**Project Proposal**

**Project Title**

Predict the Yield(Pass/fail) of semiconductor manufacturing process

**Introduction**

A complex modern semiconductor manufacturing process is normally under constant surveillance via the monitoring of

signals/variables collected from sensors and or process measurement points. However, not all of these signals are equally valuable in a specific monitoring system. The measured signals contain a combination of useful information, irrelevant information as well as noise. Engineers typically have a much larger number of signals than are actually required. If we consider each type of signal as a feature, then feature selection may be applied to identify the most relevant signals.

**Dataset Description**

signal-data.csv : (1567, 592)

The data consists of 1567 datapoints each with 591 features.

The dataset presented in this case represents a selection of such features where each example represents a single production

entity with associated measured features and the labels represent a simple pass/fail yield for in house line testing. Target

column “ –1” corresponds to a pass and “1” corresponds to a fail and the data time stamp is for that specific test point.

Dataset : Provided in the Files section

**Project Aim/Objective**

We will build a classifier to predict the Pass/Fail yield of a particular process entity and analyse whether all the features are required to build the model or not.

**Tools**

In this project code implementation done using Python Jupyter Notebook. Jupyter notebook is the famous tool which is widely used for data science and machine learning projects in recent time.

**Data Analysis and Modelling**

In data analysis task segregate predictors vs target attributes and checking imbalance data. After this perform train test split to separate train and test data.

For modelling here, I used the supervised machine learning techniques with k-fold cross validation and also apply hyper parameter tuning techniques. To enhance the model performance here I also use the dimensionality reduction techniques.

**Conclusion**

Compare all the models designed with their train and test accuracies and finally selecting the best trained model.