

КОНКУРС ДЛЯ НАЧИНАЮЩИХ АНАЛИТИКОВ ДАННЫХ

GIS CHALLENGE 2020

One Hot Team

Александр Хохлявин

Виктория Стрельцова

Симар Муратов

[1 0 0]

[0 1 0]

[0 0 1]

WHO WE ARE



Симар
Муратов



Александр
Хохлявин



Виктория
Стрепельцова

Предметная область

Что? Мониторинг автомобилей

Для чего?

- Оптимизация расходуемых ресурсов
- Пресечение мошеннических действий
- Выявление атипичного поведения водителей
- Улучшение (максимизация результата при минимизации затрат) логистических бизнес-процессов

Как? С помощью специальных датчиков

Задачи

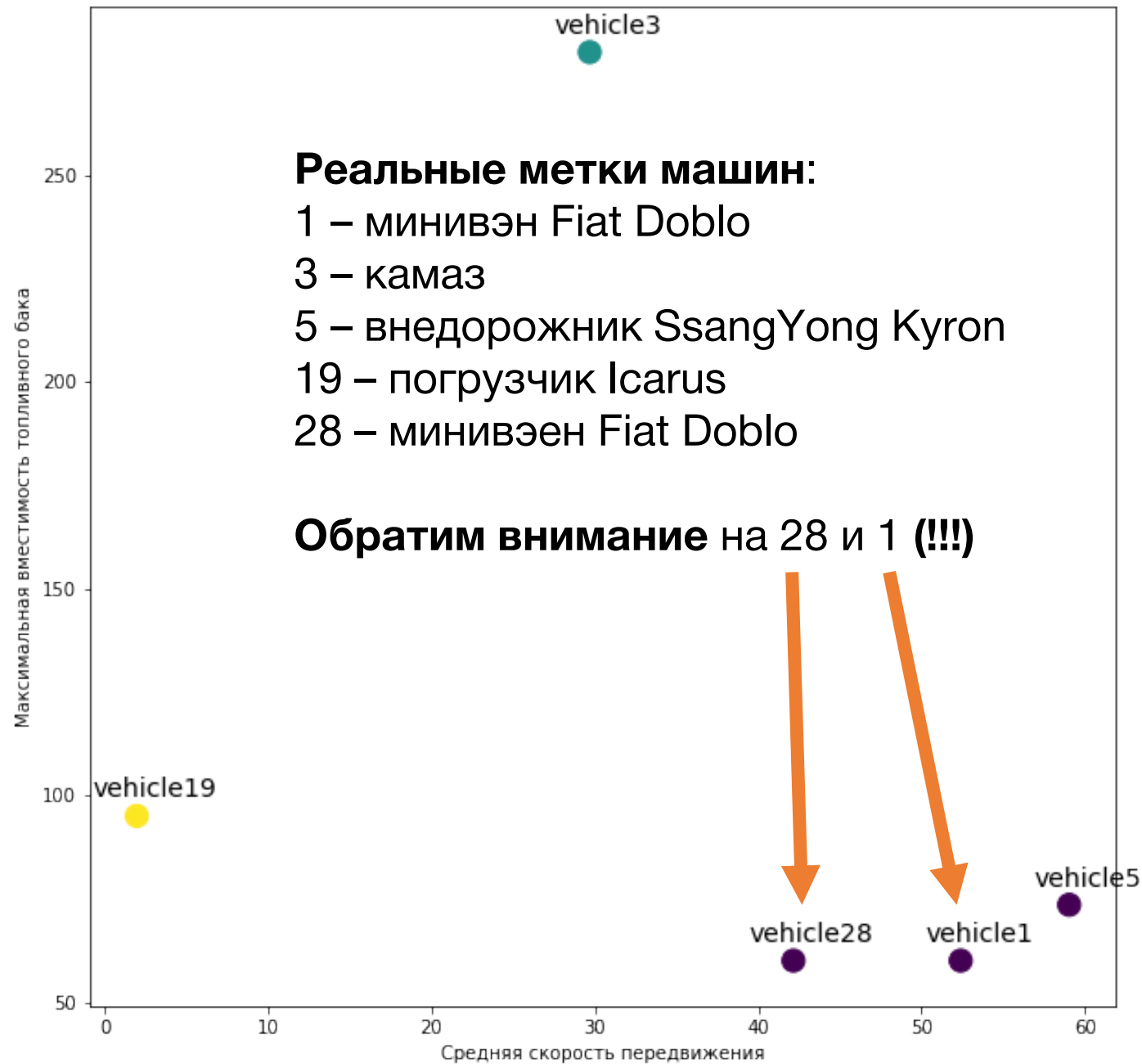
- Выявить заправки топлива с точностью до 5%
- Выявить подозрения на хищения топлива (сливы)
- Выявить режим работы техники с наибольшим расходом топлива
- Выявить режимы перевозки грузов
- Определить стиль вождения
- Классифицировать транспортные средства по типам

