REPORT

Predictive Maintenance for Industrial Equipment

Project Overview:

The aim of this project was to develop a predictive maintenance system for industrial machinery. By analyzing key factors contributing to machine failure, the goal was to create a model that could accurately predict equipment failures and provide actionable insights to optimize maintenance schedules. By doing so, we aim to reduce unplanned downtime, optimize operational efficiency, and improve overall equipment reliability.

Objectives

1. Understand the factors that contribute to equipment failure.
2. Build a predictive model to forecast potential failures.
3. Provide actionable insights to optimize maintenance schedules and reduce downtime.

Dataset Description

The dataset consists of 363 rows and 8 columns, which capture the operational and environmental conditions of machines over time. Below are the key features:

- **Machine ID**: Unique identifier for each machine.
- **Timestamp**: The date and time when the data was recorded.
- **Temperature**: The operating temperature of the machine in degrees Celsius.
- **Pressure**: The pressure inside the machine in PSI (Pounds per Square Inch).
- **Vibration**: The vibration level of the machine in mm/s (millimeters per second).
- **Operational Hours**: The total number of hours the machine has been in operation.
- **Maintenance History**: A binary indicator (Yes/No) of whether the machine has undergone maintenance.
- **Failure**: A binary indicator (Yes/No) of whether the machine has failed.

Data Exploration and Preprocessing

During the data exploration phase, the following tasks were performed:

1. **Null Values & Duplicate Rows**: No missing data or duplicate rows were found.
2. **Data Types**: Ensured that each column's data type matched the expected values.
3. **Anomalous Data Removal**: During standardization of the Timestamp column, one row had an unusual date that could not be standardized. This row was removed.
4. **Label Encoding**: Binary columns like Maintenance History and Failure were converted into numerical values using LabelEncoder for further processing.

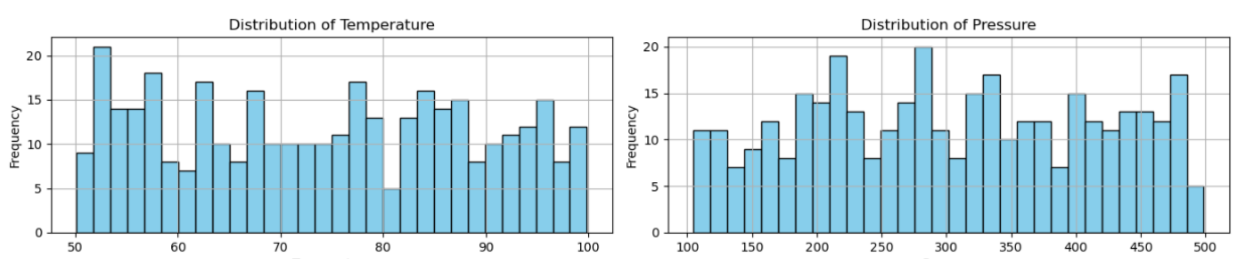
To improve model accuracy, the following features were created:

