

Видеокурс от Мегафон

Курсовой проект

Ходаковская Мария

2023

Задача:

При выполнении проекта необходимо было реализовать алгоритм, который определяет вероятность подключения услуги для пользователей.

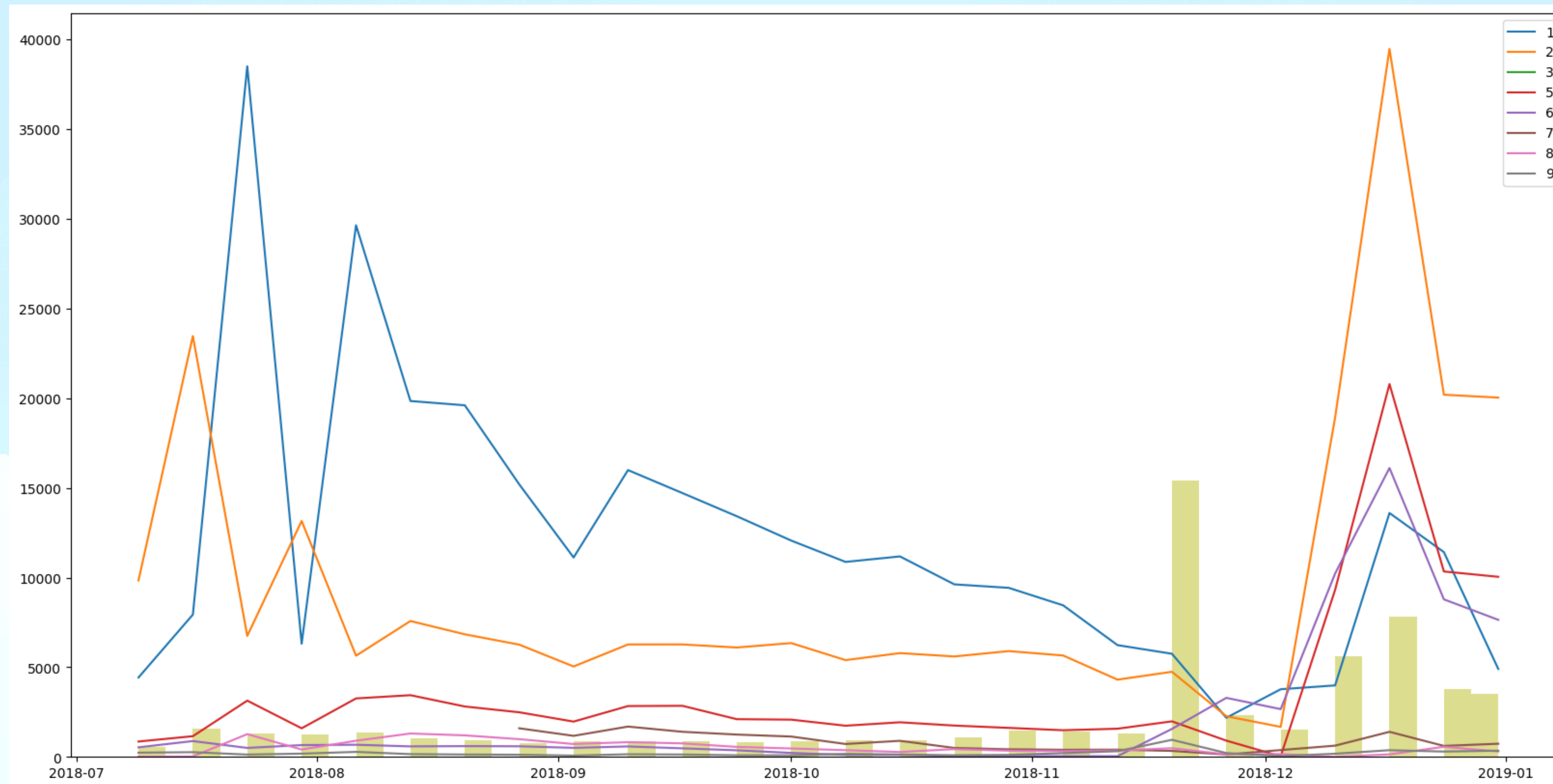
Исходные данные :

В качестве исходных данных предоставлена информация об отклике пользователей на предложение одной из услуг.

