

Задача по предсказанию, сколько потратит клиент в свой первый визит на следующей неделе

Бобров Евгений

Московский государственный университет имени М.В. Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра математических методов прогнозирования

4 октября 2017

Первое решение

Первое решение состояло из двух этапов. Выделение клиентов, которые перестали посещать магазин, и разметка их нулём. Таких было около 10% от всей выборки. Оставшиеся клиенты прогнозировались модой. Было достигнуто качество 0.37672 точности.

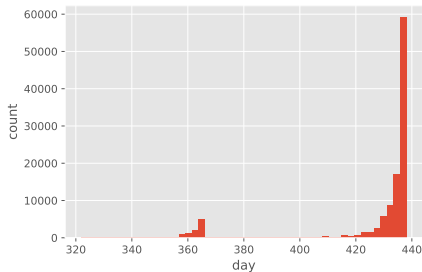


Рис.: Распределение числа последних дней в записях

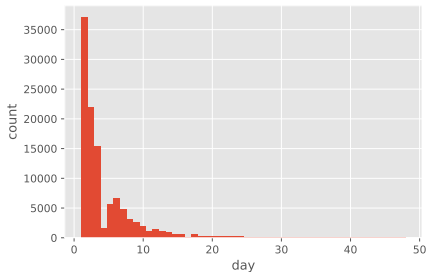


Рис.: Распределение числа первых дней в записях

Дальнейшие рассуждения

Последний классы встречаются крайне редко они были слиты в один, для большей сбалансированности. Последнюю неделю я достаточно безуспешно пробовал обучать нейросети, результаты которых едва ли обходили моду. Тогда было решено обучить градиентный бустинг на стандартных признаках, что позволило с небольшим перевесом преодолеть бейзлайн на 0.39090.

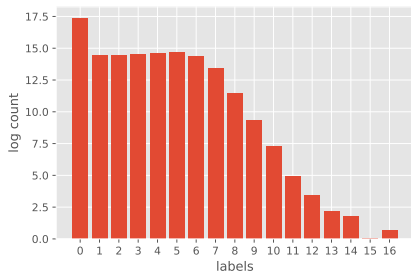


Рис.: Распределение числа меток в логарифмической шкале

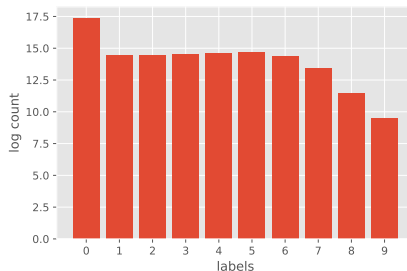


Рис.: Метки после предобработки в логарифмической шкале

Что за данные?

