

Отчет об соревновании Sberbank Data Science Journey

Дмитриев Даниил
ocksumoron(DMIA)

Решение

Это моё первое нормальное соревнование, так что ничего сложного я не использовал.

Я писал решение только для первой задачи, со второй как-то не пошло :(

Были генерированы фичи и на них построен xgboost, параметры которого подбирались вручную.

ФиЧи

- По mss_code и tr_type: для каждого потребителя посчитал количество использований каждого из типов
- Для amount я считал отдельно для трат и доходов следующие статистики - минимум, максимум, медиану, среднее арифметическое и сумму
- В начале я отправил такое решение и оно получило 0.854545 ROC-AUC

Новая фишка

- Появилась идея посмотреть на периодичные траты и доходы одного размера - то есть либо выплаты за кредиты, либо зарплата. Так как в семье обычно за кредиты платит мужчина, и зарплата в среднем у мужчин больше, то это должен получиться хороший разделяющий признак.
- Эта гипотеза была проверена на трейне, визуализации были отправлены в канал на слаке
- С этими двумя признаками и вручную настроенным xgboost'ом решение набирало 0.885206 ROC-AUC на прайват лидерборде