



Кейс VK Машинное обучение на графах

Предсказание интенсивности взаимодействия между друзьями в социальной сети ВКонтакте

цифровой
прорыв 

сезон: III

Команда NETWORK

профи

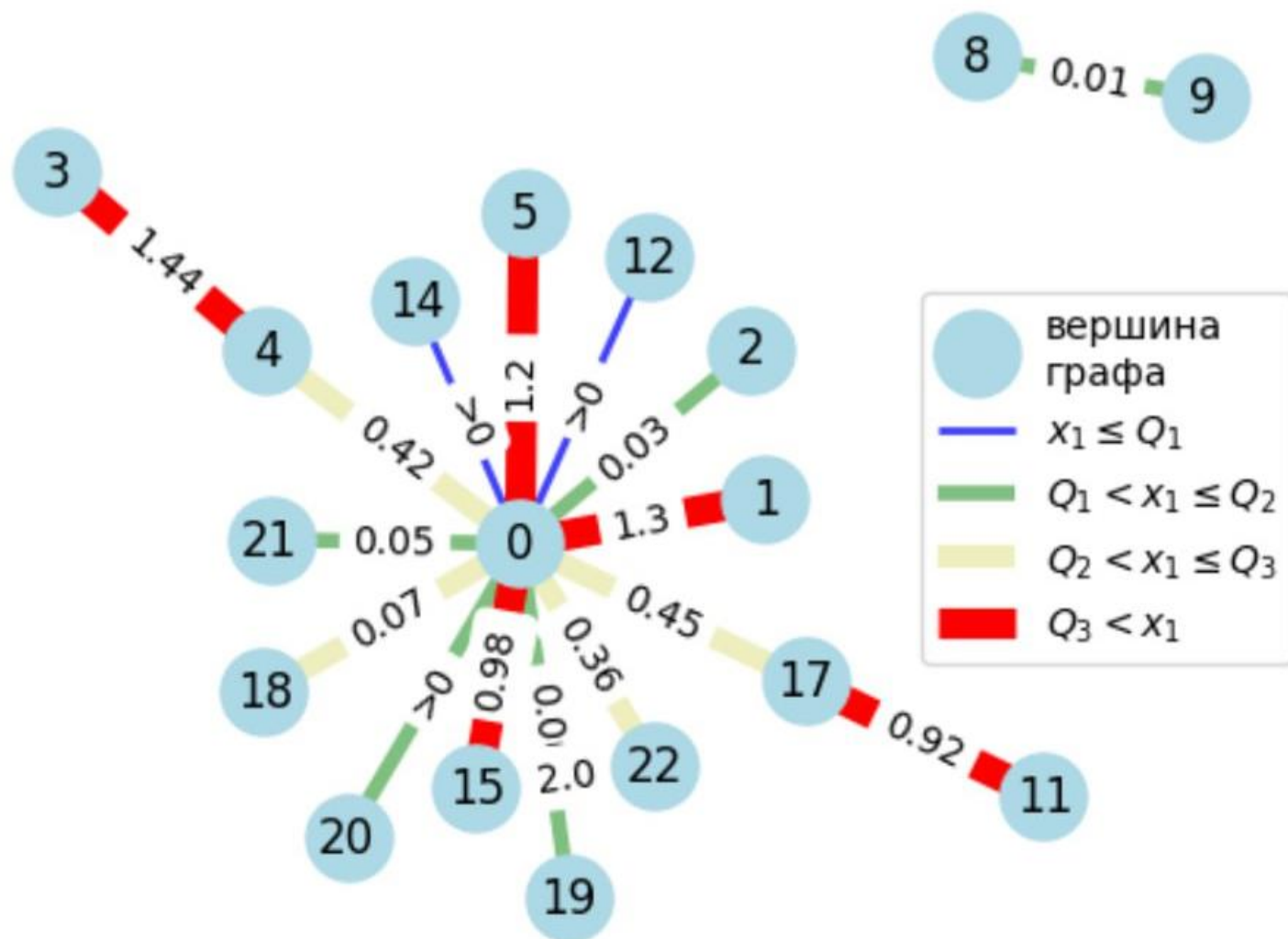
Проблематика

Качество социального взаимодействия между пользователями ВКонтакте – почти всей аудиторией рунета

- Задача регрессии



Типичный граф для конкретного ego_id



Толщина линий рёбер графа пропорциональна коэффициенту активности взаимодействия между пользователями x_1 .

Также применена цветовая градация рёбер по возрастанию активности: синий, зеленый, желтый, красный.