Преоброцессинг данных

- Транспортных средств не много,
но датасетов **много**, а данных **ещё больше**
- **Изобретаем** точечную нотацию: `vehicle.dataset.value`
- Поправить типы данных для дробных, целых и дат

Классификация транспортных средств **по типам**

- Объектов немного – признаков **мало**
- **Медианное** и **среднее** значения оборотов двигателя за весь период наблюдений
- **Размах** (разность между максимальным и минимальным значениями) высот
- **Максимальная** и **средняя** скорость передвижения транспортного средства
- **Максимальные** значения уровня топлива (что может быть интерпретировано как **вместимость бака**) и некоторые другие

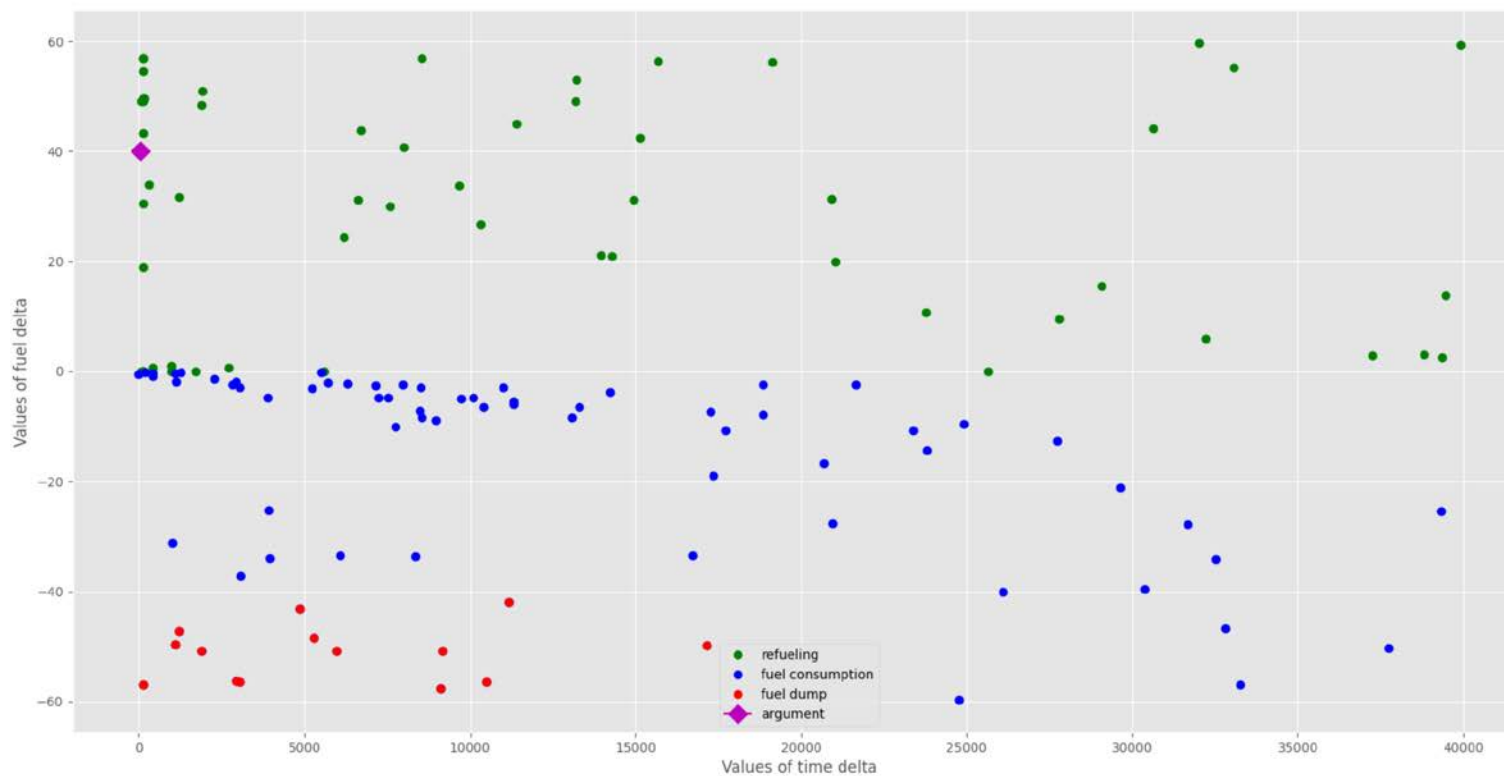


Выявление сливов/заправок

Эвристики

- Это **слив**, если:
 - есть сильный расход топлива за небольшой промежуток времени
 - выключено зажигание и/или обороты двигателя постоянны (нулевые)
- В противном случае – либо заправка, либо обычный режим
- **Заправка**, если:
 - Есть большое изменение уровня топлива со знаком +
 - Выключено зажигание
- Иначе **обычная работа**, если:
 - Включено зажигание, расход топлива небольшой за недолгое время

Классификация методом k-ближайших соседей



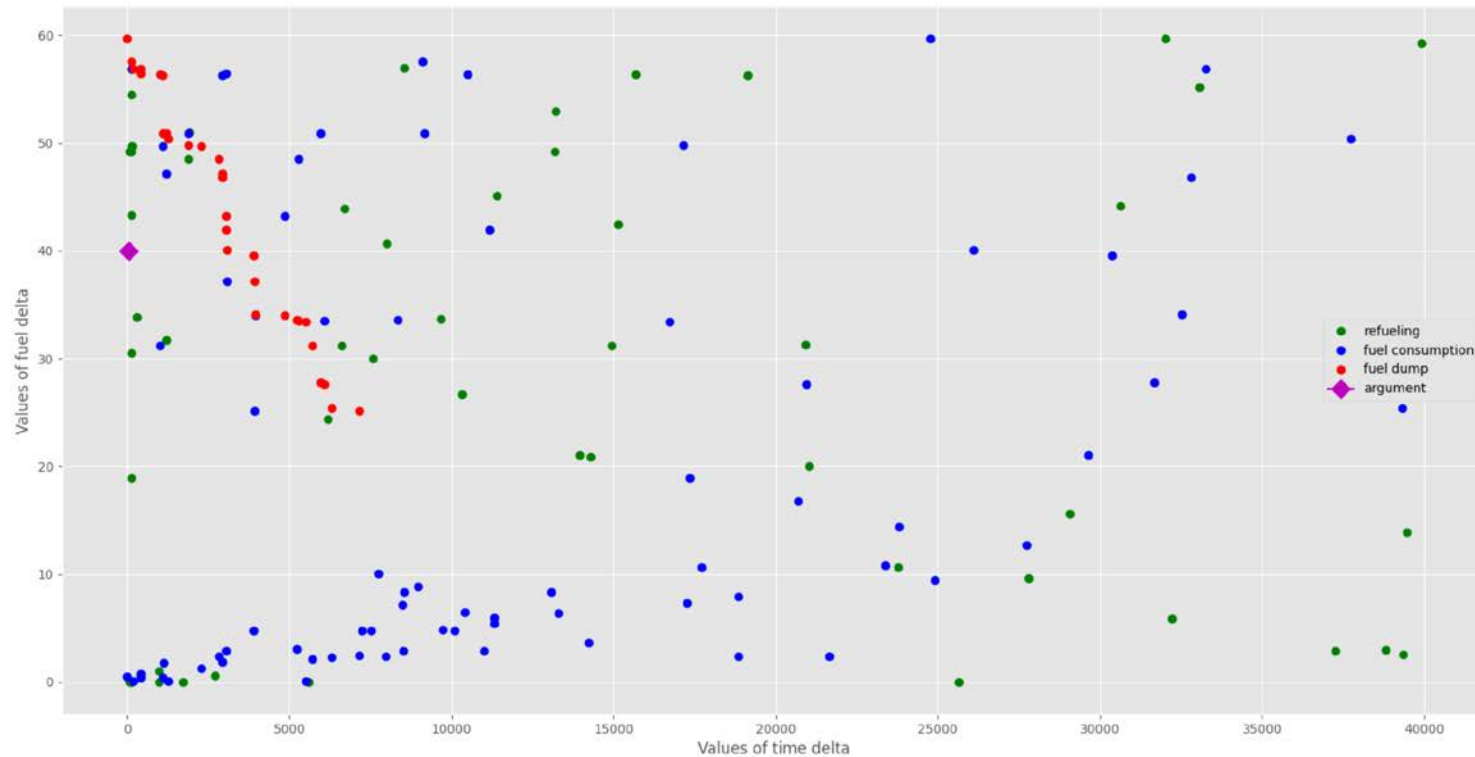
Результат применения первого типа классификации

