Решение задачи "Прогнозирование вероятности невозврата кредита"

Амир Мирас

МГУ имени М.В. Ломоносова

15 января 2018 г.

Основная идея

1. Сделать датасет **кредит** $\times$ **признак**
2. Для каждого клиента агрегировать статистики по признакам его кредитов
3. Обучить на полученных признаках модель

Обработка признаков

1. Признаки с датой: добавим всевозвожные разности, выделим день недели, месяц, день
2. PAYMENT_DISCIPLINE_REVERSE: *TFIDF* с $ngram = (1, 2)$
3. Категориальные признаки: добавим счетчики
4. Некоторые осмысленные признаки: $CREDIT\_SUM - CREDIT\_SUM\_DEBT, CREDIT\_SUM\_OVERDUE - CREDIT\_MAX\_OVERDUE$ и т.д.

Итоговая модель

1. Для каждого клиента будем агрегировать следующие статистики: среднее, медина, минимум, максимум, стандартное отклонение
2. На полученных признаках обучим LGBMClassifier

Результат на private leaderboard: **0.70334** (21 место)