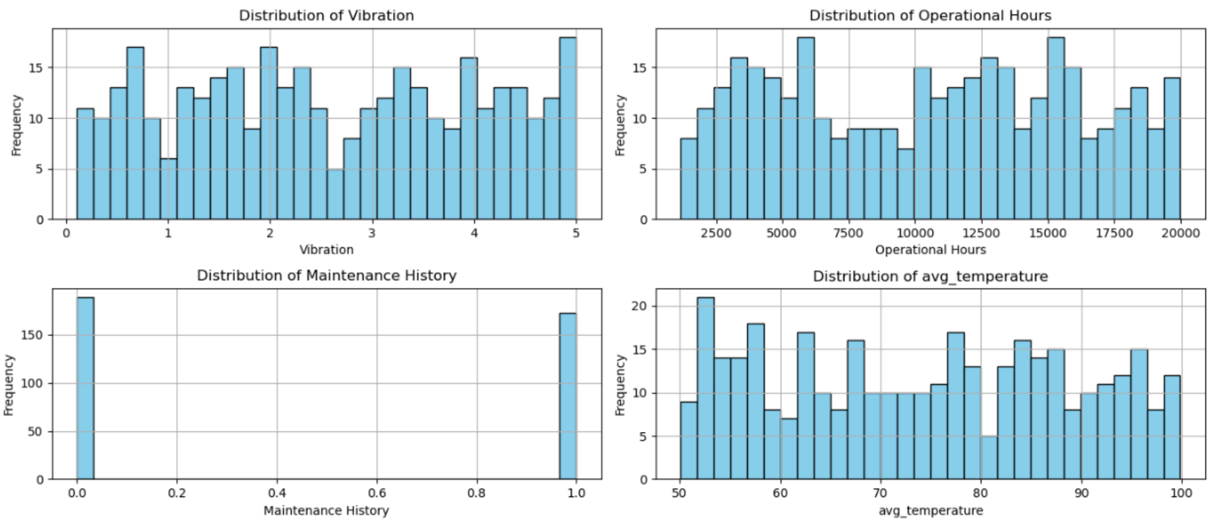
1. **Average Temperature**: Grouped by Machine ID to calculate the average temperature.
2. **Cumulative Vibration**: Calculated the cumulative sum of vibration levels over time.
3. **Cumulative Operational Hours**: A new feature that tracks the cumulative operational hours until failure.

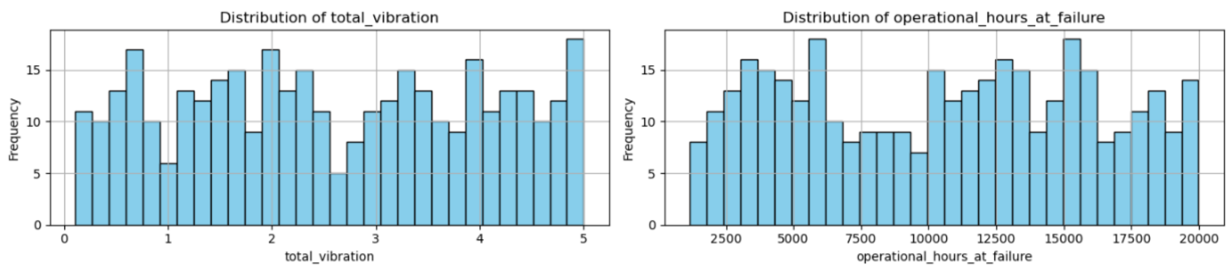
Exploratory Data Analysis (EDA)

Several visualizations and statistical summaries were generated to understand the data better and discover potential patterns contributing to machine failure:

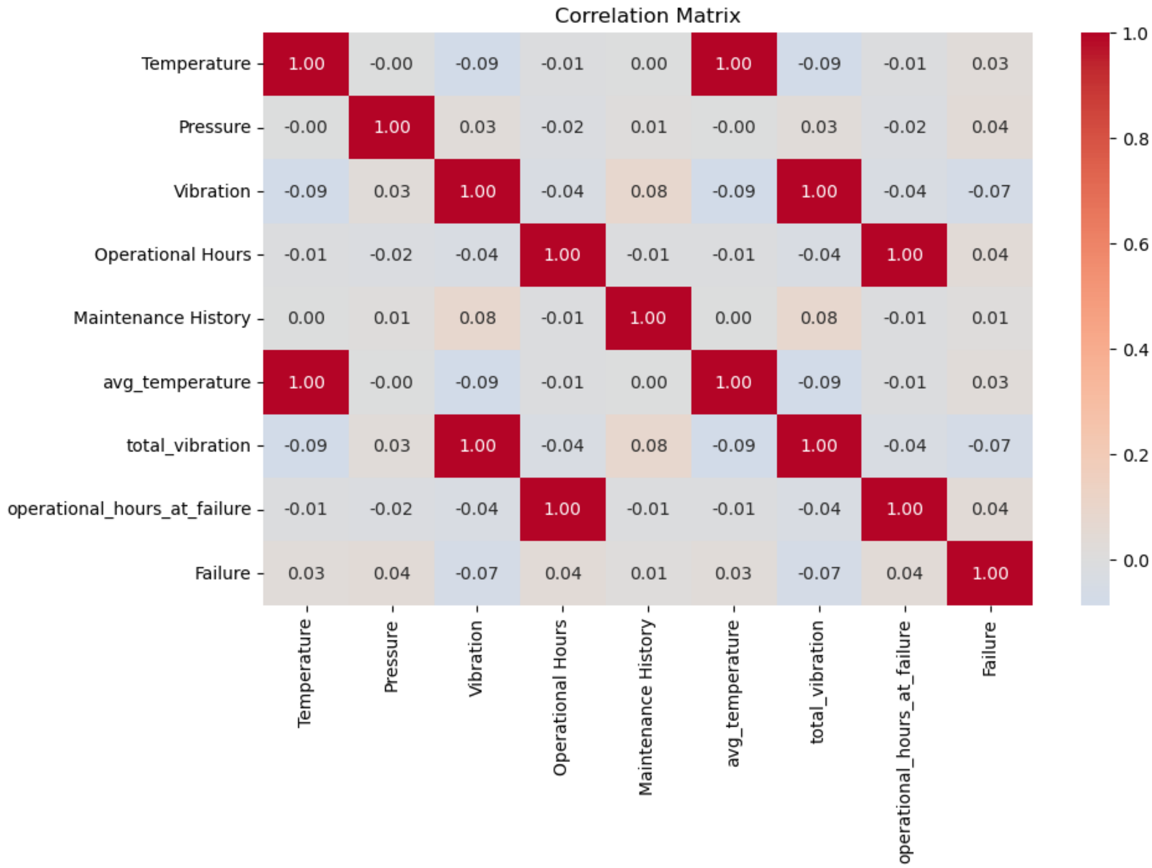
- **Histograms**: Plotted for all numerical features to understand their distributions.