Классификация методом k-ближайших соседей

```
for i in range(0, len(data_delta_fuel_level_neg), 1):  
    if abs(data_delta_fuel_level_neg[i]) - average_delta_fuel_level_neg > scatter * average_delta_fuel_level_neg \   
    and data_delta_time_neg[i] - average_delta_time_neg < scatter * average_delta_time_neg:  
        fuel_dumps.append(data_delta_fuel_level_neg[i])  
        time_dumps.append(data_delta_time_neg[i])
```

Основное условие первого типа классификации

Классификация методом k-ближайших соседей



Результат применения второго типа классификации

Классификация методом k-ближайших соседей

```
max_fuel=sorted(data_delta_fuel_level_neg,key=float,reverse=True)
min_time=sorted(data_delta_time_neg,key=float)
while max_fuel[j]>average_delta_fuel_level_neg+(scatter*average_delta_fuel_level_neg/100
    fuel_dumps.append(max_fuel[j])
    time_dumps.append(min_time[j])
    j+=1
```

Основное условие второго типа классификации

Выявление режима работы техники
с **наибольшим** расходом топлива

Параметры, характеризующие режим работы:

- скорость движения (км/ч)
- скорость оборотов двигателя (оборотов/мин)

Целевой признак:

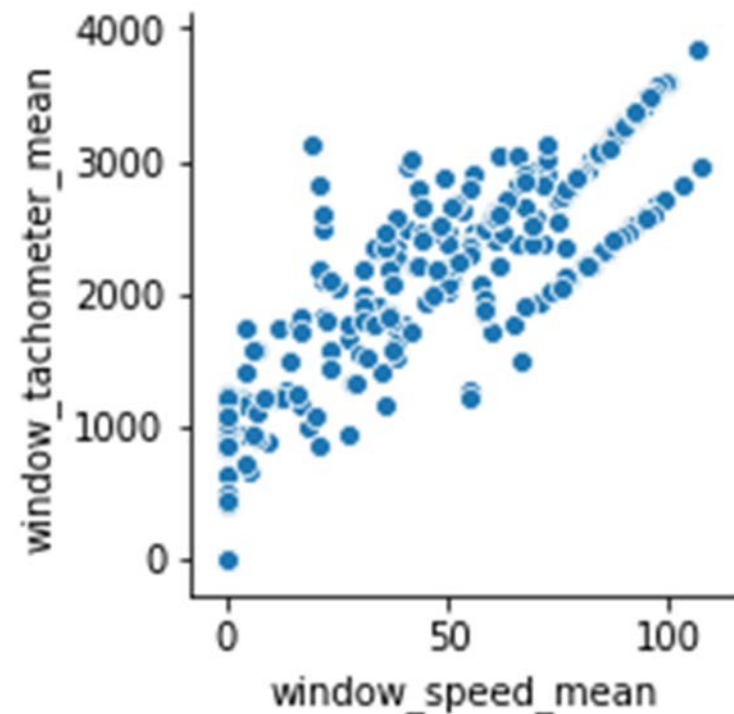
- расход топлива

Методы:

- метод плавающего окна
- линейная регрессия

Выявление режима работы техники с **наибольшим** расходом топлива

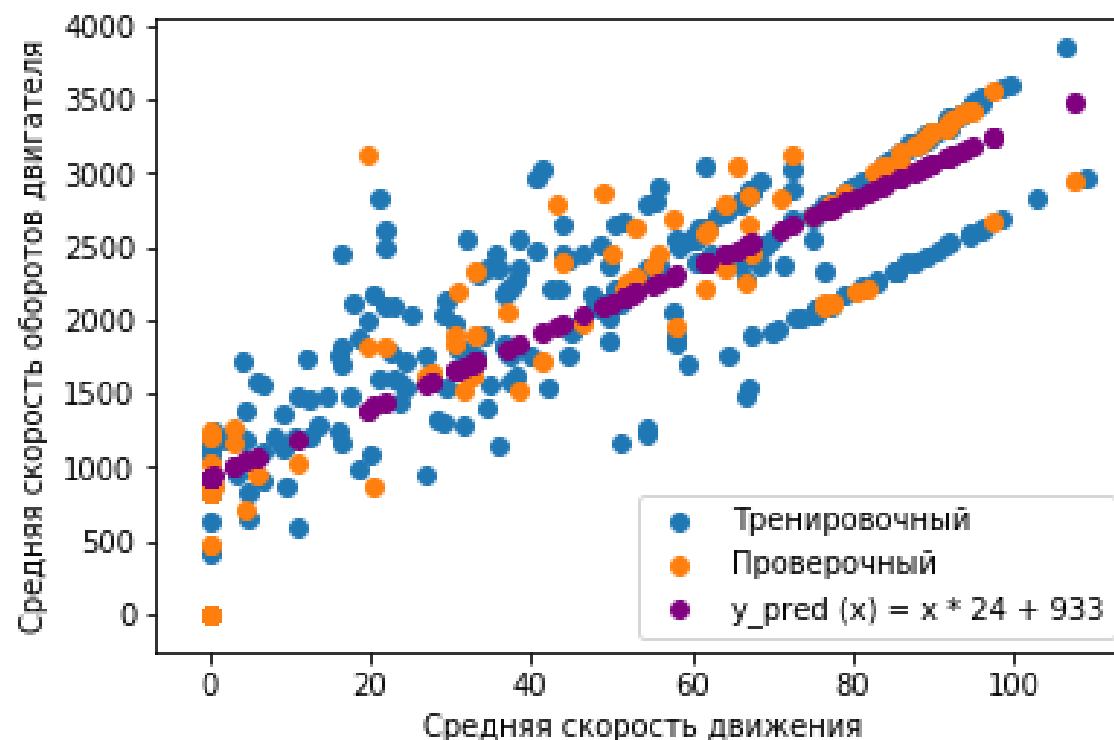
- Были оставлены данные только с максимальным значением расхода топлива
- Распределение значений скорости и оборотов двигателя в этом случае показало, что максимальный расход дает не пара этих значений и не интервал, а зависимость
- Зависимость определяем с помощью линейной регрессии



Выявление режима работы техники с **наибольшим** расходом топлива

- С помощью метода **линейной регрессии** была определена зависимость между скоростью и оборотами
- Полученная зависимость:

$$rotation = 24 * speed + 933$$



Определение **стиля вождения**

Предполагаемые **стили**:

- безопасный
- умеренно-опасный
- опасный

Ключевые **признаки**:

- скорость движения (км/ч)
- высота отн. уровня моря (м)
- скорость вращения двигателя (оборот/мин)

• **Методы**:

- плавающего окна

Метод плавающего окна в определении стиля вождения

- **Предположение:** резкие скачки скорости, показателей высоты и оборотов двигателя говорят об умеренно опасном или опасном стиле вождения
- **Решение:** на протяжении движения вычисляем разности характеристик в два близлежащих момента времени (движемся по данным «окном»), фиксируем разности, превышающие заданные планки.
- Ширина «окна» индивидуальна для каждого параметра и определяется эмпирически.



