Detailed Report

On

**Predictive Maintenance for Industrial Equipment Project**

-By Anshika Tyagi

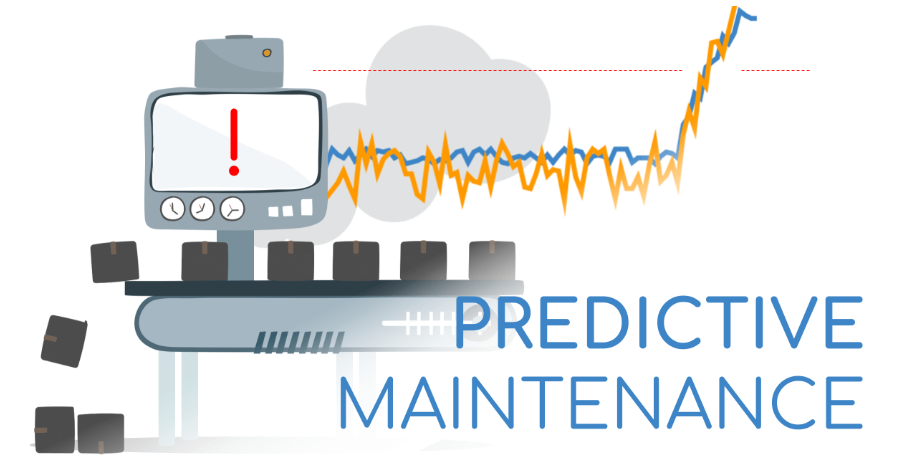
Date – 22 August ,2024

**Problem Description**: Our task is to develop a predictive maintenance system that can forecast potential failures in industrial equipment. The goal is to analyze historical equipment data to predict future maintenance needs, thereby preempting equipment failures.

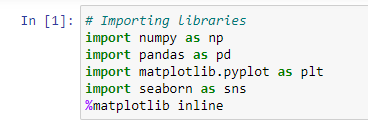
**Overview :** This report is about Predictive Maintenance for Industrial Equipment Project , a machine learning model which is performed on Predictive Maintenance dataset which is taken from website Kaggle in Jupyter Notebook.

The report presents entire process which involved in building the model,model selection criteria, challenges encountered and key insights obtained**.**

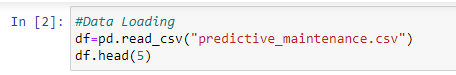
**Dataset link :** **https://www.kaggle.com/datasets/shivamb/machine-predictive-maintenance-classification**

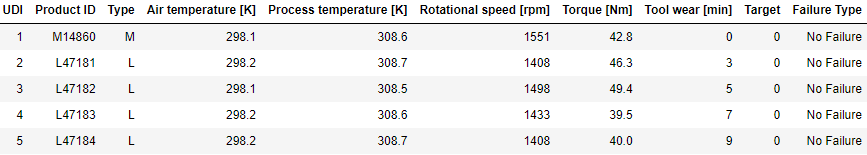


* **Steps or Process which is used for building model :**
* Importing Necessary Libraries **:** Firstly ,import all python libraries such as numpy ,pandas ,matplotlib ,seaborn and sklearn etc.

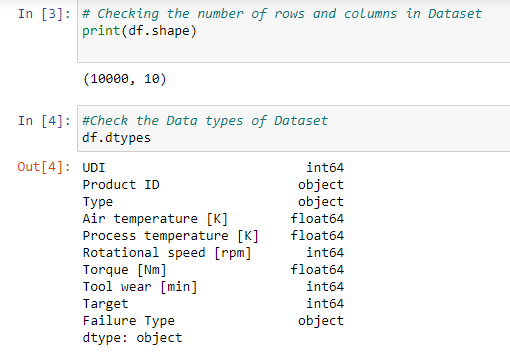


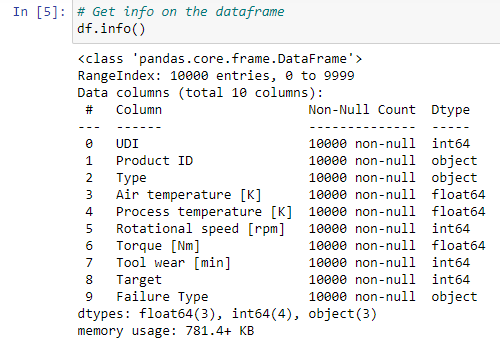
* Loading Dataset : Then load dataset in the jupyter notebook which is in csv file.