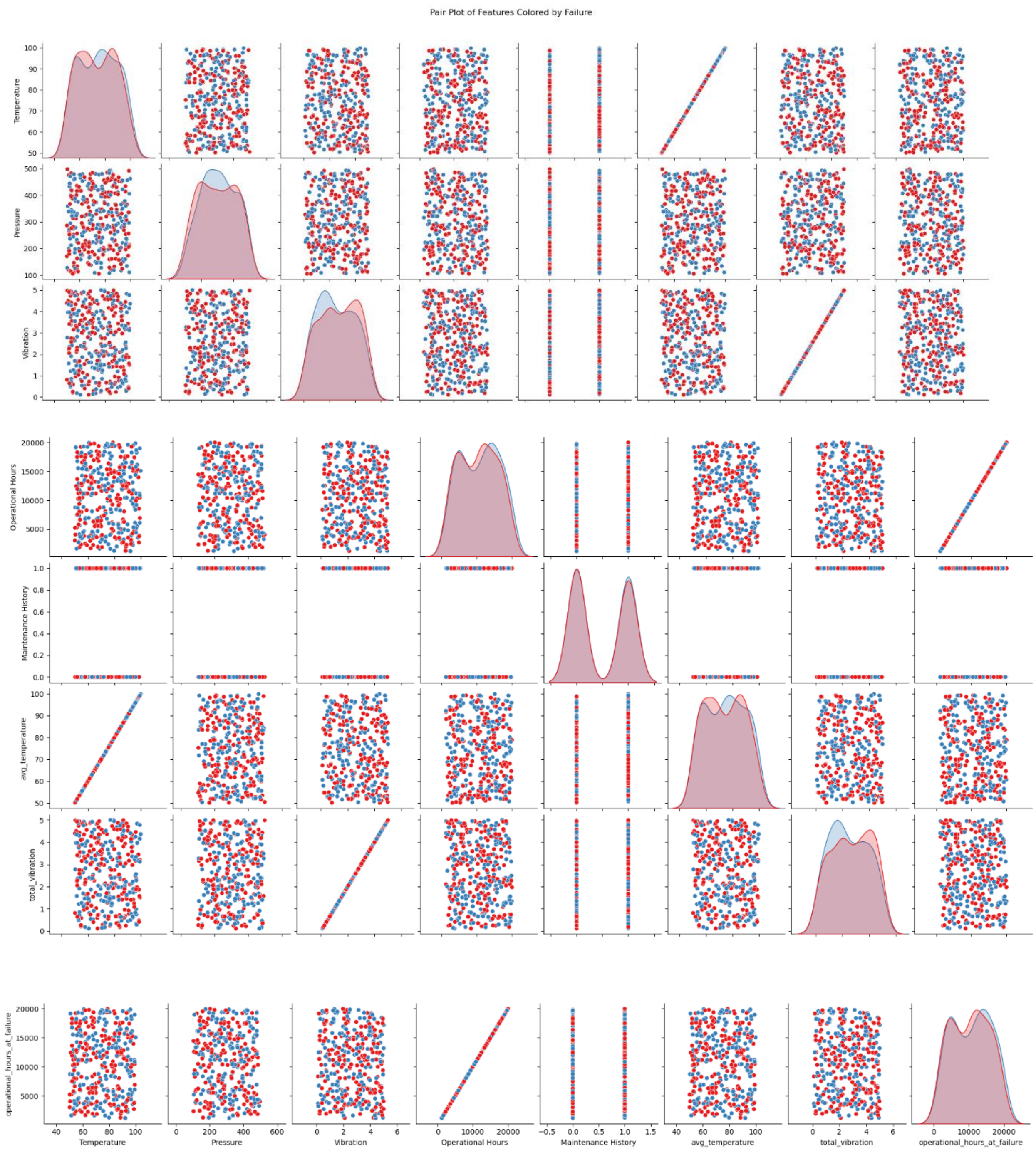

Distribution of Vibration
Frequency
Vibration
Distribution of Operational Hours
Frequency
Operational Hours
Distribution of Maintenance History
Frequency
Maintenance History
Distribution of avg_temperature
Frequency
avg_temperature


Distribution of total_vibration
Frequency
total_vibration
Distribution of operational_hours_at_failure
Frequency
operational_hours_at_failure