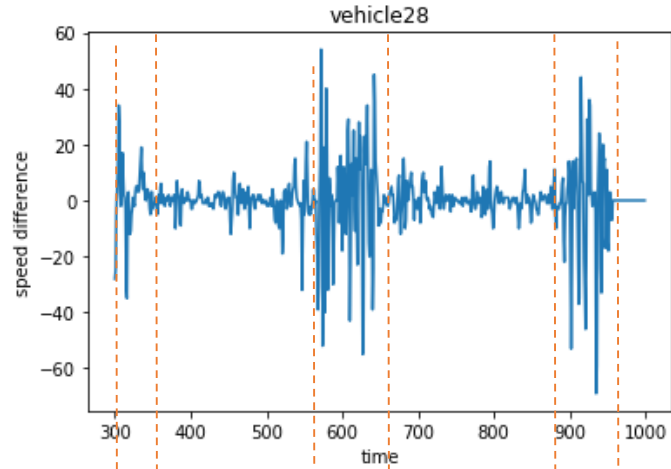


График изменения **скорости**
Ширина окна – 8 секунд

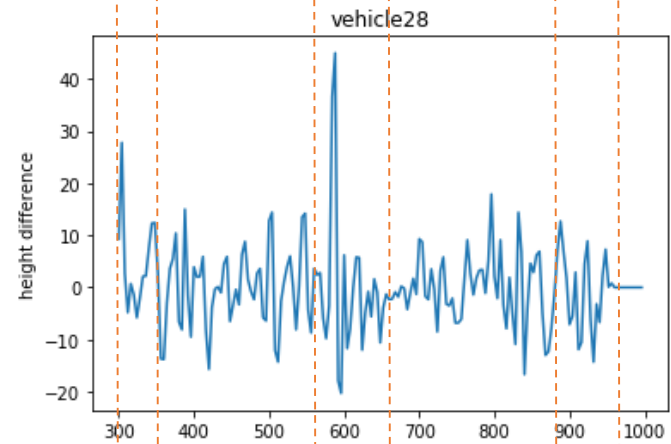


График изменения **высоты**
Ширина окна – 10 секунд

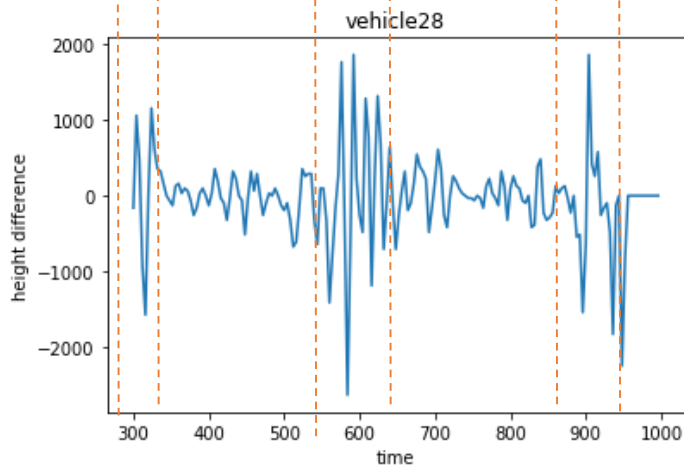


График изменения **числа оборотов двигателя**
Ширина окна – 8 секунд

Итоги

- Проведена **очистка, препроцессинг** и подготовка данных для подачи в модель машинного обучения
- Представлен **универсальный подход** для загрузки данных: точечная нотация данных – N ТС, K датчиков.
- Применены **эвристические** и **алгоритмические** подходы к решению всех представленных в конкурсе задач
- Получены **результаты** классификаций, сравнили с **реальными** данными

Нестареющая классика

