

Все любят пиццу

Большое домашнее задание по курсу «Введение в анализ данных» № 2

1 июня 2019 г.

1. Общая постановка задания

Мистер Пануччи владеет сетью пиццерий. Для доставки пиццы мистер Пануччи использует довольно большую сеть из курьеров. Он, как мудрый менеджер, понимает, что его текущая система работает неоптимально. В одних районах курьеры регулярно простаивают, в других не хватает рук, и заказы опаздывают. Хочется с помощью машинного обучения решить эту проблему, и каждую неделю перераспределять силы курьеров так, чтобы они не простаивали, и все заказчики получали свою пиццу вовремя.

Вам предлагается подумать о том, как эту проблему можно было бы решить. Для того, чтобы решить эту задачу, вам нужно ответить на следующую серию вопросов:

1. Как можно измерить, сколько денег сейчас теряет мистер Пануччи из-за неоптимальности курьерской сети?
2. **Какую именно целевую переменную надо научиться прогнозировать, чтобы решить проблемы, связанные с курьерами? Что за задачу машинного обучения придётся для этого решить (классификация/регрессия и тп) ?**
3. Как надо использовать прогнозы, чтобы решить проблемы пиццерии? Чтобы ответить на этот пункт, нужно описать схему того, как ваши прогнозы помогают оптимально распределить курьеров по районам.
4. **Какие переменные можно было бы использовать в качестве объясняющих? Откуда их взять (попробуйте отобрать те переменные, которые реально раздобыть)?**
5. 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?
6. **Как можно измерить пользу от внедрения новой системы распределения курьеров?**

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 04.06.2019
- Задание сдаем 13.06.2019 в 23.30 на anytask.org

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:



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. Далее участники команды делят эти баллы между собой в любой пропорции в зависимости от вклада каждого участника в итоговую работу. Если в команде два участника, то 8 соответственно превращается в 16.

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