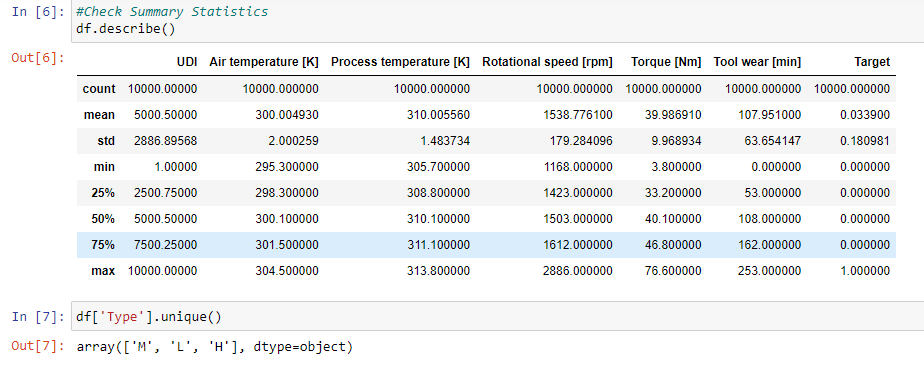




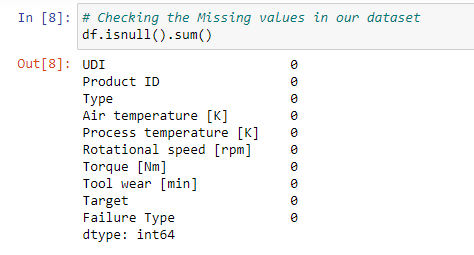
* Know about dataset : After loading , we try to know about the dataset such as number of rows and columns ,their datatypes and summary statistics.



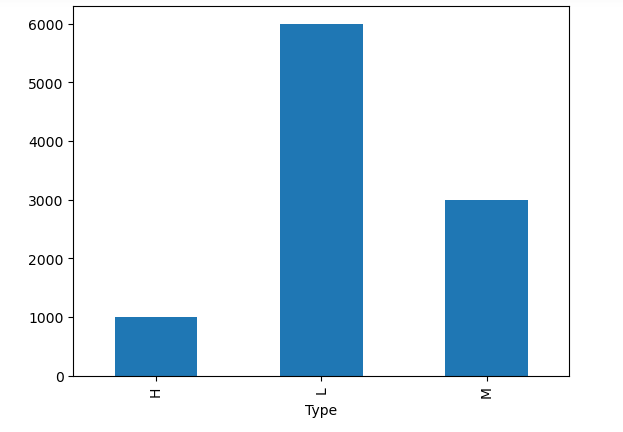




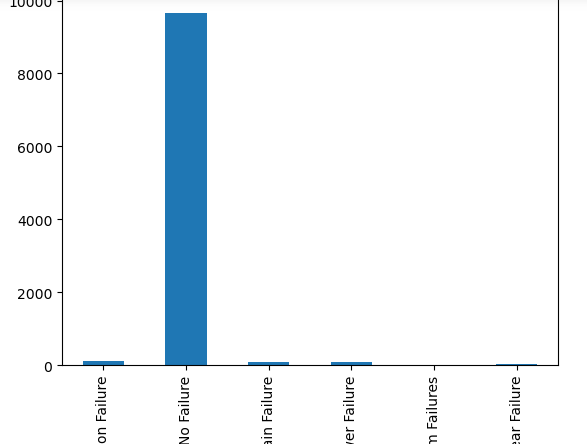
* Data Preprocessing and Visualization : In this step, clean and preprocess the data to make it suitable for analysis and modeling.This includes handling missing/null values,encoding categorical variables , data visualization and extract new features from existing ones.