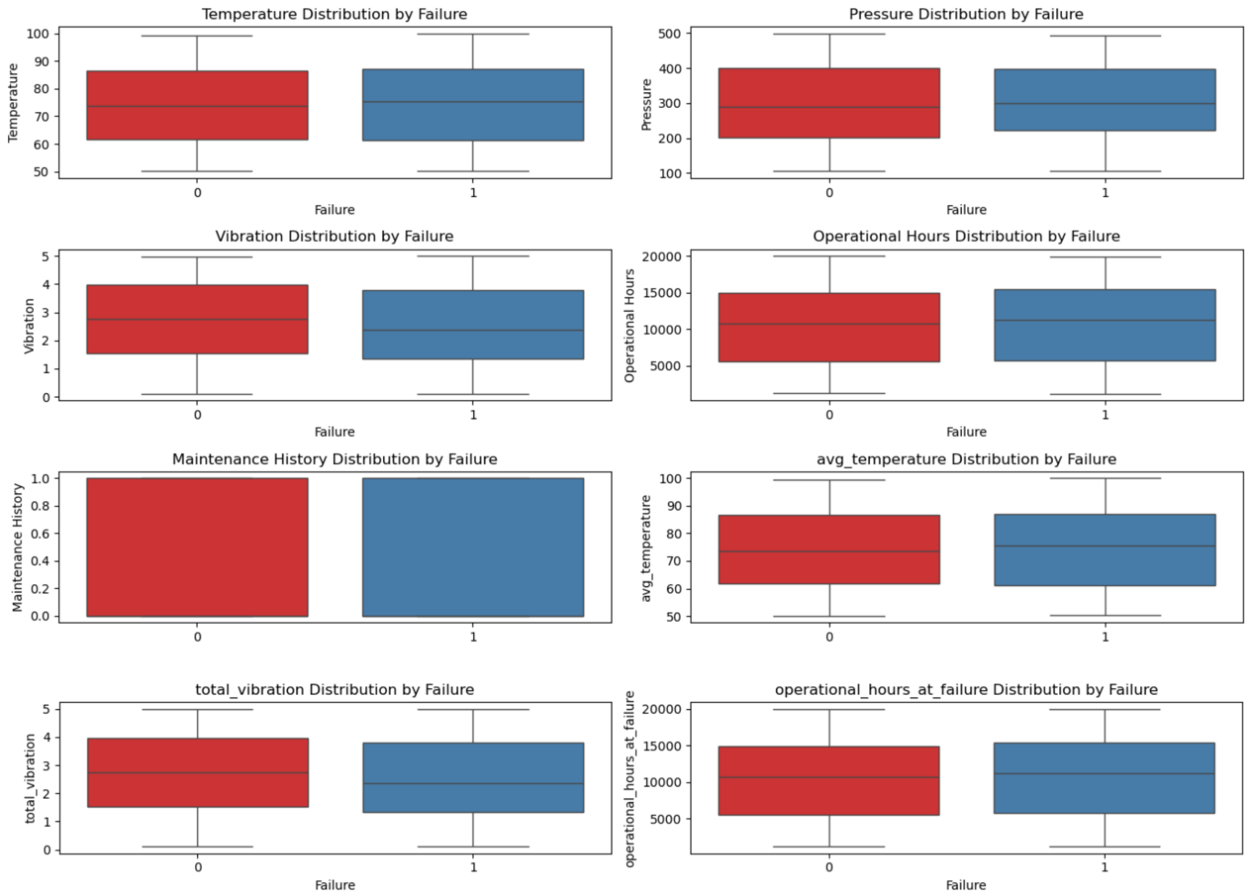
- **Correlation Matrix**: Visualized correlations between features.