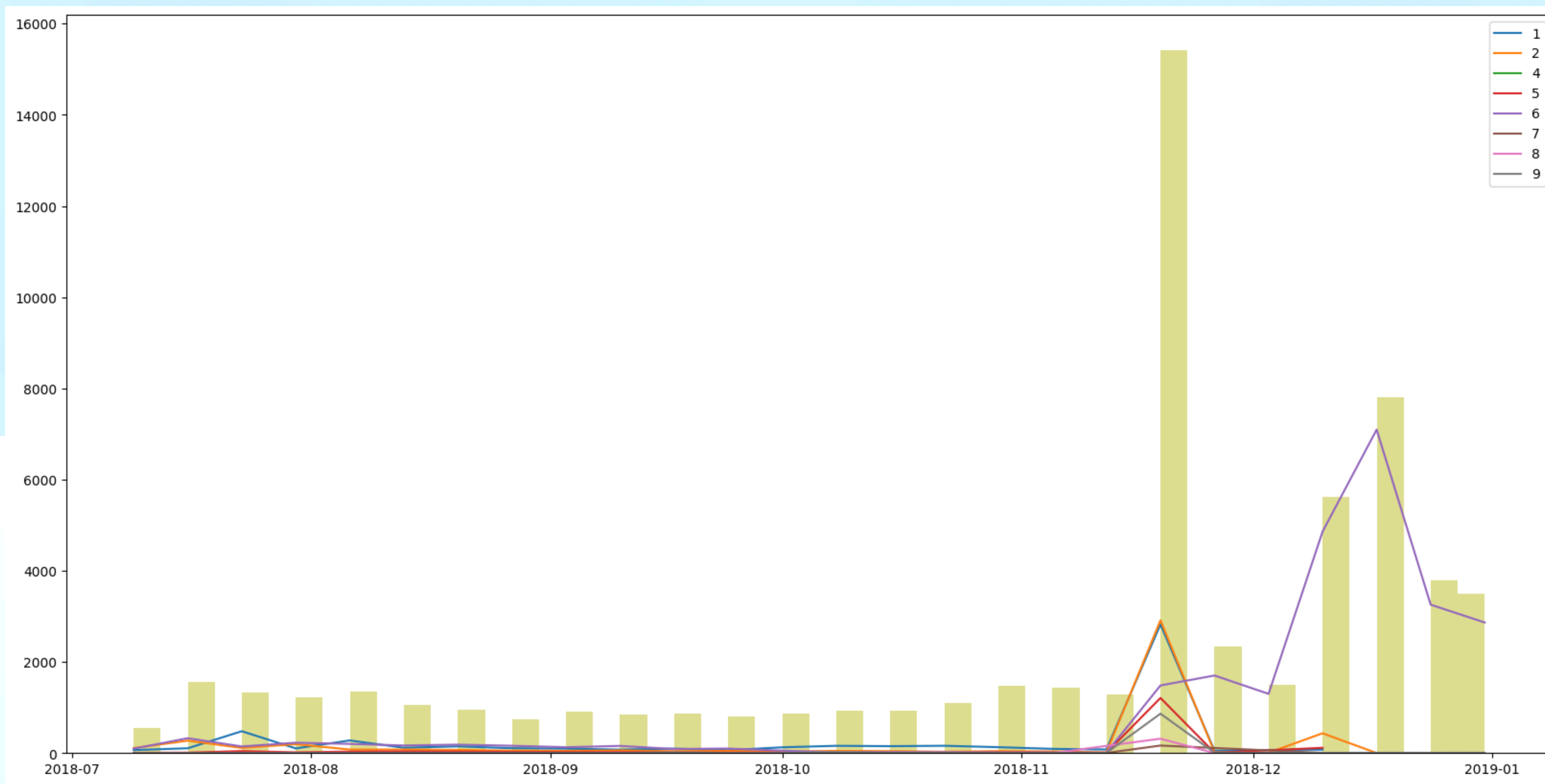
Каждому пользователю предложения о подключении услуги могут поступать несколько раз, каждое из которых он может принять или отклонить.

