МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ

(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)

Институт №8 «Информационные технологии и прикладная математика»

Кафедра 810 «Информационные технологии в моделировании и управлении»

**Лабораторная работа №4**

**по курсу «Интеллектуальный анализ данных»**

**Градиентный бустинг.**

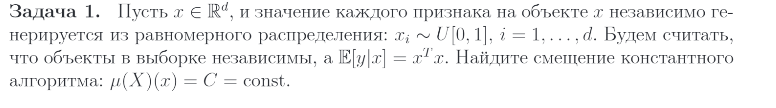
Выполнил: А.С.Бобряков

Группа: М8O-103М-19

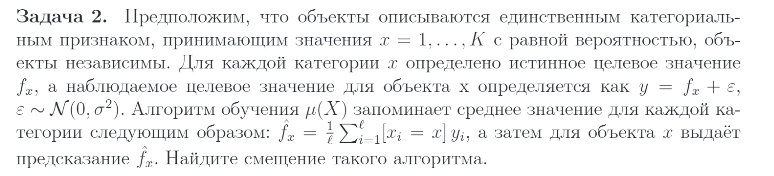
Преподаватель: К.К.Абгарян

Москва, 2020

**Задание:**



**Решение:**



**Решение:**

Лабораторная работа

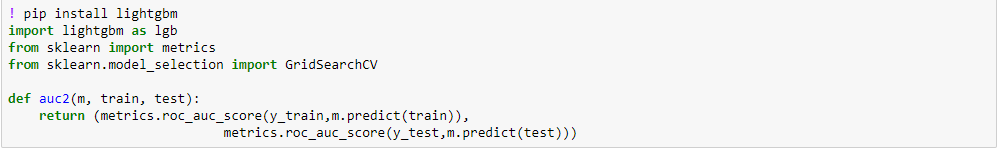
**Задание:**

Обучить реализации градиентного бустинга LightGBM и Catboost на вещественных признаках. Подобрать оптимальные гиперпараметры: глубину деревьев, количество деревьев, темп обучения. Визуализировать важности признаков, посчитанные этими алгоритмами.

**Решение:**

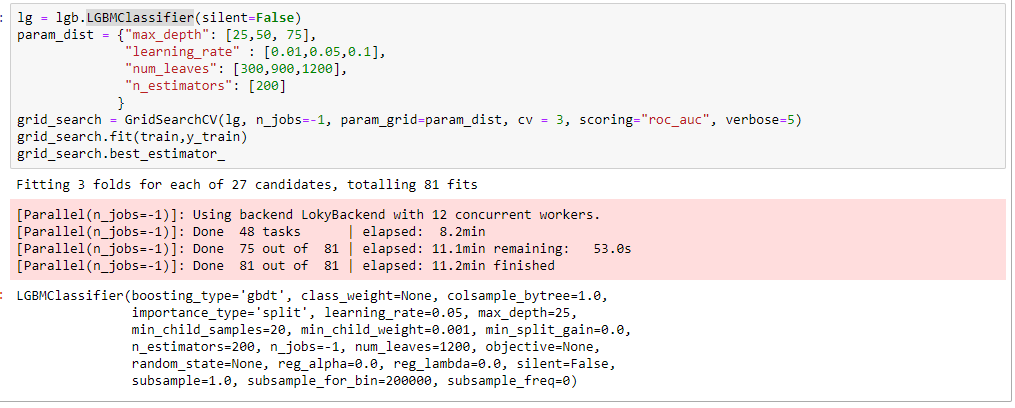
Данные: 

Метрика качества – AUC-ROC:

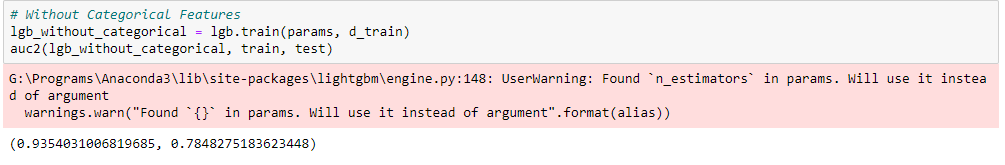


LGBMClassifier

Ищем оптимальные параметры для LGBMClassifier:

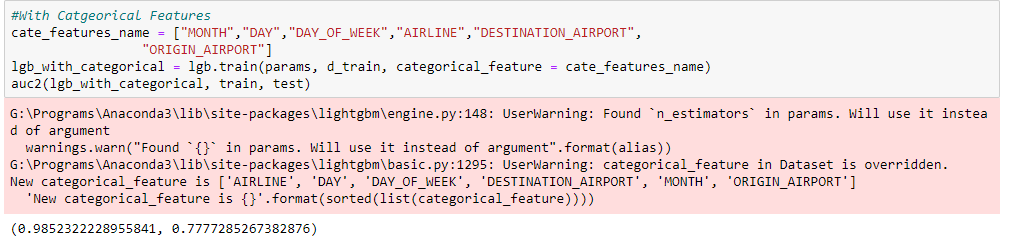


Обучаем LGBMClassifier с оптимальными параметрами БЕЗ категориальных признаков:



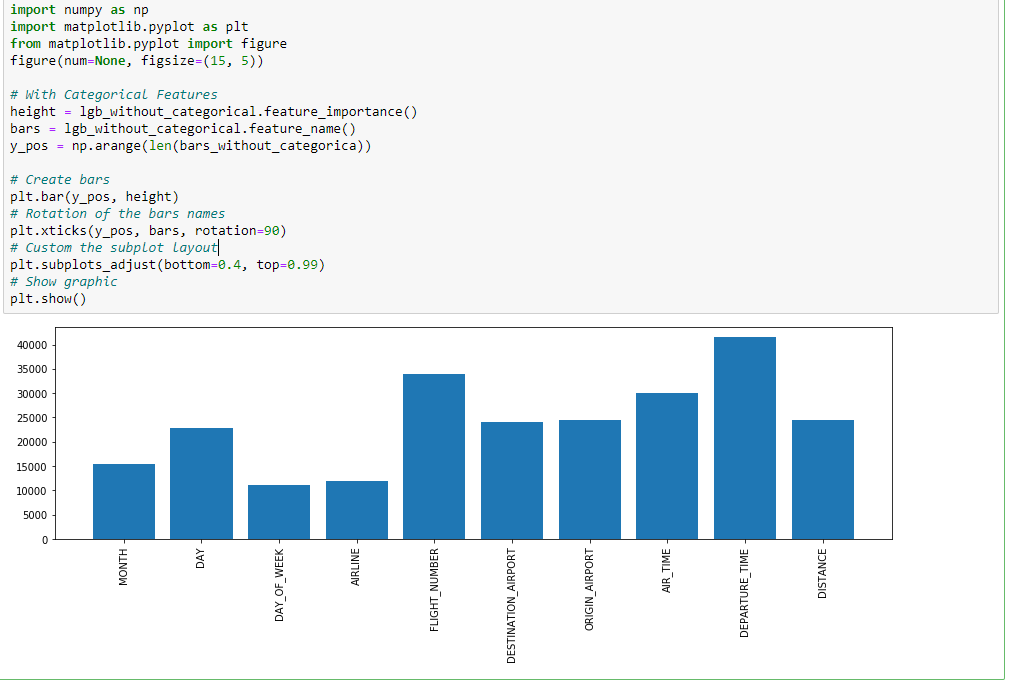
AUC-ROC на тренировочном наборе достиг 0.935. На тестовом – 0.785

Обучаем LGBMClassifier с оптимальными параметрами С категориальными признаками:



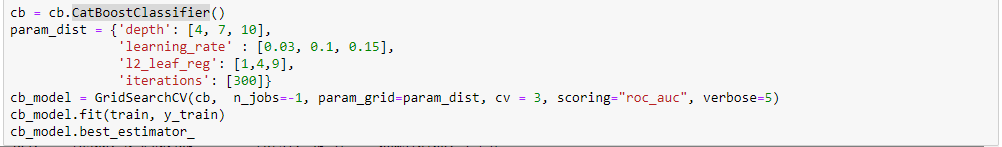
AUC-ROC на тренировочном наборе достиг 0.985. На тестовом – 0.778

Важность признаков LGBMClassifier:

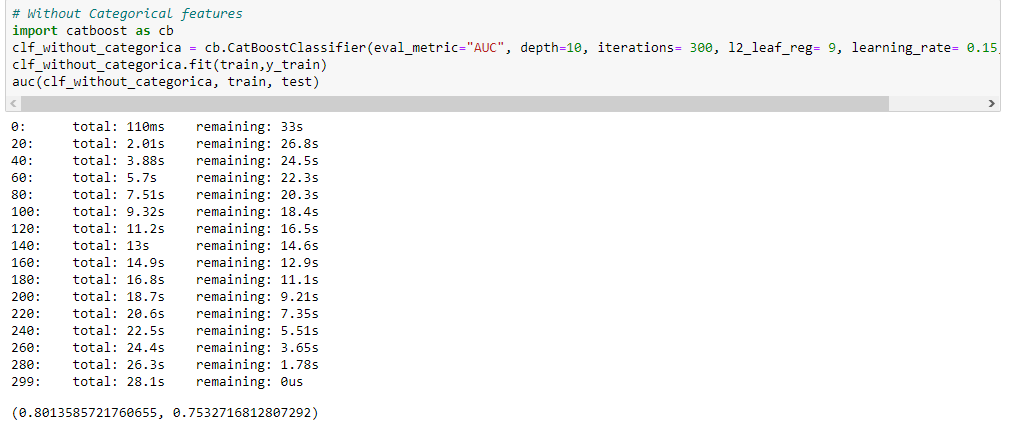


CatBoostClassifier

Ищем оптимальные параметры для CatBoostClassifier:

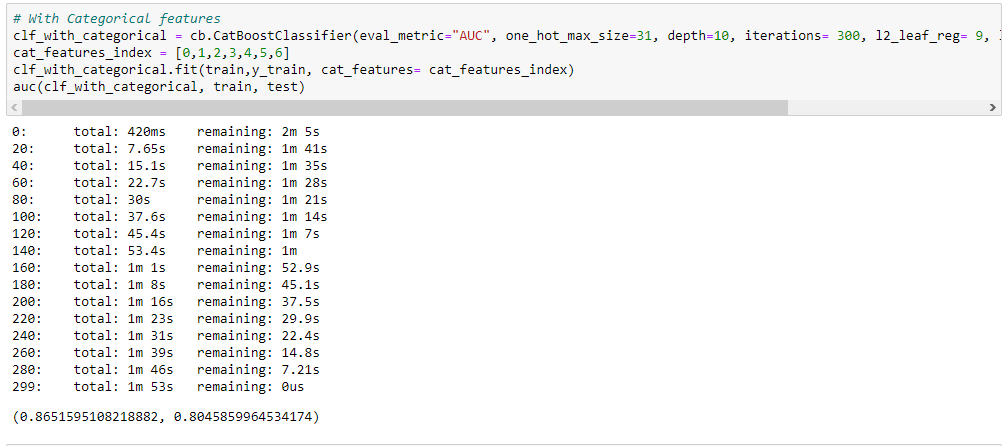
 

Обучаем CatBoostClassifier с оптимальными параметрами БЕЗ категориальных признаков:



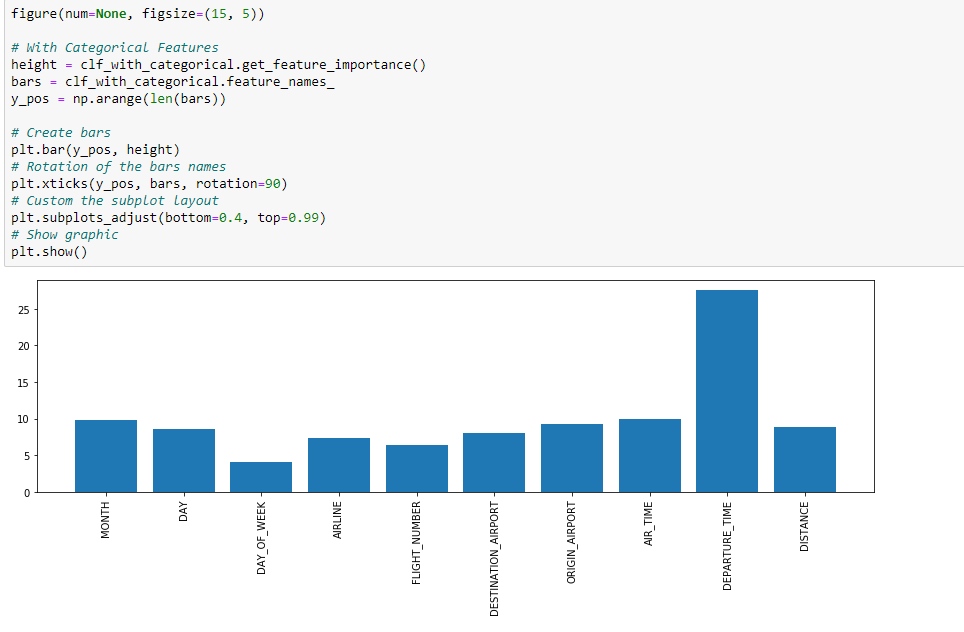
AUC-ROC на тренировочном наборе достиг 0.801. На тестовом – 0.753

Обучаем CatBoostClassifier с оптимальными параметрами С категориальными признаками:



AUC-ROC на тренировочном наборе достиг 0.865. На тестовом – 0.805

Важность признаков CatBoostClassifier:



Выводы

В ходе лабораторной работы были изучены методы градиентного бустинга LightGBM и Catboost на вещественных и категориальных признаках. Были визуализированы важности признаков, а также произведено сравнение качества через AUC-ROC для моделей с и без категориальных признаков.