Требования к решению заданий:

1. Язык программирования: R или Python.
2. Воспроизводимость: пользователь Вашего решения должен иметь возможность запустить Ваш код и получить результаты схожие с Вашими.
3. По первому заданию необходимо подготовить аналитический отчет с программным кодом, по второму заданию – аналитический отчет с программным кодом и csv-файлом с предсказаниями на тестовой выборке.

## **Задание 1**

Сотрудник службы оценки качества продукции заметил, что число бракованных листов стали в партии для стали марки A больше, чем для стали марки B. Также он обратил внимание, что при средней скорости прокатки более 4 м/с число бракованных листов больше. В соответствии с этими наблюдениями, предлагается снизить скорость прокатки и ввести дополнительные меры контроля качества для стали марки A.

Обоснуйте, что:

1. Более 3 бракованных листов на партию выходит значимо чаще для стали марки А, чем для стали марки B.
2. При скоростях прокатки более 4 м/с свыше 3 бракованных листов стали на партию выходит значимо чаще, чем при меньших скоростях прокатки.

Данные, необходимые для анализа, содержатся в файле «Статистика за 2018 год». Количество листов стали в каждой партии предполагается одинаковым.

## **Задание 2**

Вам необходимо построить модель, которая на основании данных, поступающих каждую минуту, определяют качество продукции, производимое на обжиговой машине.

Обжиговая машина представляет собой агрегат, состоящий из 5 одинаковых по размеру камер, в каждой камере установлено по 3 датчика температур. Кроме этого, для данной задачи Вы собрали данные о высоте слоя сырья и его влажности. Высота слоя и влажность измеряются при входе сырья в машину. Сырье проходит через обжиговую машину за час.

Данные с показателями работы обжиговой машины содержатся в файле X\_data.csv:

|  |  |
| --- | --- |
| **Название тега** | **Описание тега** |
| T\_data\_1\_1 | 1-й датчик в 1-й камере |
| T\_data\_1\_2 | 2-й датчик в 1-й камере |
| T\_data\_1\_3 | 3-й датчик в 1-й камере |
| T\_data\_2\_1 | 1-й датчик во 2-й камере |
| T\_data\_2\_2 | 2-й датчик во 2-й камере |
| T\_data\_2\_3 | 3-й датчик во 2-й камере |
| T\_data\_3\_1 | 1-й датчик в 3-й камере |
| T\_data\_3\_2 | 2-й датчик в 3-й камере |
| T\_data\_3\_3 | 3-й датчик в 3-й камере |
| T\_data\_4\_1 | 1-й датчик в 4-й камере |
| T\_data\_4\_2 | 2-й датчик в 4-й камере |
| T\_data\_4\_3 | 3-й датчик в 4-й камере |
| T\_data\_5\_1 | 1-й датчик в 5-й камере |
| T\_data\_5\_2 | 2-й датчик в 5-й камере |
| T\_data\_5\_3 | 3-й датчик в 5-й камере |
| H\_data | Высота слоя |
| AH\_data | Влажность сырья |

Качество продукции измеряется в лаборатории по пробам, которые забираются каждый час, данные по известным анализам содержатся в файле Y\_train.csv. В файле указано время забора пробы, проба забирается на выходе из обжиговой машины.

Вы договорились с заказчиком, что оценкой модели будет являться показатель MAE, для оценки модели необходимо сгенерировать предсказания за период, указанный в файле Y\_submit.csv (5808 предиктов).

Все необходимые данные лежат по ссылке:

<https://www.dropbox.com/s/lo2w549fv8lo3oc/Test2.7z>