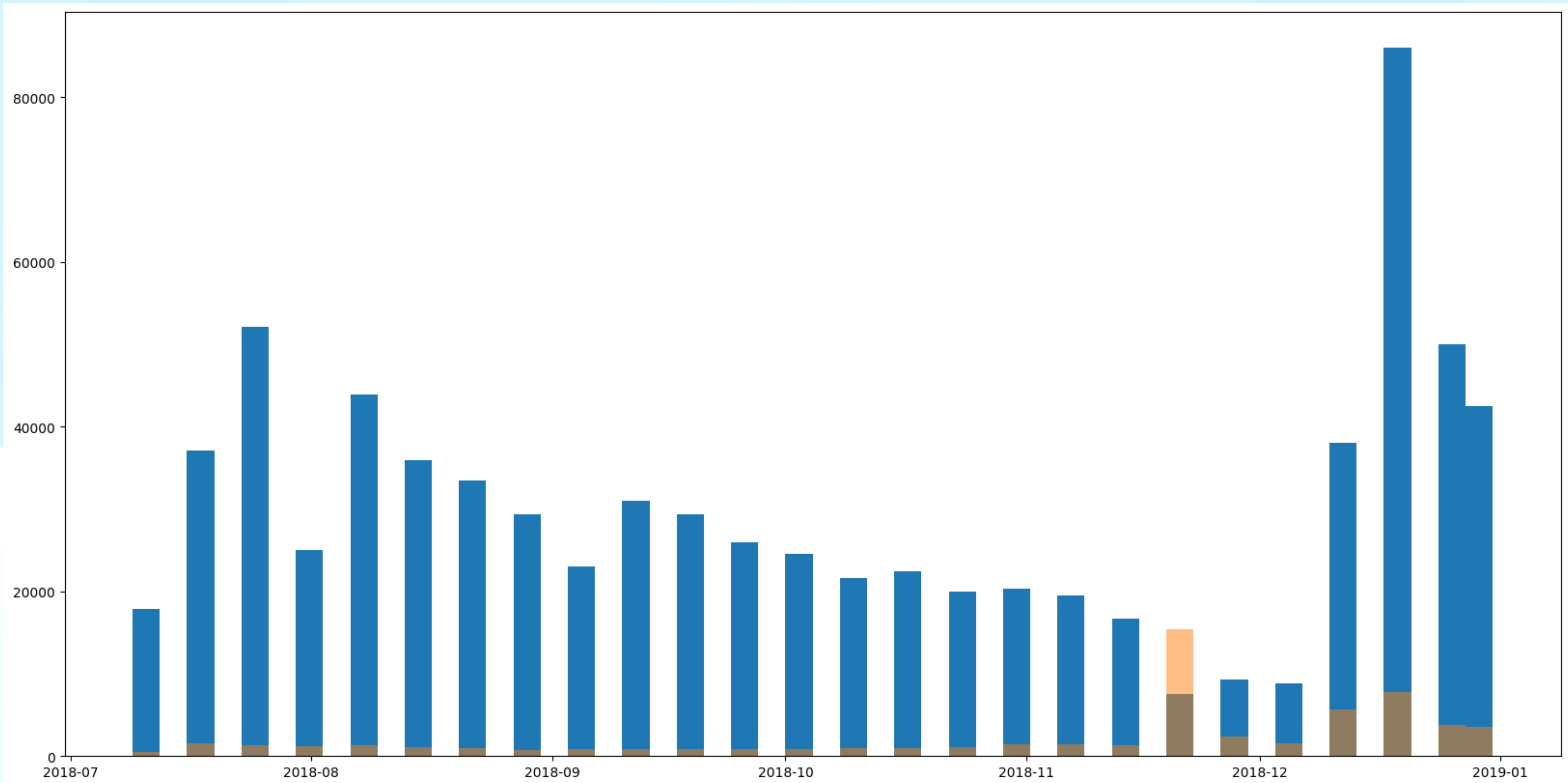
Отдельным набором данных является нормализованный анонимизированный набор признаков, характеризующих профиль потребления пользователя.

Анализ исходных данных



- Самыми частыми услугами являются 1 и 2
- В преддверии нового года наблюдается всплеск активности
- Количество положительных откликов практически неизменно до конца ноября





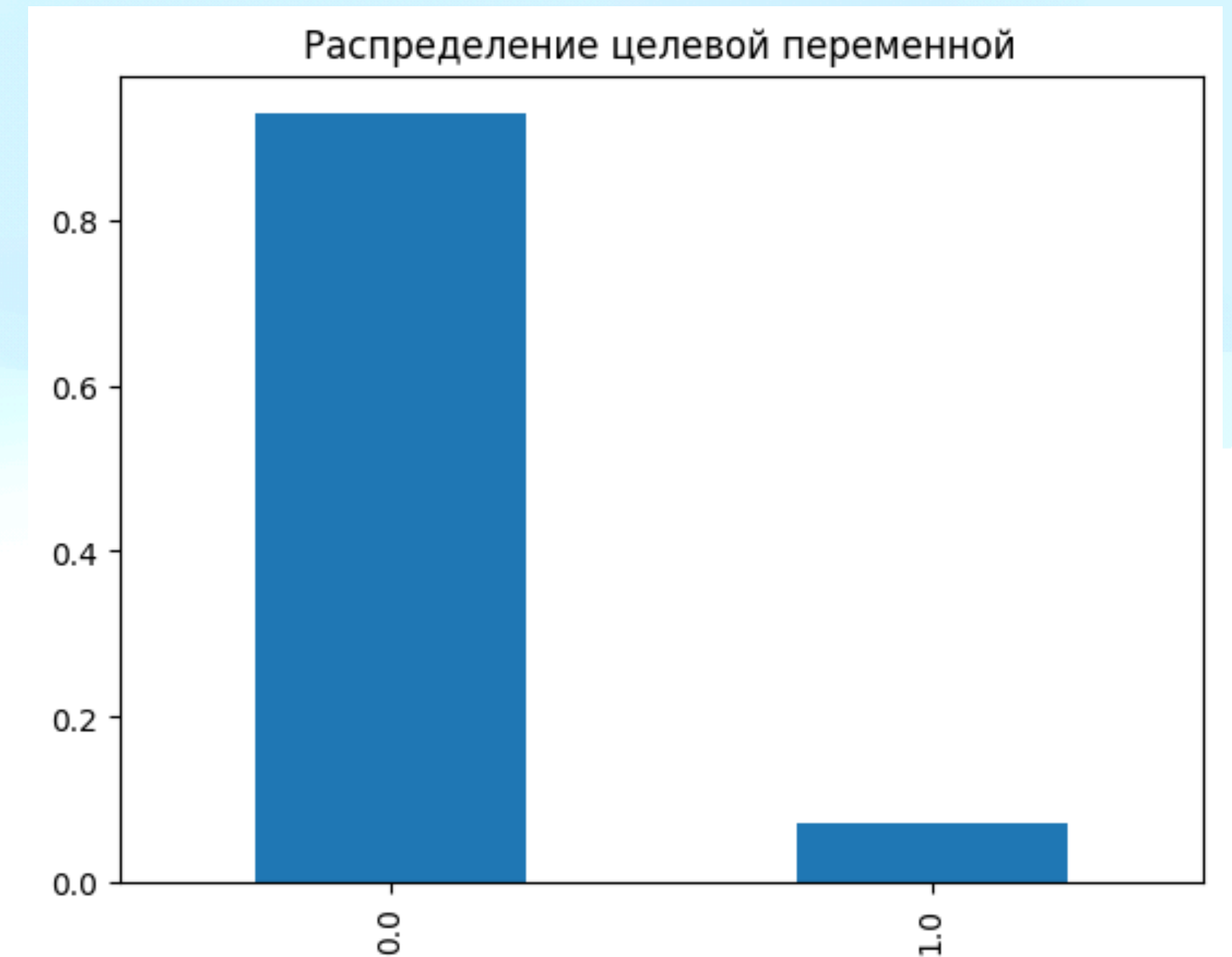
Популярность услуг

	vas_id	0	1
0	1.0	0.982	0.018
1	2.0	0.981	0.019
2	4.0	0.746	0.254
3	5.0	0.982	0.018
4	6.0	0.573	0.427
5	7.0	0.986	0.014
6	8.0	0.974	0.026
7	9.0	0.817	0.183

