

## **Видеокурс от Megafon + курсовой проект**

Андрианов Алексей Юрьевич

### **Задача**

У нас появился запрос из отдела продаж и маркетинга. Как вы знаете «МегаФон» предлагает обширный набор различных услуг своим абонентам. При этом разным пользователям интересны разные услуги. Поэтому необходимо построить алгоритм, который для каждой пары пользователь-услуга определит вероятность подключения услуги.

### **Данные**

В качестве исходных данных вам будет доступна информация об отклике абонентов на предложение подключения одной из услуг. Каждому пользователю может быть сделано несколько предложений в разное время, каждое из которых он может или принять, или отклонить.

Отдельным набором данных будет являться нормализованный анонимизированный набор признаков, характеризующий профиль потребления абонента. Эти данные привязаны к определенному времени, поскольку профиль абонента может меняться с течением времени.

Данные train и test разбиты по периодам – на train доступно 4 месяцев, а на test отложен последующий месяц

### **Алгоритм**

Исходя из данных и поставленной задачи наиболее оптимальным будет создание алгоритма классификация с нахождением вероятности положительного класса. В наших данных присутствуют временные метки, так что при разделении данных для тестирования модели мы будем учитывать это.

## Модель

В качестве бейзлана выбрана модель дерева решений, так как она показывает хороший результат скорость. Основными моделями для решения задачи классификации модели градиентного бустинга, они являются наиболее эффективными сейчас. CatBoost, Light Gradient Boosted Machine и XGBoost применяются чаще всего, так что были использованы они.

### Результаты модели

Для оптимизации гиперпараметров был применен Hyperopt, так как он позволяет уменьшить число экспериментов.

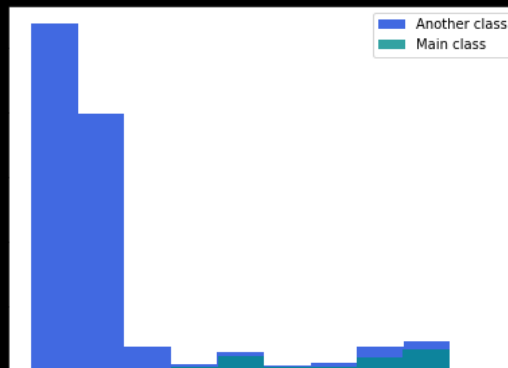
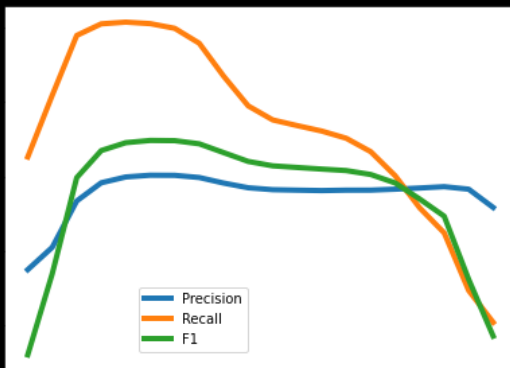
XGBoost показал результаты лучше остальных моделей.

F1 score = 0.716

## Принцип составления индивидуальных предложений

Так как стоимость рассылки предложений во много раз меньше прибыли если пользователь подключит услугу, то нам менее страшна ошибка 1-го рода. Так что при составлении индивидуальных предложений рекомендуется максимизация Recall.

Наибольший Recall получается при  $\text{Threshold} = 0.226$



f1	precision	recall	probability
0.748	0.702	0.905	0.311
0.748	0.701	0.899	0.353
0.745	0.699	0.907	0.268
0.744	0.698	0.879	0.395
0.735	0.691	0.905	0.226
0.732	0.691	0.834	0.437
0.72	0.685	0.795	0.479
0.714	0.682	0.776	0.521
0.712	0.681	0.768	0.563
0.71	0.681	0.761	0.605
0.708	0.681	0.751	0.647
0.703	0.682	0.733	0.689
0.698	0.667	0.889	0.184
0.691	0.683	0.701	0.732
0.67	0.684	0.658	0.774
0.647	0.686	0.624	0.816
0.569	0.605	0.809	0.142
0.562	0.683	0.547	0.858
0.485	0.658	0.504	0.9
0.46	0.575	0.726	0.1

