Курсовой проект Megafon

Андрей Баталов

Часть 1. Исследование и предварительная подготовка данных

В качестве исходных данных предоставлена информация об отклике абонентов на предложение о подключении одной из услуг. Данные train и test разбиты по периодам: на train доступно 4 месяца, а на test отложен последующий месяц.

В данных нет пропущенных значений.

Наблюдается значительный дисбаланс классов целевой переменной.

Features

Данные представляют собой отдельный нормализованный анонимизированный набор признаков, характеризующий профиль потребления абонента.

В данных также отсутствуют пропуски, однако работа с файлом осложняется его размером - файл содержит 4 512 528 значений и в распакованном виде занимает 20 гб пространства.

В виду большого количества данных в датасете отфильтруем бесполезные для нас и затем объединим с датасетами train и test по признаку id.

Объединенный train разделим на тренировочную и валидационную части для обучения модели и оценки результатов.

Часть 2. Выбор модели, обучение и оценка.

В качестве модели возьмем CatBoostClassifier. Эта бустинговая модель хорошо зарекомендовала себя при работе с разнородными данными, в том числе категориальными, умеет обрабатывать возможные пропуски и работать с несбаласированными классами, что актуально в нашем случае. Для этого мы применим параметр auto_class_weights='Balanced'.

Скоринг работы модели будем осуществлять при помощи функции f1, для наглядности и более полной информации применим classification_report.

Полученные результаты и идеи для дальнейших исследований

Результат по метрике f1_score(average-’macro’) получили равным 0.715.

Для улучшения модели можно попробовать рекомендовать клиентам услуги:
-из топ-популярных;
-услуги, приобретенные схожими клиентами;
-комбинировать рекомендации со скидками на определенный период (например на месяц).
Значительное число клиентов имеют 2 и более sim-карт. Вероятно ими пользуются родственники или члены семьи. Можно делать персональные предложения для каждой такой sim-карты, даже если id принадлежит одному клиенту.

Благодарю за внимание

Андрей Баталов