

Применение методов ИИ в сетях связи

Мы с вами попробовали в работе несколько инструментов для анализа данных:

- Microsoft ML Studio.
- Python + Scikit Learn и сопутствующие библиотеки.
- Python + Tensorflow и Keras.

Каждый из этих инструментов имеет свои плюсы и минусы, и вы в процессе выполнения лабораторных, возможно, уже поняли, что вам понравилось использовать больше. И, безусловно, этими инструментами не ограничивается все многообразие инструментов в области ИИ. Однако мы изучали инструменты машинного обучения при помощи заранее подготовленных наборов данных и известных команд. В реальной жизни чаще всего нет ни того, ни другого.

Теперь я предлагаю вам попробовать проанализировать реальные данные, и проанализировать без списка шагов, который почти наверняка сработает.

Ниже приведены две задачи и несколько наборов данных к каждой из задач. Вам нужно взять только одну из этих задач (на ваш вкус) и наиболее понятный/перспективный/красивый набор данных на ваш взгляд. Далее, вам надо попробовать построить модель при помощи одного из изученных ранее инструментов, обучить ее при помощи одной части данных и протестировать на другой (большинство наборов данных не разбито на две части, то надо сделать самостоятельно).

Задача 1. Предсказание оттока клиентов оператора связи (т.н. churn). Иногда клиенты переходят от одного оператора связи к другому, и операторов то чаще всего не радует. В приведенных ниже наборах данных приведены различные параметры взаимоотношений оператора связи со своими клиентами, а также информация о том, ушел ли клиент от оператора связи (churn). Попробуйте сделать модель, предсказывающую уход клиента на основе этих данных:

<https://www.kaggle.com/abhinav89/telecom-customer>

<https://www.kaggle.com/becksddf/churn-in-telecoms-dataset>

<https://www.kaggle.com/zagarsuren/telecom-churn-ibm-watson>

Задача 2. Предсказание роста числа устройств или трафика в сетях связи. На основе этих данных попробуйте построить регрессионную модель, которая может предсказывать ту или иную характеристику сети (например, число подключенных мобильных устройств или число посещений страницы в интернете) в зависимости от года и или дня (в том числе и в будущем):

<https://www.kaggle.com/taniaj/world-telecommunications-data>

<https://www.kaggle.com/c/web-traffic-time-series-forecasting/data>

Результат можно предоставить в свободной форме.

Важно: на каждой из приведенных страниц сайта Kaggle.com на вкладке Kernels есть примеры работы с этими данными от других пользователей. Изучать и заимствовать их можно, но если вы берете чей-то код, вы должны понимать, как он работает и написать это в виде подробных комментариев. Если комментарии будут просто переводом комментариев с сайта, такая работа получит максимум оценку «удовлетворительно».