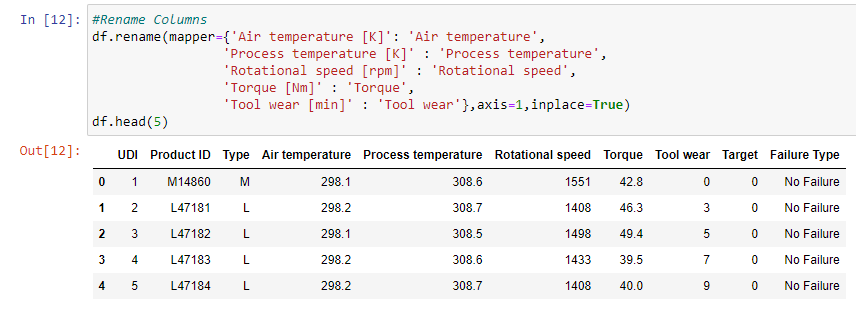


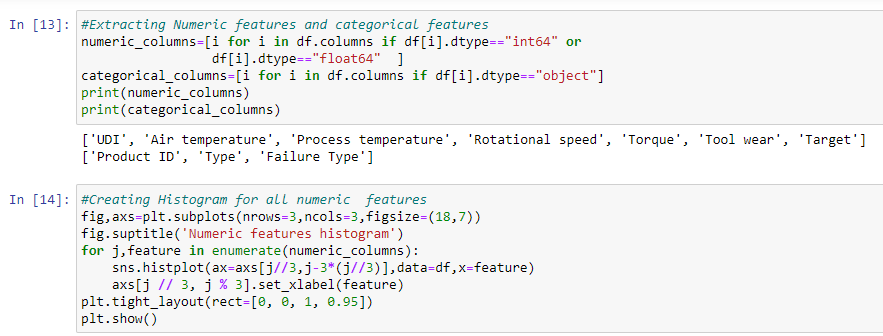


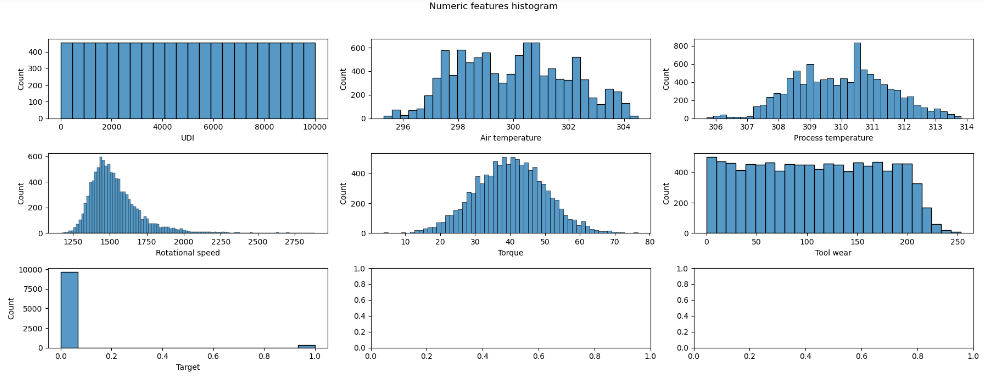






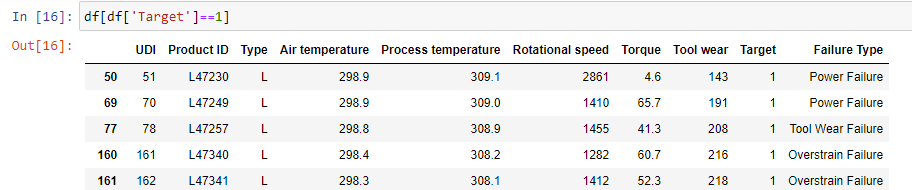


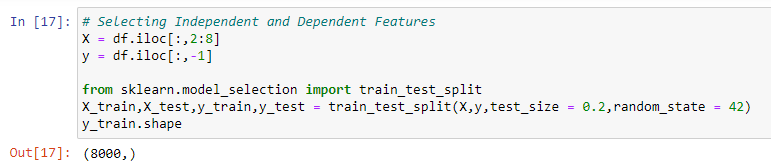


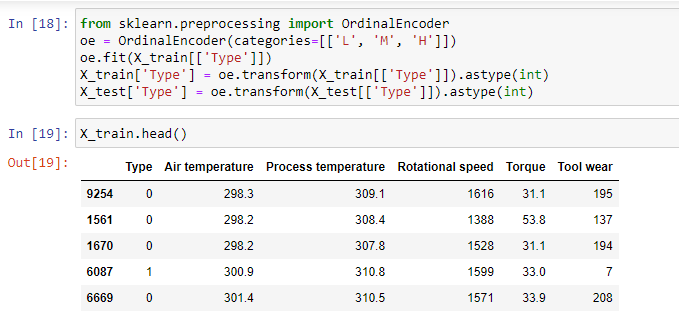


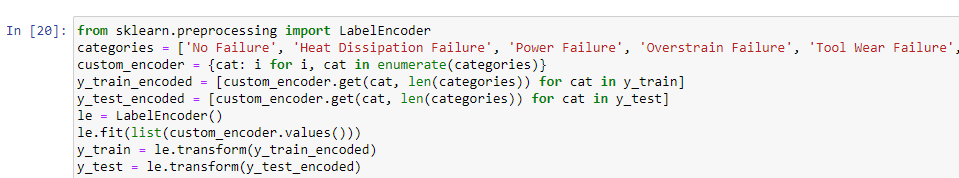


* Feature Selection and Ordinal Encoding : In this step ,  we have column Target in our dataset which has 2 values 0 or 1, which indicates failure and no failure. So , we take those rows in which Target value is 1, in order to predict type of failure occur. Also we will convert categorical columns to numeric columns.

