

Про борьбу с унынием

Домашнее задание по курсу «Введение в анализ данных» № 8

18 мая 2020 г.

1. Общая постановка задания

Иногда так бывает так, что клиенты приуныли. Это называется **оттоком**. Если человек приуныл, он уходит из нашего магазина и больше ничего никогда не покупает, мы получаем меньше прибыли. Было бы классно понимать, что человек начал унывать и предпринимать какие-нибудь меры по его удержанию. Например, если Васе наш сотовый оператор больше не нравится, можно дать ему скидочку или новый тариф или подарить новую Nokia. Тогда Васе сразу станет лучше, и он передумает уходить к другому оператору.

Скрудж МакДак владеет сотовым оператором "МакДак-телеком". Он, как мудрый менеджер, понимает, что терять клиентов из-за оттока плохо. Ему хочется с помощью машинного обучения сделать так, чтобы отток клиентов уменьшился. Он согласен предлагать тем, кто потенциально может сбежать скидки на мобильную связь. За всё время существования, его фирма накопила довольно большой массив данных о том, как именно люди пользовались услугами его сотового оператора.

Вам предлагается подумать о том, как решить проблемы Скруджа. Для этого вам нужно ответить на ряд вопросов:

1. **Что такое отток?** Мы могли бы понять, что человек ушёл от нас, если бы он пришел к нам в отдел сотовой связи и расторг с нами договор. Чаще всего так не происходит. Более того, расторжение договора — последний этап ухода человека. Нам хотелось бы понять, что он собирается бросить нас хотябы за месяц, чтобы заранее предложить ему скидку.
Представим себе, что данные Скруджа — это табличка в пандасе. Там есть информация о всём, что люди делали внутри нашей сети (сколько смс-ок отправили, как часто звонили и тп). Как по этим данным понять, что пользователь приуныл? Как бы вы проверили на данных — хорошее ли правило вы придумали для определения оттока?
2. Переменную, которую вы определили в предыдущем пункте мы и будем стараться спрогнозировать. **Что за задачу машинного обучения мы решаем** (классификация/регрессия и тп) ?
3. **Какие переменные можно было бы использовать в качестве объясняющих?** Откуда их взять (попробуйте отобрать те переменные, которые скорее всего точно есть у телеком-компаний)? Понятное дело, что точно вы знать не можете. Просто предположите.
4. Какие этапы предобработки нужно провести с данными перед тем, как обучать для них модель? С какими проблемами вы скорее всего столкнётесь (переобучение, выбросы, пропуски и другое)? Как вы будете с ними бороться?

5. Как бы вы рекомендовали Скруджу использовать прогнозы, которые выдаёт модель для новых пользователей? Чтобы ответить на этот пункт, нужно описать схему того, как ваши прогнозы помогают оптимально распределить бюджет, выделенный на скидки по проблемным людям.
6. Как можно измерить пользу от внедрения новой системы скидок и понять, что модель предсказания оттока правда помогает удерживать клиентов?
7. Скрудж хочет попробовать удерживать людей не только скидками, но и добавлением в старые тарифы новых индивидуальных функций. Например, если человек часто звонит в другую страну, вместо скидки на текущий тариф, ему предлагают в тариф дополнительные минуты на подобную связь. Как бы вы могли сравнить в рамках текущей модели, какой из двух способов лучше?

На выходе должен получиться небольшой текст, в котором находятся ваши рассуждения. Пожалуйста, структурируйте их так, чтобы было понятно где и на какие вопросы вы отвечаете. **Внимание!** Не надо писать огромный мануал про машинное обучение! Постарайтесь изложить все свои мысли на 2-3 страницах **структурированного понятного текста**. Работа сдаётся в формате pdf в anytask.

2. Организационные мелочи

Команды

Задание выполняется в командах по 2-3 человека. Индивидуальные работы не будут засчитаны.

Дедлайн

- Задание выдаем 18.05.2020
- Задание сдаем 05.06.2020 в 23.30 на anytask.org

Оценивание

Работа оценивается по 10-бальной шкале. Она ставится на команду и нормируется на число человек в ней. Так, если в команде три человека и она получает оценку 8, она превращается в 24 балла. Если в команде два участника, то 8 соответственно превращается в 16. Далее участники команды делят эти баллы между собой в любой пропорции в зависимости от вклада каждого участника в итоговую работу.

Например. Команда из Рона, Гермiony и Гарри получила 24 балла. Большую часть работы выполнила Гермiona, она забирает себе 10 баллов. Ещё 8 баллов себе забирает Гарри, так как он талантливый. Рон получает 6 баллов, так как он раздолбай и почти ничего не делал. Команда посоветовалась и решила, что так распределить баллы будет справедливо. Больше 10 баллов забрать себе нельзя.

Ваша итоговая оценка зависит от того, насколько глубоко проработана задача, насколько чётко обоснован каждый этап работы, каждая метрика.

В каком виде сдаём?

1. Ждем ваши мысли в виде pdf-ки
2. Пожалуйста, укажите в своей работе всех членов команды и номер группы.
3. Постарайтесь делать осмысленные выводы! Не будьте мудрыми, как король:

