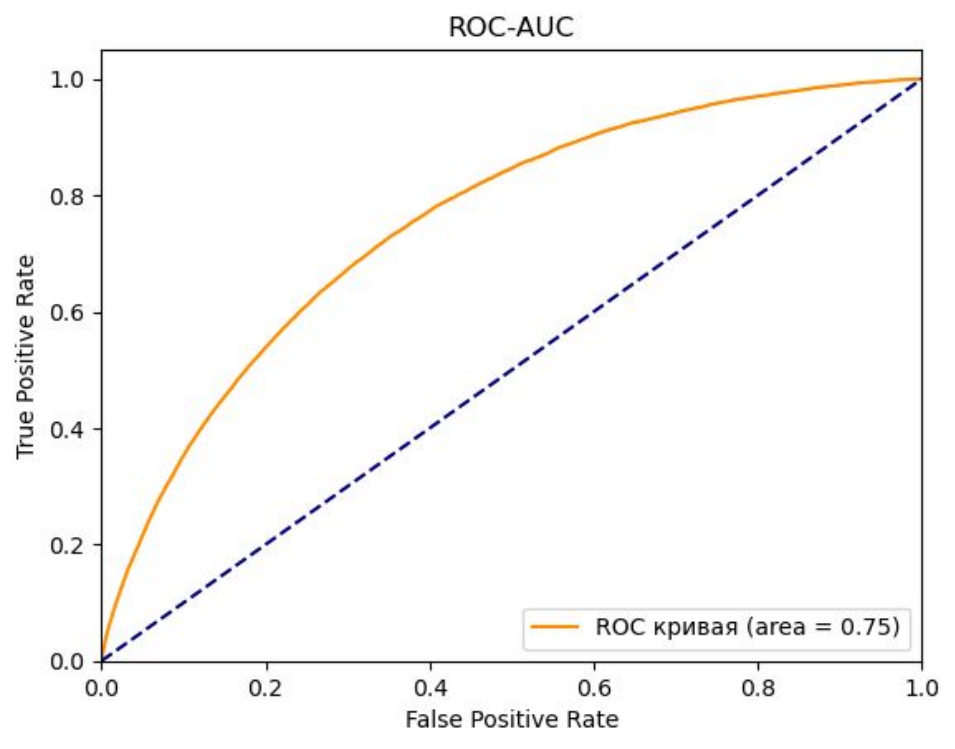
- ROC-AUC 0,753
- Recall 0,714



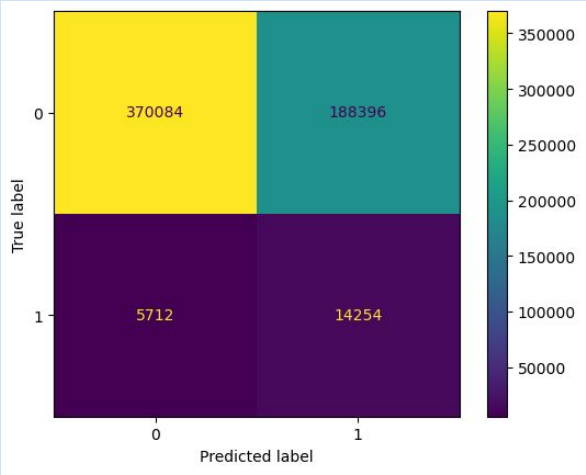
Исходя из необходимости предсказания риска дефолта и минимизации ошибок первого рода в качестве наилучшей модели, для дальнейшего внедрения выбрана **LGBMClassifier***

*подбор гиперпараметров не оказал влияния на ключевые метрики

ROC-AUC LGBMClassifier test sample



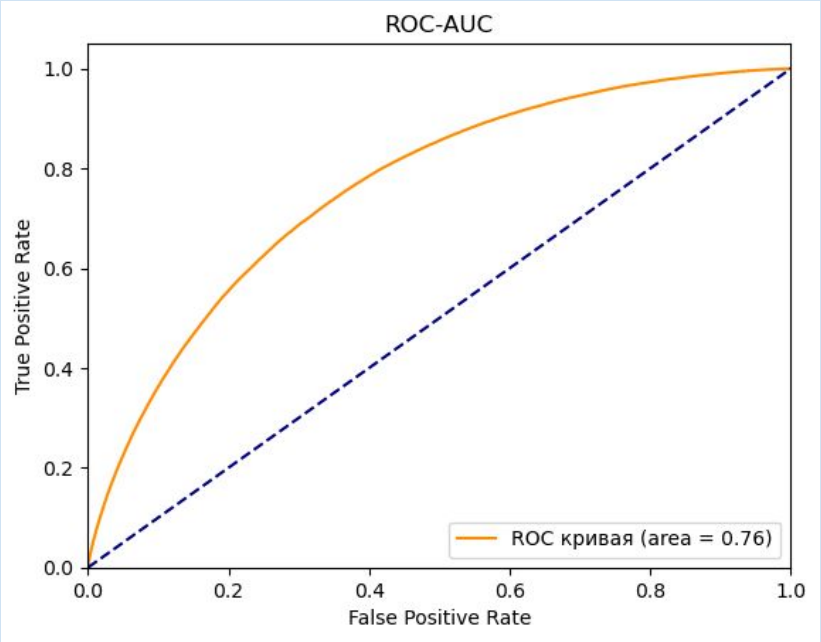
Confusion matrix



Внедрение модели

С помощью Sklearn.Pipeline реализован пайплаин:

- по вызову fit подготавливаются данные, обучается модель
- по вызову predict пайплайн делает предсказание на заданном наборе данных



LGBMClassifier Pipeline**:**

•ROC-AUC 0,761

•Recall 0,726