Equipment Failure Prediction in a Manufacturing Plant

# Executive Summary:

This project tackles the challenge of predicting equipment failure in a manufacturing plant using machine learning techniques. The ability to forecast failures before they occur is essential for minimizing downtime and reducing maintenance costs. By analyzing operational, usage, and environmental data, we aim to identify the key factors that contribute to equipment failures and provide actionable insights to optimize maintenance strategies.

# Project Objectives:

* **Predict equipment failure** using historical operational data.
* **Identify factors** that influence equipment failure.
* **Provide insights** to assist in preventive maintenance and reducing downtime.

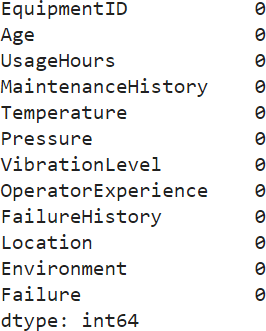
# Dataset Overview:

The dataset includes various features representing the operational data of equipment, structured as follows:

|  |  |
| --- | --- |
| **Feature** | **Description** |
| **EquipmentID** | Unique identifier for each piece of equipment |
| **Age** | Age of the equipment (in years) |
| **UsageHours** | Total hours the equipment has been used |
| **MaintenanceHistory** | Number of maintenance activities performed in the last year |
| **Temperature** | Average operating temperature of the equipment (°C) |
| **Pressure** | Average operating pressure (psi) |
| **VibrationLevel** | Average vibration level (mm/s) |
| **OperatorExperience** | Operator’s experience (in years) |
| **FailureHistory** | Number of failures in the past |
| **Location** | Location of the equipment in the plant (Section A, B, etc.) |
| **Environment** | Operating environment (Humid, Dry, etc.) |
| **Failure** | Target variable indicating if the equipment failed (Yes/No) |

**Data Preprocessing Steps**

1. **Missing Value Treatment**: Applied strategies to handle missing data and ensure integrity.

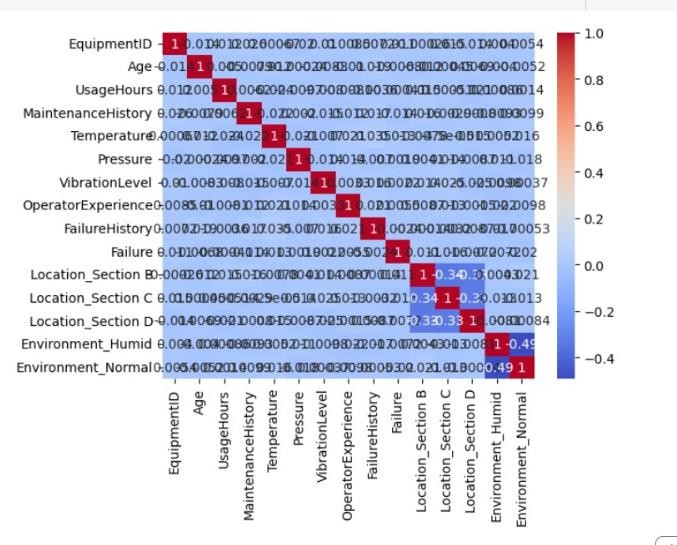
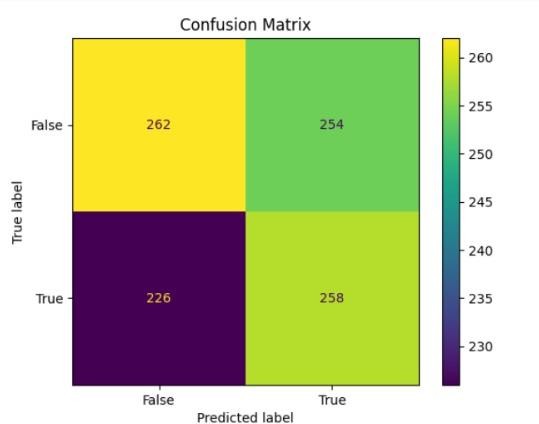


