

1 Введение

- Получение фич
- Создание модели
- Предсказание

2 Получение фич

Считываем данные в `pandas.DataFrame`.

lang: На основе обучающей выборки посчитаем среднее количество ретвитов для каждого значения `'user.lang'`. На основании этого получим значения для тестовой выборки. Если для какого-нибудь значения `'user.lang'` в тестовой выборке не было таковых в обучающей, то выберем в этом случае среднее средних значений по всем значениям `'user.lang'`.

text: Из текста возьмём количество '@', 'http', '#', '?'.

description: И отсюда то же, что из текста.

time zone: Аналогично `'user.lang'`.

Можно получить данные

3 Создание модели

4 Предсказание