

Прогнозирование маршрутов передвижения пассажиров Московского метрополитена на основании данных о валидации транспортных карт

Липко Иван
ID: 1220663
ivanlipko@yandex.ru



Описание задачи

Ежедневно в России миллионы пассажиров совершают поездки на различных видах транспорта, формируя большой пул данных о своих перемещениях. Например, Московский метрополитен в сутки перевозит до 7 млн человек. А к 2030 году планируется открыть дополнительно до 200 новых станций. Около 60% пользователей приложения «Московский транспорт» при построении мультимодальных маршрутов используют метро Москвы.

Управление пассажирскими потоками – сложный процесс, который требует анализа постоянно увеличивающегося объема информации, в том числе о ежедневных маршрутах пассажиров. При решении подобных аналитических задач на помощь приходят технологии искусственного интеллекта, которые позволяют на основе больших данных изучить транспортное поведение пользователей, рассчитать нагрузки на ветки метрополитена, собрать статистику о наиболее популярных направлениях.

На основании данных, полученных в результате валидации пассажирами проездных билетов/банковских карт на станциях метро, участникам чемпионата нужно определить наиболее популярные маршруты для пользователей приложения «Московский транспорт», которые передвигаются по городу на метро, и научиться их предсказывать. Особенностью задачи является то, что конечную точку и путь пассажира нужно предсказать, основываясь на данных о станции отправления и его прошлой истории поездок. Прогнозирование маршрутов позволит в будущем уведомлять пассажиров о произошедших изменениях, предлагать альтернативные варианты перемещения по городу.

Основная идея

- ОтФильтруем выбросы
- Не учимся на малых данных – всё для большинства
- Экспериментируем с параметрами модели
- Уменьшаем объём используемой памяти
-
- Кривые обучения для контроля качества

История качества

- # 0.4897555334746157 -- baseline
- # 0.2678758686368177 -- **baseline, обучение на всех данных. На паблице 0.092150**
- # 0.0028946815886393275 0.4765418328846045
0.2397182572366219 -- **logit, linreg public -0.177917**
- # 0.051367635290150336 0.5337106363746371
0.2925391358323937 -- RandomForestClassifier,
RandomForestRegressor

