**Wafer-Fault-Detection**

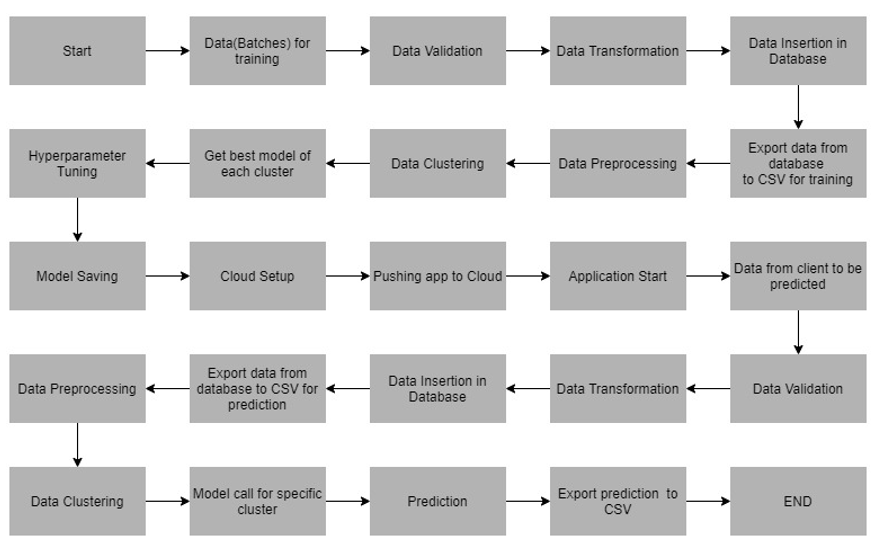
**Problem Statement:**

**To build a classification methodology to predict the quality of wafer sensors based on the given training data as earlier it was difficult to predict which sensor is faulty hence once any sensor is faulty they had to check all the sensors manually and had to shut down then entire production during that time.** 

The inputs of various sensors for different wafers have been provided. In electronics, a wafer (also called a slice or substrate) is a thin slice of semiconductor used for the fabrication of integrated circuits. There are two classes: +1 and -1.

* +1 means that the wafer is in a working condition and it doesn’t need to be replaced.
* -1 means that the wafer is faulty and it needs to be replaced.

**Architecture**

[](https://github.com/Vinay-Kalmoodkar/Wafer-Fault-Detection/blob/master/architecture.png)

**Data Validation**

W perform different sets of validation like filename validation, number of columns, name of the columns,data type of each columns and other kind of validations as per the agreement done between us and client on the schema of filenames and column names etc.

**Data Insertion in Database**

In this step we perform the following things

1. Database Creation and connection
2. Table creation in the database
3. Insertion of files in the table

The above database operation is only done once the validation part is completed.

**Model Training**

1. Data Export from Db - The data in a stored database is exported as a CSV file to be used for model training.
2. Data Preprocessing - Checking for null values, imputation for null values using KNNImputer, removing the features with zero standard deviation etc.
3. Clustering - The idea behind clustering is to implement different algorithms.To train data in different clusters. The Kmeans model is trained over preprocessed data and the model is saved for further use in prediction.
4. Model Selection - After clusters are created, we find the best model for each cluster. Two algorithms are used *RandomForest* and *XGBoost*

**Model Prediction**

Here also all the above steps like Data Validation, Data Insertion in Database, Data Preprocessing and Clustering is performed. Based on the cluster group, the model is loaded and prediction is made.