Наиболее популярны услуги 4, 6 и 7

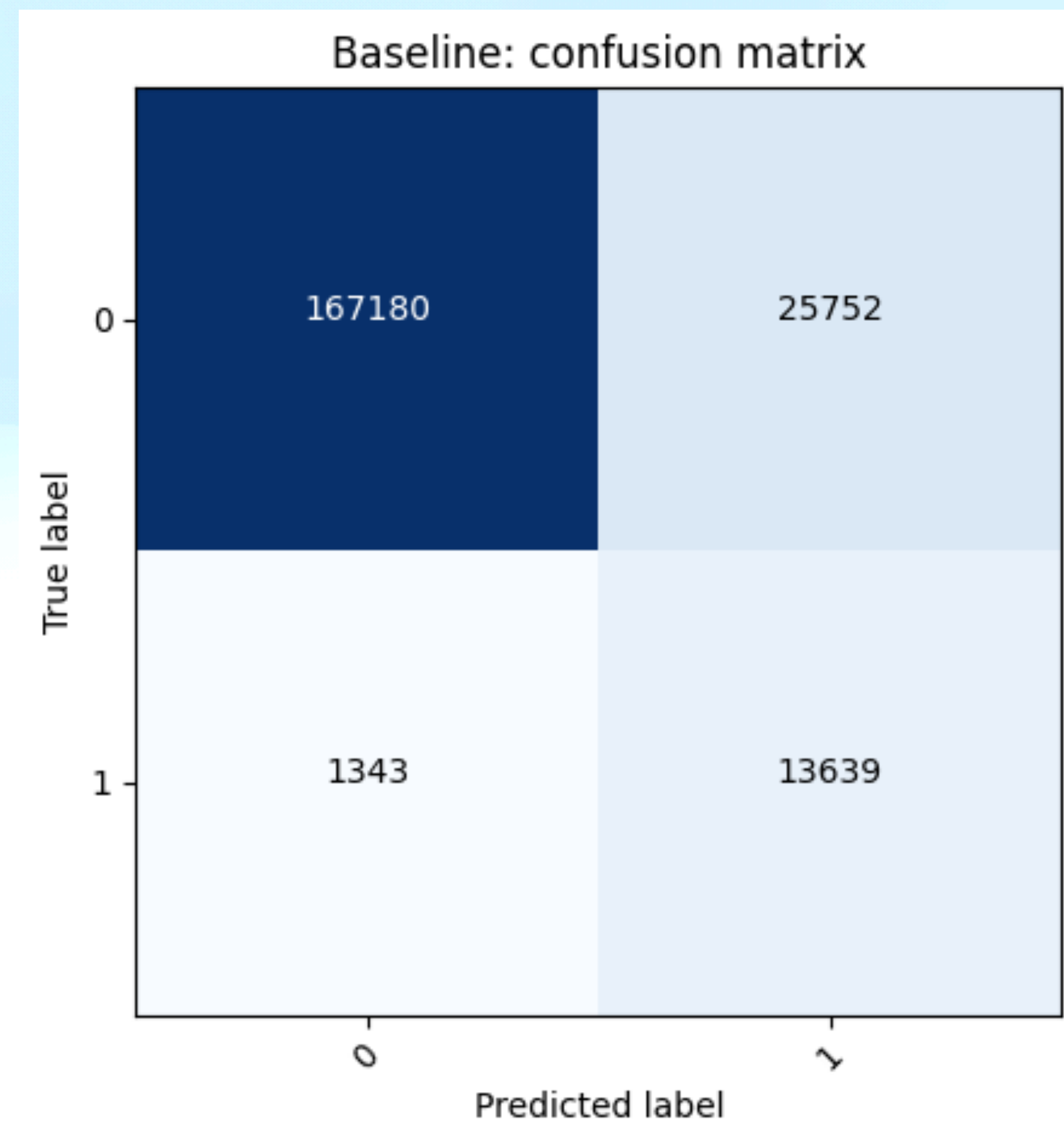
Подготовка данных

- Из исходной таблицы с признаками абонентов (features), удалены записи, отсутствующие в data_train/data_test;
- Выполнено объединение таблиц по user_id;
- Данные из обучающего набора разделены на тренировочную и тестовую выборки;
- Выполнена балансировка данных