01 Базовая модель



CatBoost



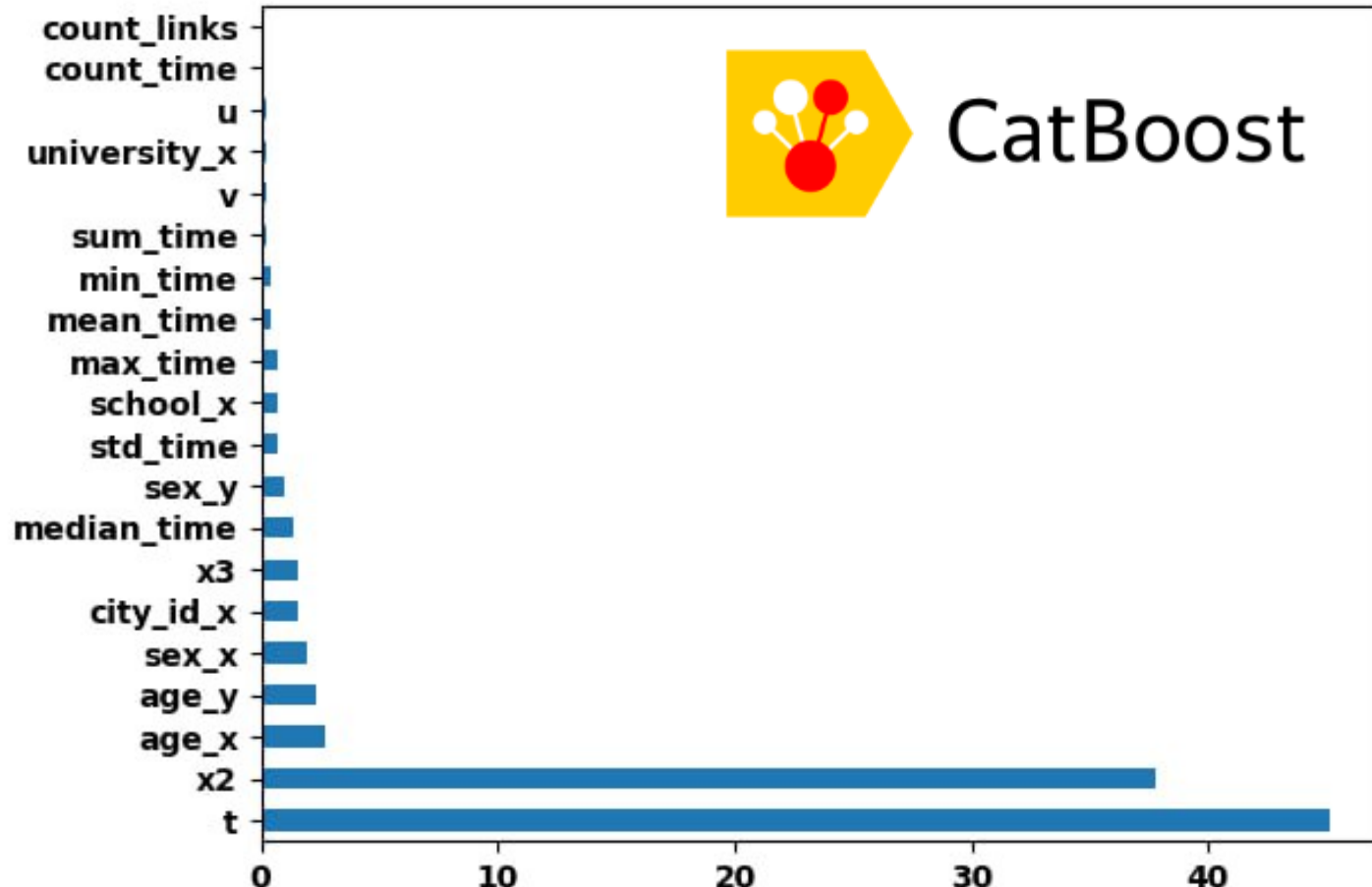
Базовая модель
была улучшена
путем генерации
НОВЫХ И
дополнительных
признаков

- 'time_sum_mean' - суммарное время эго-графа
- 'time_std' - стандартное отклонение эго-графа
- 'count' - количество связей в эго-графе
- 'mean' - среднее количество связей в эго-графе
- 'median' - медиану связей в эго-графе
- 'count_mean' - среднее количество связей в эго-графе
- 'time_sum_mean' - среднее количество связей в эго-графе на единицу времени

