

# Digital Reputation Challenge

Сомов Иван  
МГУ им. Ломоносова, ВМК, ММП

# Данные

Три таблицы с неизвестными столбцами (кроме user id):

- X1 – какие-то агрегаты по поведению пользователей
- X2 – факты посещения сайтов пользователями (гипотеза)
- X3 – по версии организаторов, какие-то агрегаты по X2

# Целевая переменная

Психотипы человека (гипотеза):

- Открытость
- Сознательность
- Экстраверсия
- Сотрудничество
- Невротизм

The slide features a header with the Mail.ru group logo, a Russian crown icon, and the text 'MOSCOW DATA SCIENCE MAJOR'. Below the header, there's a section titled 'Эмоциональная стабильность' (Emotional stability) with a photo of a woman speaking at a podium. To the right, there's a section titled 'Big Five модель' (Big Five model) with a table describing the five traits:

Trait	Description
Openness – открытость опыта	Восприятию нового, изучение чего-то нового, предрасположенность к фантазиям, креативность, воображение.
Conscientiousness – сознательность	Включает надежность, ответственность, скрупулезность в работе, дисциплинированность, самоконтроль, организованность.
Extroversion – экстраверсия	Коммуникабельность, общительность, разговорчивость, стремление к компании и т.д.
Agreeableness – сотрудничество	Характеризует то, приятна ли личность в общении, старается ли найти общий язык с другими или, напротив, грубит, отталкивает, обосабливается.
Emotional Stability – эмоциональная стабильность	Показатель эмоциональной стабильности/нестабильности, чувствительности, нервозности, беспокойности, раздражительности.

Below the table, there are small icons of people engaged in various activities like reading, writing, and socializing. At the bottom left is the name 'Ирина Деева' and a footer note: 'Вычислительное прогнозирование психометрик пользователя на основе его цифрового следа в социальных сетях'.

# Еще один инсайт

Опросник BigFive <https://psytests.org/bigfive/5pq.html>

Отечественная адаптация методики выполнена **А. Б. Хромовым** в Курганском Государственном Университете в 2000 году:

- 75 вопросов
- Шкала ответов от -2 до 2
- Шкала результатов от 1 до 5



Люди считают меня отзывчивым и доброжелательным человеком

-2      -1

0

1      2

Некоторые люди считают меня холодным и черствым

интроверсия

пассивность

подчиненность

замкнутость

избегание впечатлений

избегание внимания

обособленность

равнодушие

соперничество

подозрительность

непонимание

самоуважение

15 20 25 30 35 40 45 50 55 60 65 70 75  
61

экстраверсия

активность

доминирование

общительность

поиск впечатлений

привлечение внимания

15 20 25 30 35 40 45 50 55 60 65 70 75  
15 15

13

8

10

15 20 25 30 35 40 45 50 55 60 65 70 75  
64

привязанность

теплота

сотрудничество

доверчивость

понимание

уважение других

14

14

11

13

12

# Первая попытка

- Baseline solution от организаторов
- Убрал X2, X3
- Оставил параметры xgb
- Результат:  
    pub: 0.581162  
    priv: 0.580663
- Подобрал параметры xgb, усреднил с предыдущем решением
- Результат:  
    pub: 0.583023  
    priv: 0.582973

# Исследования данных, отбор признаков

- Отбор признаков по feature importance не помог
- 19 признак содержал только 4 метки 1 (далее был выброшен)
- X3 вообще не давала прироста
- Визуальный анализ признаков попарно не дал результатов

# Прорыв

## Матрица X2:

- Рассматриваем как user-item матрицу X2 + X2\_test
- Получаем с помощью ALS user эмбеддинги (размерность 20)
- Результат на xgb (без подбора параметров):

pub: 0.601319  
priv: 0.60527

# Стекинг

- Взял user эмбеддинги от ALS с разными размерностями
- Подобрал параметры для XGBoost, Random Forrest, FF-NN, SVM, Logistic Regression
- Мета-классификатор – Logistic Regression
- Результат:  
**pub:** 0.610988  
**priv:** 0.612616

# ушел не туда...

- UMAP с cosine-расстоянием для X2 + X2\_test
- Добавил crossval-predict ответы к ответам с ALS
- Финальный результат  (19 место): **pub: 0.612335**  
**priv: 0.612814**