Name: Renuka Hatwalne

Title: Predictive Maintenance for Industrial Equipment Using Machine Learning

- **Features are there in Dataset: Temperature, Pressure, Vibration, Humidity, Failure**
  - There are 5 features available in experimental dataset.
- **Research About Feature:**
  - There are 5 features available in experimental dataset.
  - **Temperature, Pressure, Vibration, Humidity** feature has continuous range of values.
  - **Failure** has two classes (0, 1).
  - Here dependent variable is Failure.
  - All 4-variable act as independent variable.
- **Performance metrics used.**

  Mean squared error

  R2score

  Accuracy
- **Data cleaning & data preprocessing techniques you can apply on dataset.**

  Null value handling as some of feature contains null value.

  Handled it by replacing it with mean.

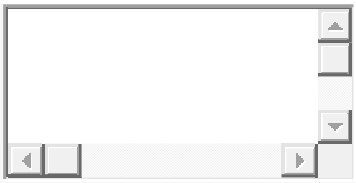
  Feature engineering

To remove duplicate records or rows to ensure data consistency but there is no duplicate record found.

- **Possible algorithms applied on dataset.** Split the dataset into training and testing sets.
  Train multiple models using different algorithms and hyperparameters.
  Random Forest, Gradient Boosting, Support Vector Machines
- **Techniques applied for hyperparameter.**
  Best accuracy i.e.0.95 is observed by SVC using “rbf” kernel out of this- ('linear', 'poly', 'rbf', 'sigmoid')
- **Experiment with ensemble techniques to improve model performance is performed on model.**

- Individual Classifier Accuracies:
- Random Forest: 1.0
- Gradient Boosting: 1.0
- Logistic Regression: 1.0
-
- Ensemble Classifier Accuracies:
- Bagging (Random Forest): 1.0
- Boosting (Gradient Boosting): 1.0
- Stacking (Voting Classifier): 1.0

- In [100]:

- 

- Evaluate the performance of each model using appropriate metrics (e.g., RMSE, MAE) is done on given dataset.
- Compare the performance of different models and identify the best-performing one.

- Random Forest Regressor:
- RMSE: 0.037193189340702336
- MAE: 0.01366666666666667
-
- Gradient Boosting Regressor:
- RMSE: 0.06593367853774076
- MAE: 0.03101948917402032
-
- Linear Regression:
- RMSE: 0.1926494080135646
- MAE: 0.1463769496530853

```
Accuracy Scores:
Random Forest: 1.0
Gradient Boosting: 1.0
Logistic Regression: 1.0
Best Performing Model: Random Forest
```

- Perform cross-validation to assess the model's generalization ability.

```
Cross-Validation Scores for Random Forest Classifier: [0.96666667 0.
96666667 0.93333333 0.96666667 1.        ]
Mean Cross-Validation Score for Random Forest Classifier: 0.96666666
66666668

Cross-Validation Scores for Support Vector Machine (SVM): [0.9666666
7 0.96666667 0.96666667 0.93333333 1.        ]
Mean Cross-Validation Score for Support Vector Machine (SVM): 0.9666
666666666666
```

```
Cross-Validation Scores for Gradient Boosting Classifier: [0.9666666
7 0.96666667 0.9        0.96666667 1.        ]
Mean Cross-Validation Score for Gradient Boosting Classifier: 0.9600
000000000002
```

- Interpret the model's predictions & provide actionable insights for maintenance planning.

Summary:

This project aims to utilize machine learning techniques for predictive maintenance in industrial equipment. By analyzing the data such as temperature, pressure, vibration, and humidity, the system predicts potential equipment failures before they occur, enabling proactive maintenance actions. The details of the tasks performed are as follows.

Data loading:

The dataset is loaded into panda dataframe using read_csv method. The dataset contains 5 columns – Temperature, Pressure, Vibration, Humidity and Failure.

Data Exploration:

Starting with the understanding of the data, *info()* method returns the datatypes of each columns, i.e., all columns are in *float64* datatype except the last column, Failure in represented in *int64.*

Verify and remove the null values*:* Isnull() method returns the columns having null values. There are few null values present in the given dataset. The values are replaced by the mean values of the respective columns using *fillna(data.mean())* methods.

Remove Duplicates: drop_duplicates() method drops the duplicate rows as duplicates create redundant data points.

Feature Engineering: Before preparing training and inference datasets, the data normalization is applied using *zscore.* We also studied the correlation of the columns in the datasets and better understanding of the datasets. For this project, no new features have been added to the original feature list as the acceptable accuracy range is achieved.

- Discuss potential limitations of the model & ways to improve its accuracy & robustness.

Given data size is relatively smaller, and therefore the model may overfit. Having a larger dataset would provide the better understanding the accuracy of the model. By analyzing the data we see t hat only 5% of the data falls into the category of failure hence any model predicting accuracy les s than 95% is futile.