02 Важность признаков

cat_features

- Sex
- City
- School
- University



02 История обучения

CatBoostRegressor()

```
test_data['x1'], feat_importances = predict_intensity(df_train, test_data, model)
```

```
Fold 1
Learning rate set to 0.407958
0:      learn: 1.0757995      test: 1.0758168 best: 1.0758168 (0)      total: 50.6s      remaining: 5h 36m 17s
100:    learn: 0.7487329      test: 0.7440791 best: 0.7440791 (100)    total: 33m 35s    remaining: 1h 39m 27s
200:    learn: 0.7414723      test: 0.7367126 best: 0.7367126 (200)    total: 1h 5m 28s      remaining: 1h 4m 49s
300:    learn: 0.7380023      test: 0.7335903 best: 0.7335903 (300)    total: 1h 36m 33s    remaining: 31m 45s
399:    learn: 0.7354581      test: 0.7313530 best: 0.7313530 (399)    total: 2h 9m 59s      remaining: 0us

bestTest = 0.7313529907
bestIteration = 399

RMSE Fold1:0.7313529907080452
Fold 2
Learning rate set to 0.407958
0:      learn: 1.0759352      test: 1.0759148 best: 1.0759148 (0)      total: 42.3s      remaining: 4h 41m 18s
100:    learn: 0.7479546      test: 0.7433079 best: 0.7433079 (100)    total: 31m 46s    remaining: 1h 34m 5s
200:    learn: 0.7412204      test: 0.7367882 best: 0.7367882 (200)    total: 1h 3m 54s      remaining: 1h 3m 16s
300:    learn: 0.7375635      test: 0.7333799 best: 0.7333799 (300)    total: 1h 42m 50s    remaining: 33m 49s
399:    learn: 0.7354597      test: 0.7317339 best: 0.7317339 (399)    total: 2h 17m 37s    remaining: 0us

bestTest = 0.7317339399
bestIteration = 399

RMSE Fold2:0.7317339398587932
RMSE mean:0.7315434900806901
```

Обработка
категориальных
признаков

Работа с
разнородными
данными

Работа с большими
наборами данных








агрегация внутри ego
графа количество
ребер у вершин
пользователей

агрегация внутри ego
графа суммарное
время пользователя

фильтруем только
нужные данные и
удаляем дубликаты


мерджим к основным
датасетам историю
по пользователям

03 Точность работы алгоритма

Участник	Результат
1  Вячеслав Пасканов vyacheslav.paskanov	0.348
2  Илья Кулешов Anonymous	0.300
3  Evgeniy Kazenov Kazenov	0.279
4  Павел Супрун (VK test) 275696	0.278
5  Капитан LoH КАПИТАНДатаСаин	0.273
6  Влад Сорокин 381715	0.264
7  VLADISLAV BALANDA BanKhv	0.254

3 место

Кейс от VK



Evgeniy Kazenov
Kazenov

Задача A

Предсказание интенсивности взаимодействия между друзьями в социальной сети ВКонтакте


Результат

0.2790388920819469



7 место

Кейс от VK



VLADISLAV BALANDA
BanKhv

Задача A

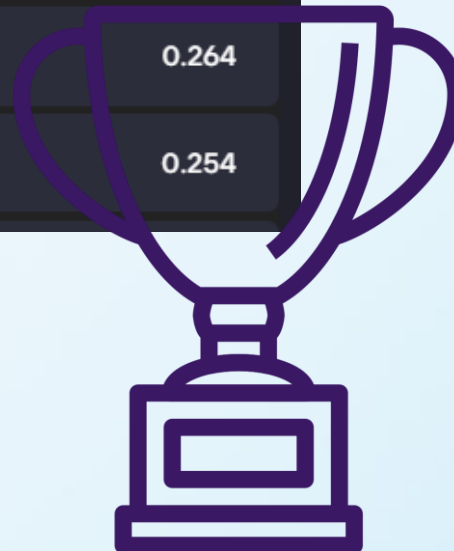
Предсказание интенсивности взаимодействия между друзьями в социальной сети ВКонтакте

Результат

0.25368002151942626



Метрика оценки – RMSE
на лидерборде (1-RMSE)



04 Адаптивность/ Масштабируемость

Используемые в работе фреймворки имеют открытый доступ, решение легко масштабируется на больших данных

Наше решение помогает пользователям находить новых друзей



Команда **профи** NETWORK



Евгений Казенов

**Стэк
ML,
Python,
NLP и др.**

@kazenovev



Дмитрий Блинов

**Стэк
Python, SQL
ML, DS**

@dima_blinov89



**Юрий
Прищепа**

**Стэк
Python, SQL
ML**

@yuprishchepa



**Владислав
Баланда**

**Стэк
Python, SQL
ML**

@Vlad2ru