1. **Feature Encoding**: Transformed categorical features using One-Hot Encoding for the

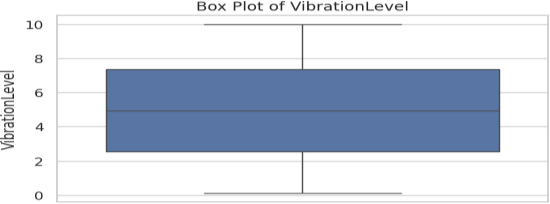
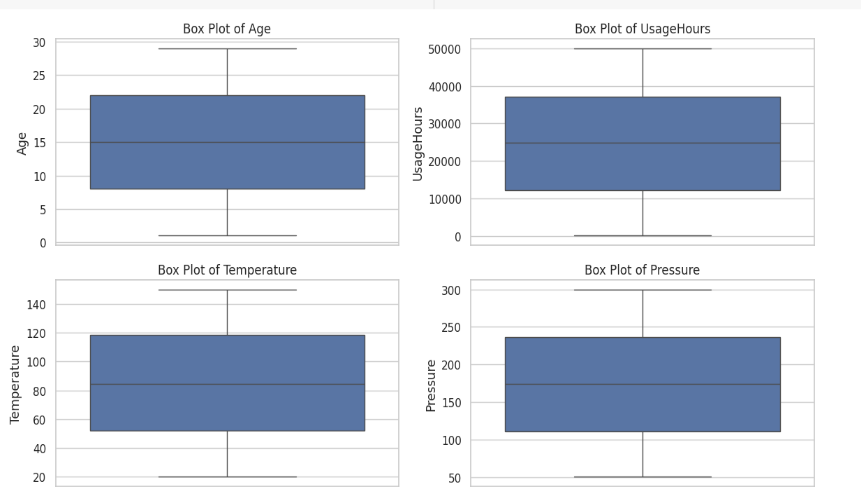
Location and Environment columns.

1. **Outlier Detection and Treatment**: Used the Interquartile Range (IQR) method to detect and remove outliers.

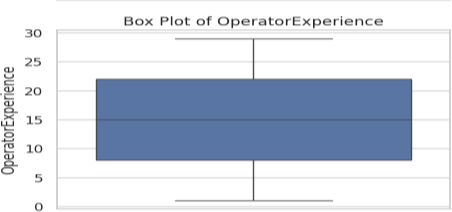
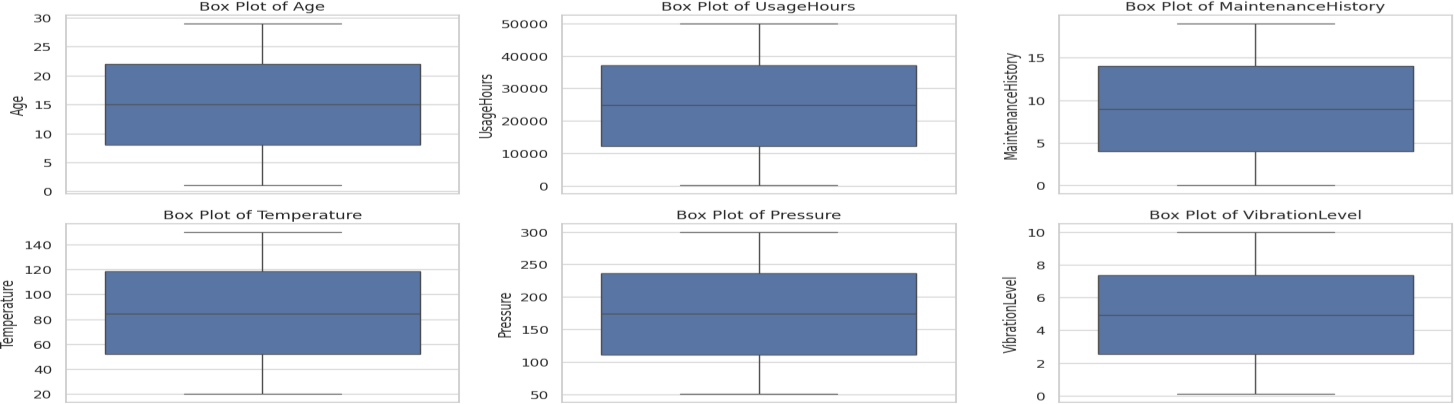
**Confusion Matrix**:



