**Machine Predictive Maintainance**

**Problem Statement**

The scope of project was to predict machine failures using sensor data and optimized corrective maintenance schedule which can identify the indicators of potential failures in production line and can help to prevent some production stops from happening.

**Domain:-** Manufacturing

**Data Gathering**

The data gathering was performed with the help of the SAP Database servers. These SAP Data Servers were connected with the sensors of the production line and the machine failure data was stored in the SAP system.

The data was obtained from the cliants server using SQL quesies.

Ex. Select <column\_names> from <table\_name>

The gathered data contained 10 columns with 10,000 rows

The features were as follows

UDI, Product ID, Type, Air Temperature, Process Temperature, Rotational speed, Torque, Tool wear and Failure Type

**Feature selection**

The unwanted features were dropped i.e. UDI, Product ID etc which was not related with the entire data set

How we have 8 features

**EDA**

During the EDA process we looked for the pattern of the dataset and the outliers.

We used Corelation plot, Heatmap ,

There was not any relation obtained between the features. But these features were able to classify and so we decided to use classification model for training.

While looking for the outliers the features like rotational speed and torque were having the outliers.

There outliers were obtained by using the IQR method. As the IQR method was resulted to be the best so this method was used.

The outliers were replaced by the mean of the respective features.

**One Hot encoding**

The one hot encoding was used for the feature ‘Type’ as there were characters in this feature.

There were 3 classes in this feature so this feature was converted to 3 separate features.

Now the model size becomes 10 features and 10,000 rows.

**Label Encoding**

The label encoding is used in the feature ‘Failure Type’ as there were characters in tgis feature.

There were 6 classes in this feature and these were replaced by numbers from 0 to 5.

**Feature Extraction**

Here the importance of all the feature with respective to the target feature is calculated and all the features were important for obtaining the target.

**Model building**

The developed dataset is trained using all the classification algorithms and the best algorithm is selected.

Here we achieved the highest accuracy of 97.56% using Random Forest algorithm.

**Model training**

The model it trained using default parameters.

Hyperparameter tuning reduces the accuracy so we didn’t used it

**Results**