Сервис расчета и моделирования пассажирского потока на основе данных о поездках и проходах на турникетах

1A4 • 25.08.2023

Обзор

Мы поняли задачу, как создание модели для анализа влияния различных факторов на пассажиропоток в метро

Для прогнозирования данных используется искусственный интеллект. В данном случае мы решили использовать CatBoost (ансамбль-модель от яндекса)

Прогнозирования количества человек относится к задачам регрессии

Источники данных

Изначальные данные

- Пропускная способность каждой станции метро
- Количество посетителей станций по часам

Обработка даты

- Федеральные праздники
- Указание дня недели

Портал Открытых Данных Правительства Москвы - data.mos.ru

• Перехватывающие парковки

Погода по часам - rp5.ru

- Температура
- Влажность
- Облачность
- Осадки

Росстат

Количество
проживающих людей
вокруг станции метро

Проблемы

Слабая стандартизация статистики

 Данные было трудно собирать и приводить к требуемому формату для датасета

Отсутствие единого реестра мероприятий

 Ближайший к желаемому это Яндекс.Афиша, но у него закрытый API

Доработка

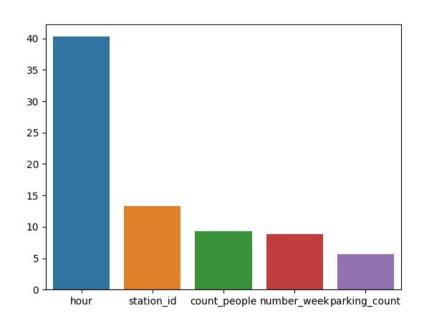
Добавление большего количества источников данных

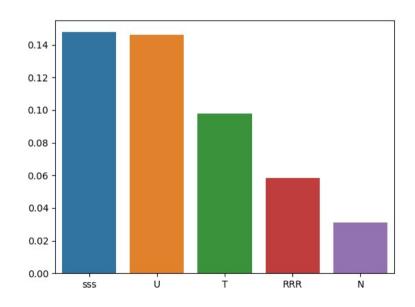
 Например, данные о наземном транспорте или данные с реестра мероприятий

Усовершенствование интерфейса

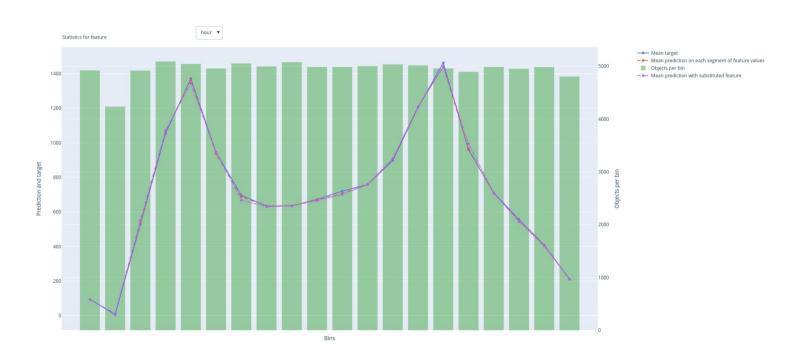
- В данный момент это просто скелет для демонстрации работы модели
- Также можно подвязать прогноз погоды к проекту
- Подвязать KudaGo.ru

Статистика модели - влияние параметров на результат

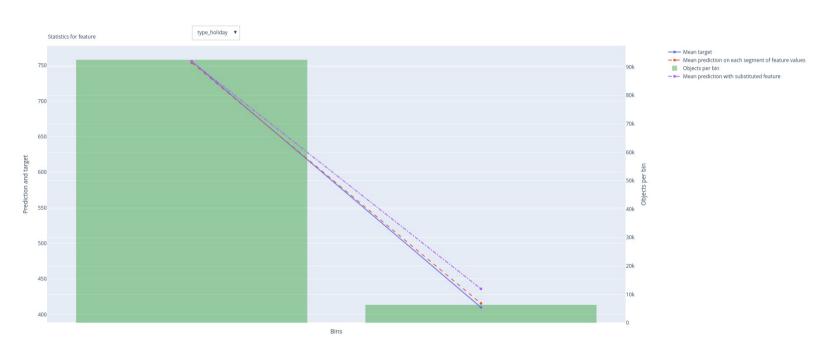




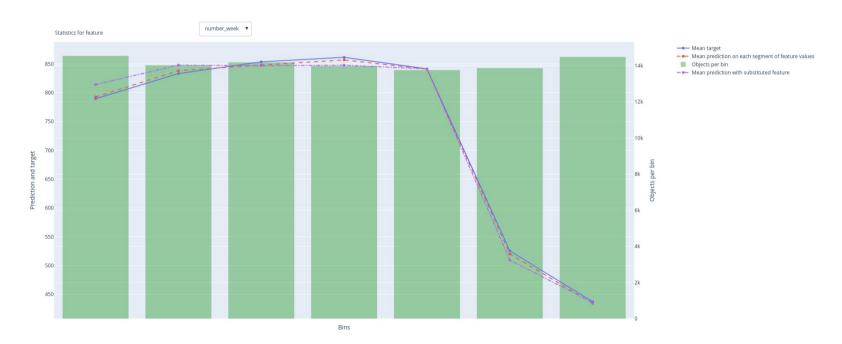
Статистика модели - влияние часов на поток людей



Статистика модели - влияние выходных и праздников



Статистика модели - влияние дня недели на пассажиропоток



Интерфейс (Демонстрация)

