

# Digital Reputation Challenge

Яворская Мария

# Данные

Описание пользователей:

- $X_1$  – количественные / бинарные / категориальные признаки
- $X_2$  – категориальные признаки
- $X_3$  – количественные признаки

Целевая переменная:

- $Y$  – 5 бинарных величин – качества пользователей

# Первое решение

- Данные: матрица  $X_1$
- Алгоритм: Random Forest
- Качество:
  - На отложенной выборке: 0.5898
  - pub: 0.5875
  - priv: 0.5832

# Второе решение

- Данные: матрица  $X_1$  + часть матрицы  $X_2$ 
  - Из матрицы  $X_2$  были выбраны значения, которые чаще встречались среди пользователей
- Алгоритм: Random Forest
- Качество:
  - На отложенной выборке: 0.6020
  - pub: 0.5864
  - priv: 0.5823