**Тестовое задание 1.**

**Постановка задачи.**

Имеется 80 моторов, каждый из которых живет определенное число циклов. Каждый цикл происходит снятие информации с датчиков S1 – S21 и происходит переустановка настроечных параметров Setting1 – Setting2.

**Необходимо:**

Провести анализ предоставленных данных по 80 моторам (файл Data\_80.csv). Сделать предположение о выходе из строя мотора на основании имеющихся данных датчиков. Далее на основании дополнительно предоставленных данных по 20 моторам (файл Data\_Add\_20.csv) сделать предположения о выходе мотора из строя на следующем цикле.

**Инструменты:**

Любые инструменты Data Science.

**РЕШЕНИЕ**

**Предположение**

В реальном мире с не работающего мотора не приходят данные датчиков. То есть я предположил, что последний по номеру отсчет это последний рабочий период мотора и в следующий период он сломался. Дополнил файл Data\_80.csv колонкой Break и соответствующим образом заполнил. Получил обучающую выборку.

**Работа с данными**

Мы понимаем, что внутри временные ряды.

Посмотрев на данные, выяснили, что s1, s10, s16, s18, s19 и s5 это константы в обоих выборках. Исключаем из данных.

**Моделирование**

С такой выборкой можно решать задачу регрессии – оценивать вероятность выхода из строя (0 – не выйдет, 1 – выйдет). И задачу классификации – оценивать в явном виде да\нет, 0 – не выйдет, 1 – выйдет.

Инструменты можно применить разные, от линейной регрессии (библиотека питона sklearn) до нейронных сетей (keras\tensorflow). Мне близок инструмент градиентного бустинга, использовал в версии для классификации.

**План атаки**

1. **Сделано** быстрое решение – взять из запасов XGBoost драфт, быстро проведено моделирование и получен результат. Все на гитхабе.
2. В дополнение в рамках текущей задачи **планирую** «развернуть» временные данные, получиться дополнительное количество фич, повторю моделирование, пришлю результаты.