## Отчет о проделанной работе

Разработка модели для определения пользователей склонных к оттоку

## Описание проблемы

Для В2С телеком компании необходимо идентифицировать пользователей, склонных к оттоку, и своевременно удержать таких пользователя от ухода. Для этого надо придумать способ идентификации отточных пользователей.

## Цель проекта

Создать модель машинного обучения, которая будет способна находить пользователей, склонных к оттоку.

## Формулировка задачи машинного обучения

• На основании выборки данных клиентов (набор признаков для каждого клиента) телеком компании с бинарным откликом(отток, не отток) построить модель машинного обучения, которая сможет предсказывать вероятность ухода клиента.

## Особенности данных

- Несбалансированная выборка 93% пользователей, не склонных к оттоку, 7% пользователей, склонных к оттоку
- Большое признаковое пространство 230 признаков
- Численные признаки плохо линейно разделимы
- Распределения значений признаков для разных классов примерно одинаковы

# Способы измерения качества и критерии успеха

Так как, выборка несбалансированная, нужно оптимизировать модель так, чтобы повысить количество верно предсказываемых отточных клиентов, при этом также минимизировать количество FP объектов. Наиболее подходящими для этого метриками будут *Roc-Auc, Recall, Precision*. Ключевой метрикой для экономической модели будет **Precision**.

#### Метрики финальной модели

Roc-Auc = 0.68

Recall = 0.734

Precision = 0.142

#### Таблица контингентности

	A = 1	A = -1
Y = 1	4624	2755
Y = -1	165	456

### Решение задачи

Выбор модели машинного обучения — Gradient Boosting classifier



Подбор оптимального количества объектов для обучения – **15к** 



Отбор категориальных признаков — наиболее коррелирующие с целевой переменной



Обработка численных признаков стандартизация признаков



T

Борьба с дисбалансом классов – **балансировка весов целевой переменной** 



Подбор оптимальных параметров для модели для обеспечения наивысшего качества модели – **Grid Search** 



Заполнение NaNзначений — **нулями** 



Выбор подходящего Кодировщика для категориальных переменных – **Label Encoder** 

## Экономический эффект от внедрения модели

#### • Параметры модели

- Кол-во денег, которое приносит пользователь в месяц - CustomerYield(CY)
- Кол-во денег, которое мы будем вкладывать в его удержание - OnCustomerSpendings(OCS)
- Какова вероятность, что предложение будет принято им - ProbabilityOfAcceptance(PoA)
- Клиенты, на которых будет рассчитана компания по удержанию ModelChurnCustomers(MCC)
- Компания рассчитана на пользователей склонных к оттоку, которых верно определила модель ActualChurnCustomers(ACC) = MCC \*
   Model Precision

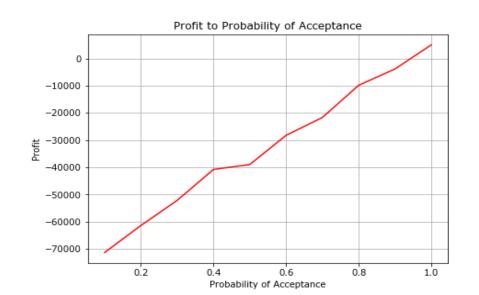
Экономическая модель:

Profit = PoA \* ACC \* (CY + N(CY, $\sqrt{CY}$ )) – OCS \* MCC

<u>Начальные параметры модели:</u>

CY = 1000; MCC = 3211;

Model Precision = 0.14; OCS = 25; PaA = 0.6



### Заключение



#### Решение



- Идентифицировать клиентов склонных к оттоку
- Построить модель машинного обучения, предсказывающую вероятность оттока клиента

- Построен классификатор на основе Градиентного бустинга с оптимальными параметрами
- Качество детектирования отточных пользователей с использованием модели повысилось в 2 раза

- Построена экономическая модель, описывающая прибыль от внедрения модели машинного обучения
- При стандартных параметрах экономической модели, внедрение модели МЛ снижает расходы на 60%, а прибыль возрастает на 40%