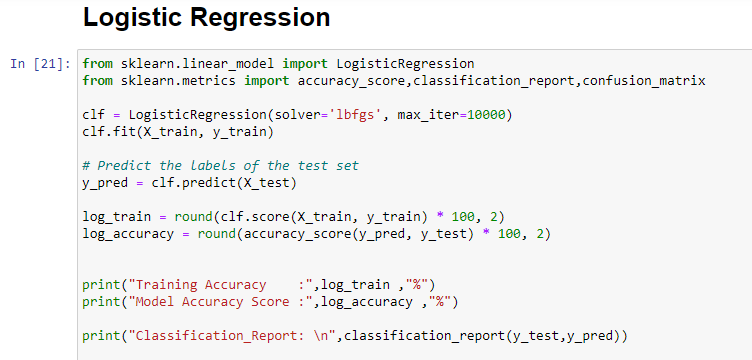


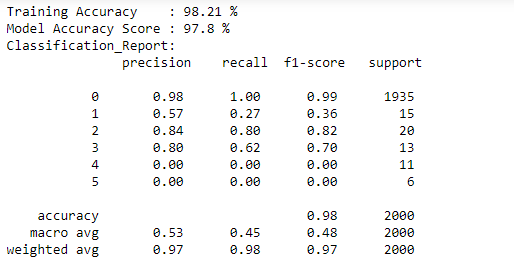


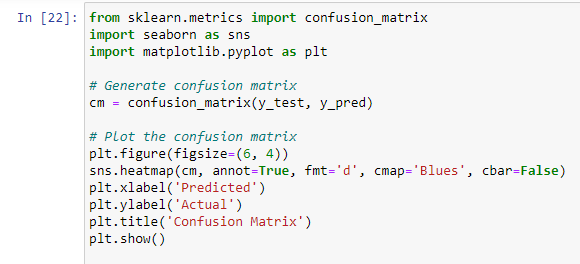


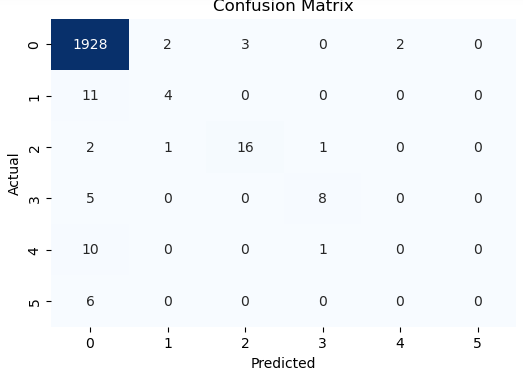


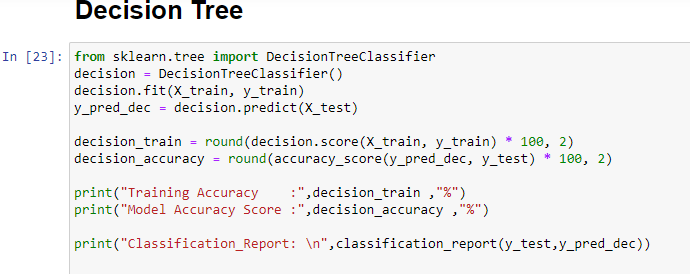
* Building Models and Evaluation : After extracting features , implement various machine learning models ,as it multi classification problem .So we use Logistic Regression , Decision tree,Random Forest and Support Vector Machines etc to predict failure type.Evaluate the performance of each model.

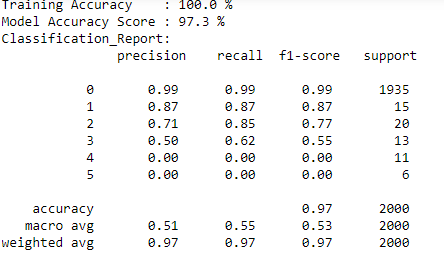


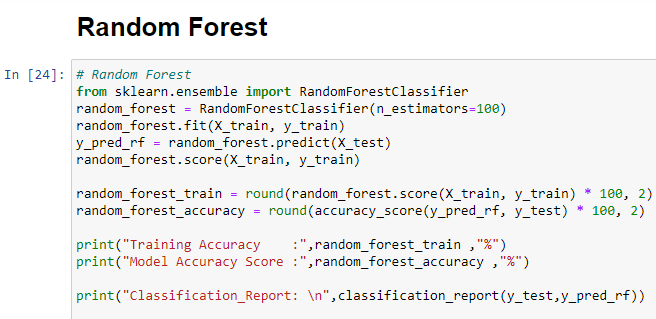


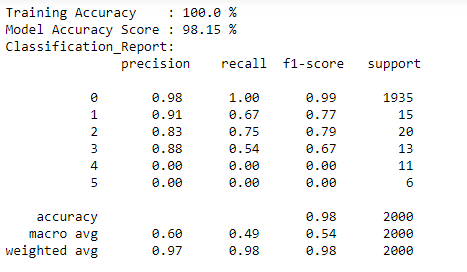


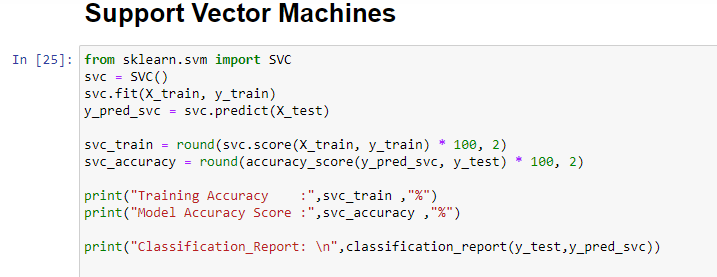


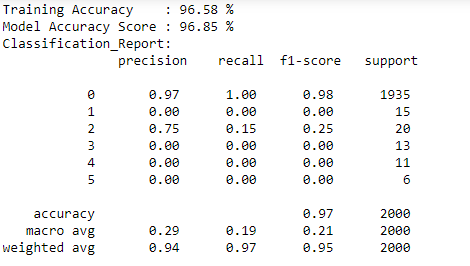




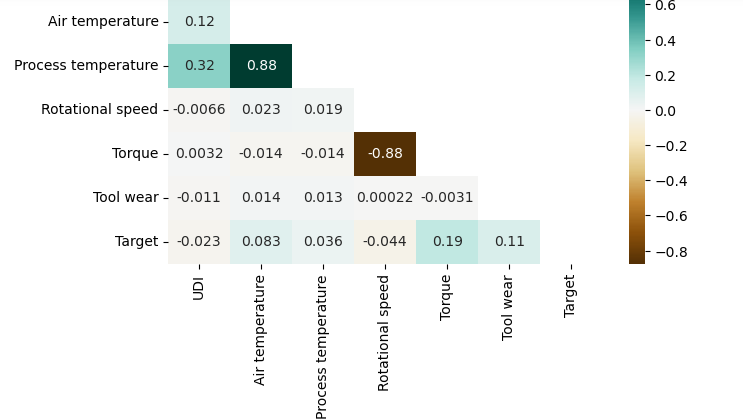












* **Model Selection Approach :** As it is multi Classification problem means we have to predict the failure type whether the it is Power failure,Overstrain failure , or tool wear failure etc .For this we use these algo to implement this…
  + - **Logistic Regression -**The accuracy of LR is 98% obtained.

* **Decision Tree** - The accuracy of DT is 100% obtained.
* **Support Vector Machine (SVM)** – The accuracy of85% obtained.
* **Random Forest** – The accuracy of RF is 100% obtained.

So , from above Decision Tree and random Forest has better accuracy, so it is more effective to predict the failure type .

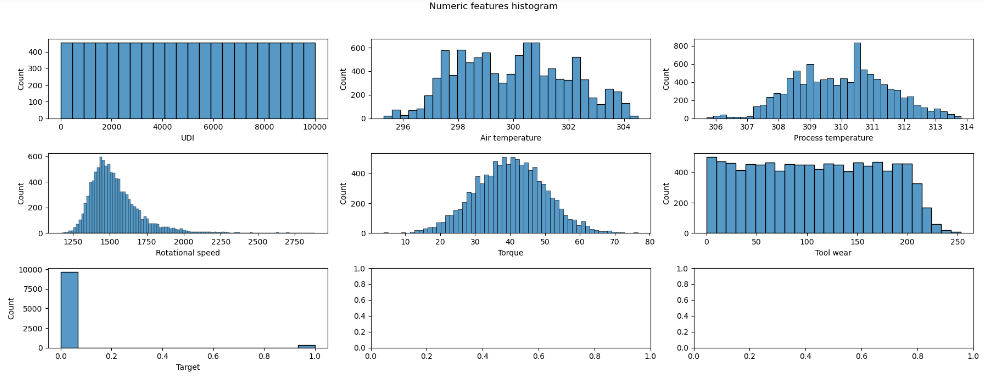
* **Challenges Occurred :**
* **Dealing with multi Class classification problem :**AS it is multi class classification problem ,to deal with the multiple classes we must select appropriate algorithm.

Approach to Handle this Challenge : As we know , Algorithms such as Random Forest and SVM are good enough to deal with multiple classes.

* **Dealing with columns** : In this dataset ,columns like Air temperature [K], Torque [Nm] etc . we have to rename these columns with correct ones.
* **Dealing with Categorical Columns :** It is also one challenge to deal with categorical columns in the dataset , Because Categorical columns with a large number of unique values can lead to sparse representations and increase the dimensionality of the dataset, making it more complex to process.