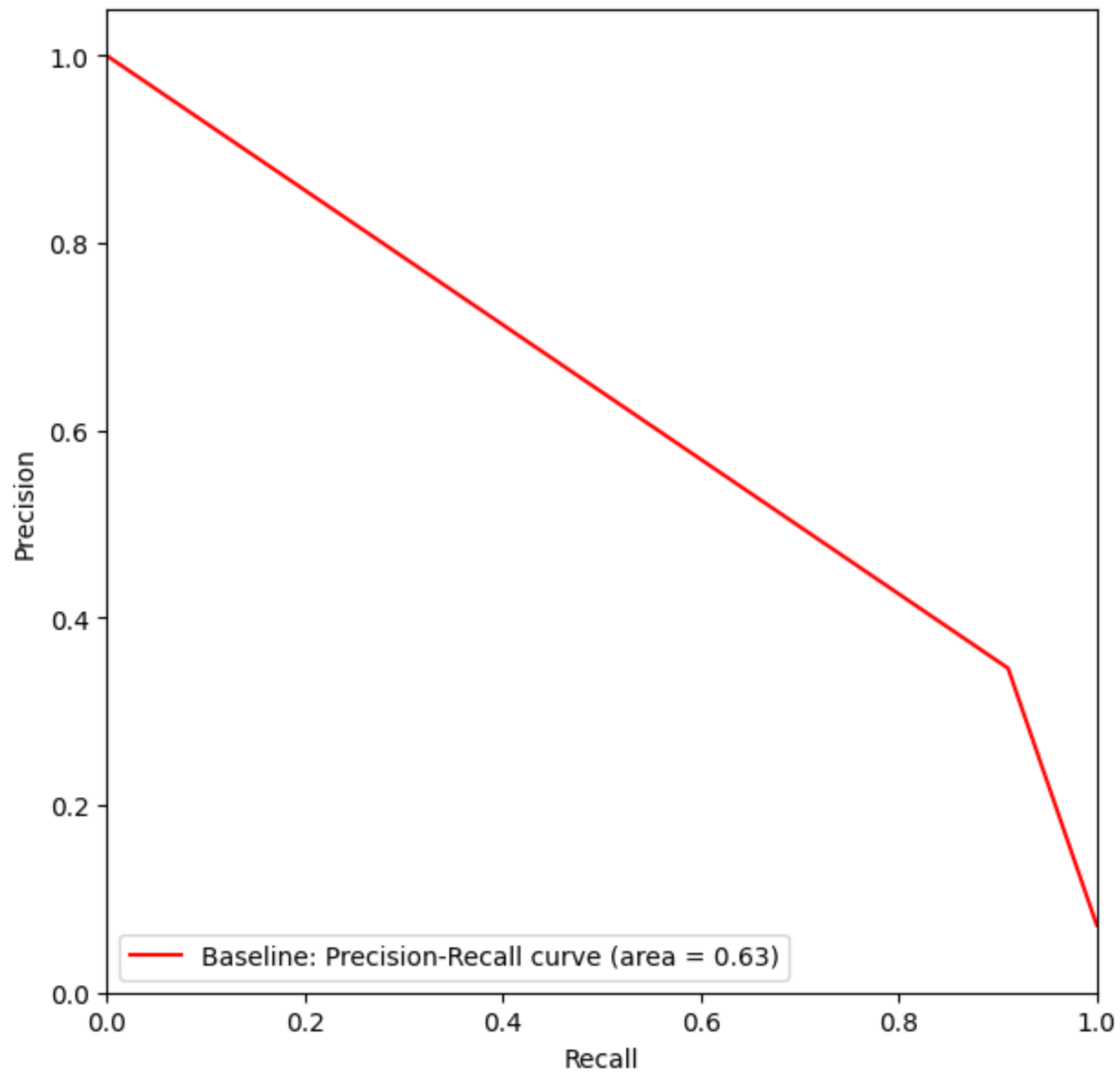


Выбор модели

В качестве модели была выбрана Catboost.



Baseline: Precision-Recall curve



Feature Importance

