Problem Definition

1. Consider the data (refer the dataset) provided in the attachment, design and develop a classification model using ML/DL concept of your choice.
2. Candidate is expected to create a Jupyter notebook and share via e-mail for verification (no google drive link or any other means are accepted). To know, how candidate is good in programming & code documentation.
3. Expected that, the F1-Score should be more than 0.95 on the train and test set

**Note:**

1. Any pre-processing operations are allowed to use to design the model (To understand how candidate is good in data pre-processing operations)
2. The split of the data should be as follows – train (80%) and test (20%)
3. The model can be either ML / DL based (To know how candidate is good in selecting models for model designing)
4. Once model is created, detail explanation about the model should be provided in a form of document (word document), which explains
   1. Each step and why this step has been chosen in the design of the model
   2. Why the specific ML/DL was selected to train the model?
   3. Analysis of the trained model
   4. Analysis of the inference carried out on the test set
   5. Limitation of the model

(To know how the candidate is good in documentation)

1. The designed model shall be tested by SAMSUNG team on new test set and based on that we expect the model will have at least of F-Score of 0.90).

(To know, how candidate is good is taking care of post model deployment)

1. If you are not able to achieve the expected target of F1-score, mention what are the possible reasons for it in the documentation (To know, how candidate is good in analysing the failure model developments).

**About dataset:**

1. The attached Dataset.csv file contains raw dataset extracted from a cameras attached to some machine during production on some products.
2. The dataset has 2041 columns and 400 rows.
3. Column A is an ID
4. Column B – Column BZM is a data (X-variables)
5. Column BZN – is the target variable (Y-variable)