Approach to Handle this Challenge **:**Techniques like Label Encoding or Ordinal Encoding applied on categorical columns with a manageable number of unique values**.**We applied Ordinal Encoding as it is ordinal data

* **Insights obtained:**
* **Air Temperature**: The distribution is relatively symmetric and spread across a range between 296 and 304. This suggests consistent air temperature readings with a slight concentration around the center of the range.
* **Process Temperature**: The distribution is bell-shaped, indicating a normal distribution with most values clustered around the mean, around 311. This could suggest stable process temperature during operations.
* **Rotational Speed**: The distribution is right-skewed, with most observations concentrated between 1250 and 1750. This implies that in most cases, rotational speeds are on the lower side.
* **Torque**: The torque feature also shows a right-skewed distribution, with most values falling between 10 and 40. This skewness indicates that higher torque values are less common.
* **Tool Wear**: This histogram shows a uniform distribution, similar to UDI. This uniformity suggests that tool wear might be an artificially generated feature or that it is evenly spread across all possible values.
* **Target**: The target variable is highly imbalanced, with most instances concentrated at 0 (non-failure). This imbalance indicates that failures (1) are rare events in the dataset, which could influence model training.



* **No Failure :** The category No failure has the highest count that shows many products do not face any kind of failure.

|  |
| --- |