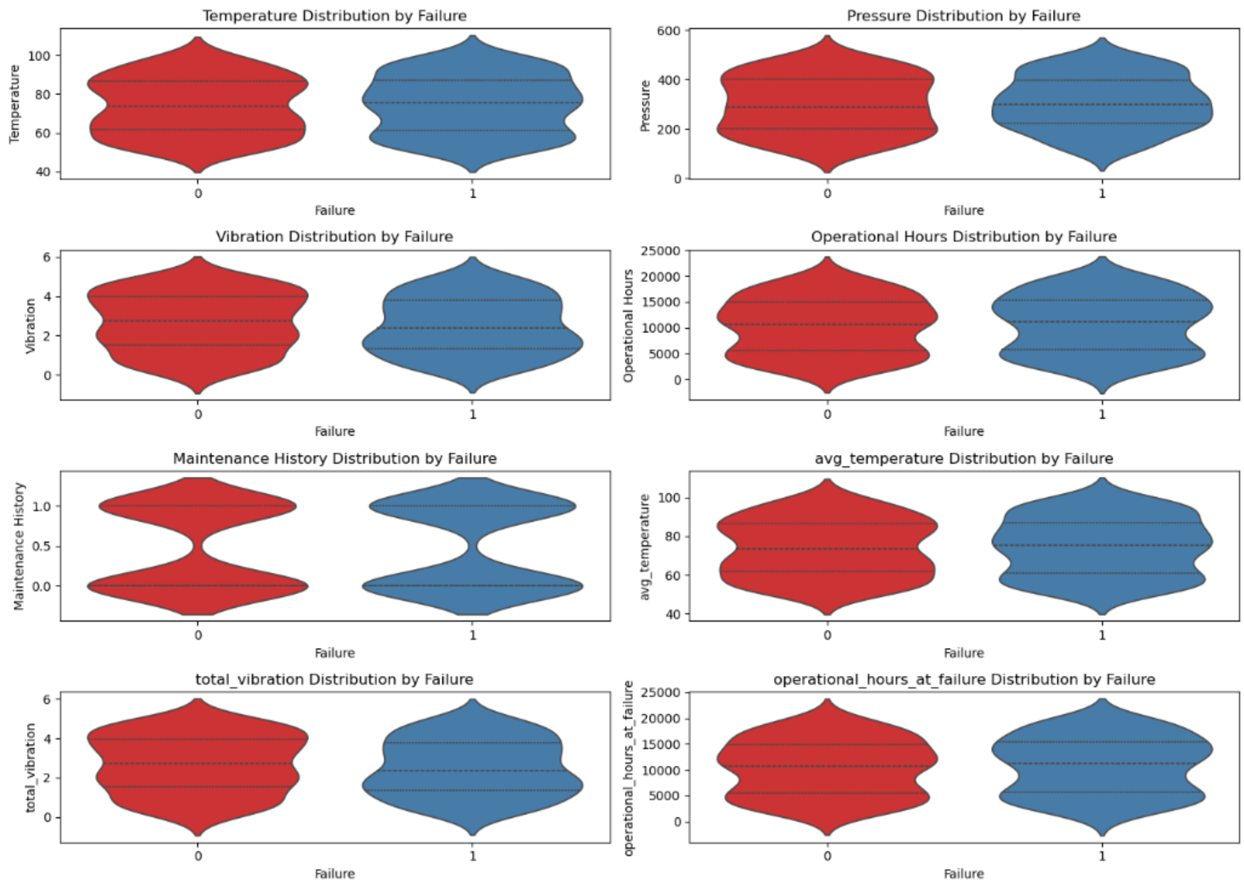


- **Pairplots**: Generated to observe interactions between different numerical features.



- **Boxplots & Violin Plots**: Used to detect outliers and visualize the distribution of features.





Modeling and Evaluation

Several machine learning algorithms were employed to predict failures. The data was split into training and testing sets, and scaled for optimal performance.

The following models were tested, and their evaluation metrics are summarized below:

Logistic Regression

1) Accuracy: 0.49
2) Precision: 0.56
3) Recall: 0.44
4) F1 Score: 0.49

Random Forest Classifier

1) Accuracy: 0.52
2) Precision: 0.57
3) Recall: 0.59
4) F1 Score: 0.58

Gradient Boosting Classifier

- Accuracy: 0.47
- Precision: 0.53
- Recall: 0.46
- F1 Score: 0.49

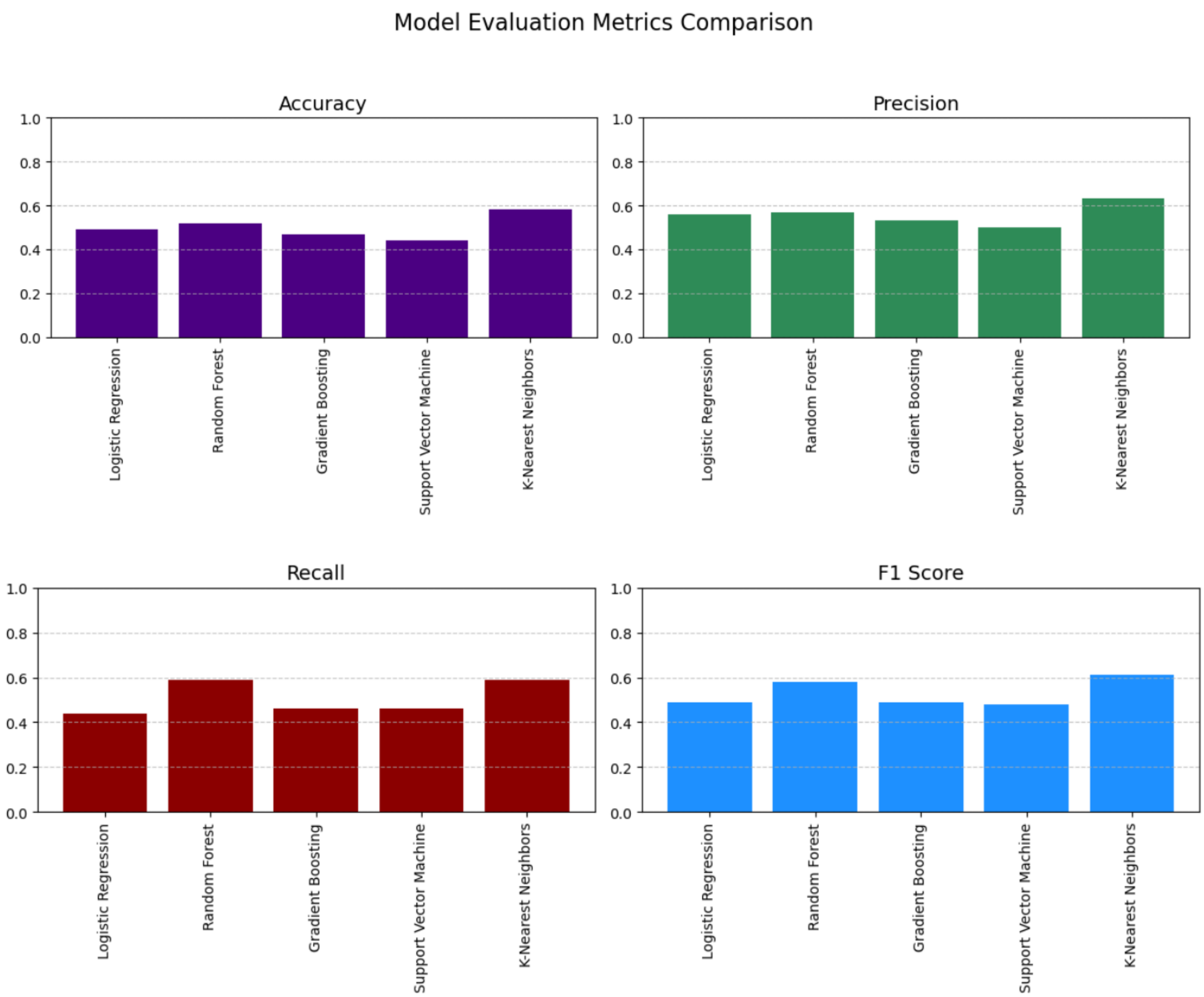
Support Vector Machine (SVM)

- Accuracy: 0.44
- Precision: 0.50
- Recall: 0.46
- F1 Score: 0.48

K-Nearest Neighbors (KNN)

- Accuracy: 0.58
- Precision: 0.63
- Recall: 0.59
- F1 Score: 0.61

Based on the evaluation metrics, K-Nearest Neighbors (KNN) provided the highest accuracy and F1 score. Therefore, it was chosen as the final model for predicting machine failures.



Maintenance Prediction Results

The model was tested on the dataset to predict machine failures and provided accurate results for both failure and non-failure scenarios (on test values):

No Failure Scenario:

1) **Predicted Class**: No Failure
2) **Recommended Action**: Operate normally, but monitor closely.
3) **Probability of Failure**: 40%
4) **Probability of No Failure**: 60%

Failure Scenario:

1) **Predicted Class**: Failure
2) **Recommended Action**: Schedule maintenance soon.
3) **Probability of Failure**: 60%
4) **Probability of No Failure**: 40%

Conclusion and Business Insights

The predictive maintenance model allows the company to:

- Predict failures at an early stage, enabling timely preventive actions.
- Optimize maintenance schedules, reducing unplanned downtime.
- Improve operational efficiency by focusing resources on machines most at risk.

By using this predictive system, the company can significantly reduce the risk of equipment failure, improve maintenance planning, and ultimately enhance operational productivity.