ML Task Part 2:

Now that you have completed Part 1, you understood that we are dealing with products that move along a production line. In a traditional Machine Learning problem, we envisage a fitting function for the output, say Y, for given inputs, say X, such that Y = f(X). In this scenario we would like you to develop a similar model where Y is now the probability of having a defect in the product for the given machine variables X. However now X constitutes of machine variables information in time.

In the folder **MachineVariables_Data**, we have information of the machine variables as the product traverses along the production line (as shown in the schematic below). For example, file **'0.csv'** has the variables information with respect to time for the product with ID '0'. The file '**ProductID_ClassLabel**' has the classification label assigned to each of the product at the end of the production line.

For example, 'Good' represents a 'tidy' product, while 'Lighting' represents product with no available information of the label and all the other labels represent an 'untidy' product.

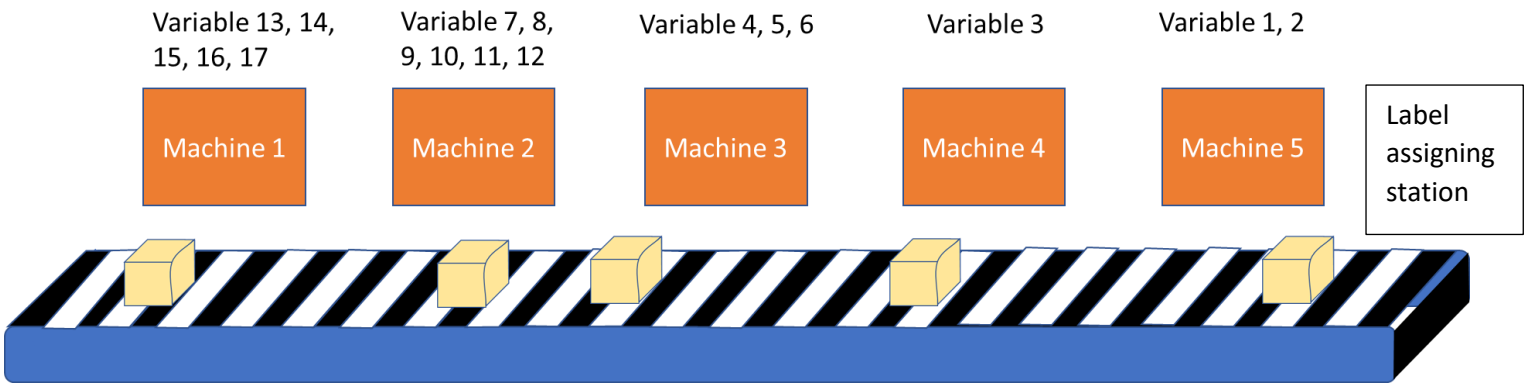
The task in here is to fit an ML model of your choice that can predict either the product is 'Tidy' or 'Untidy', given the Machine variables information with time.

NOTE:

This is a binary classification problem and make sure you don't include products with classification label as Lighting.

HINT:

1. It can be noticed that majority of the variables do not deviate much from their mean value.
2. To have a better accuracy of the model, you should understand that the data is imbalanced.



Questions:

1. Develop a machine learning that uses all the time information of the variables and predicts the probability of a product being 'Tidy' or 'Untidy'.
2. We highly encourage trying out multiple models and comment on the accuracy and other high-level metrics.
3. Identify the critical variables that are affecting the product.
4. Briefly report your findings and mention any other approaches you would have tried given ample time.
5. We would also like to here about how you would extend this to a real time scenario and how you would deploy the model