

Курсовой проект по данным «Мегафона»

Постановка задачи

Задача: предсказание отклика ответа абонентов на предлагаемые услуги.

Исходные данные:

- история откликов на предложения услуг;
- профили абонентов (анонимизированные).

Метрика: F1_score, невзвешенная.

Выбранная модель

CatBoost со следующими основными параметрами:

Параметр	Значение
depth	
learning_rate	
iterations	
eval_metric	F1
early_stopping_rounds	20

Параметры depth, learning_rate, iterations подбирались с помощью поиска по сетке с кросс-валидацией.

Подобранные параметры

Наилучший f1_score показала модель CatBoost со следующими параметрами:

Параметр	Значение
depth	4
learning_rate	0,03
Iterations	10
eval_metric	F1
early_stopping_rounds	20

Сравнение с другими моделями

Выбранная модель CatBoost показала самые высокие результаты по метрике F1_score с кросс-валидацией.

Модель	F1-score с кросс-валидацией
CatBoost	0,888
Логистическая регрессия	0,827
Решающие деревья	0,589

Подход компании к формированию предложений

Выводы на основе анализа исходных данных:

- Одному пользователю предлагаем только одну услугу за раз.
- Можно повторно предложить услугу, от которой абонент ранее отказался.

Формирование предложений

Предлагаемый алгоритм:

1. Отсортировать вероятности подключения услуг по убыванию;
2. Если для пользователя предложено несколько услуг с одним buy_time, то оставить только одну услугу с наибольшей вероятностью;
3. Проверить, что пользователю в последнее время не предлагалась данная услуга и не подключена в настоящее время;
4. Установить порог вероятностей для предложений (по умолчанию 0.5) исходя из: бюджета на рекламную компанию, математического ожидания количества откликов и пр. Необходимо больше информации от бизнеса.