**Before Removal of Outlier:**



**After Removal of Outlier:**



1. **Feature Scaling**: Standardized continuous features like Age, UsageHours, Temperature, Pressure, and VibrationLevel to ensure uniformity.

# Feature Engineering:

We created additional features to enhance the model’s ability to predict failures:

* + **UsageRate**: Calculated as UsageHours / Age, representing the intensity of equipment use.

# Model Development and Evaluation:

We implemented several machine learning models to evaluate their performance in predicting equipment failures:

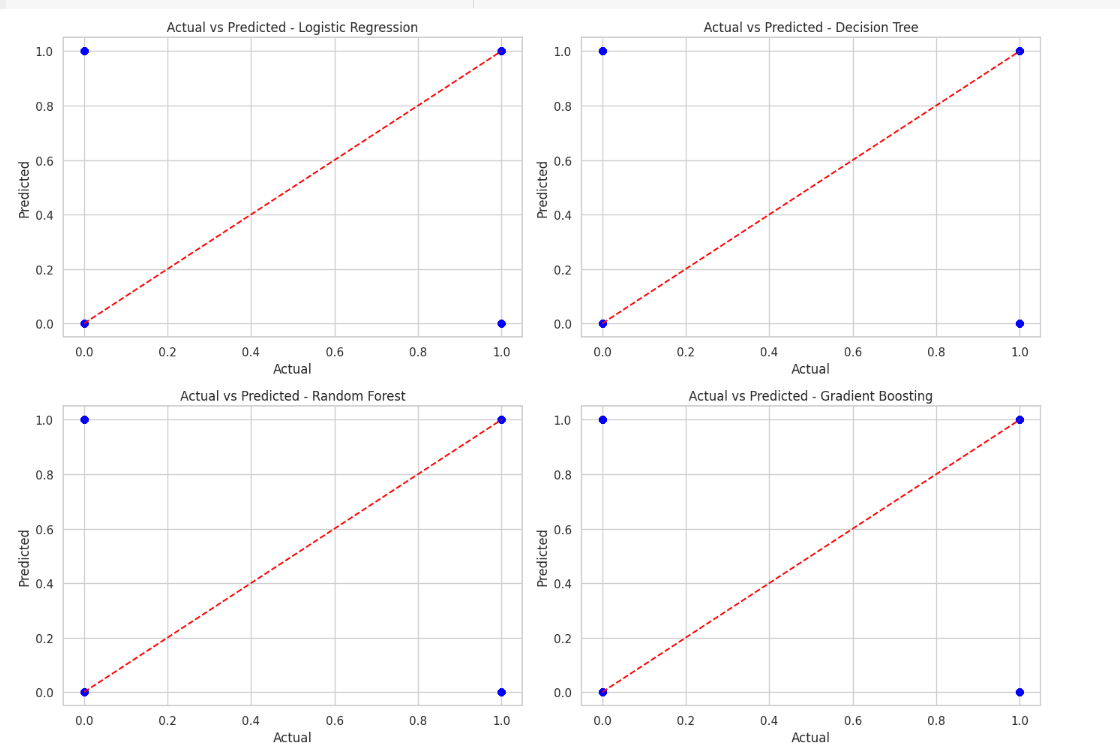
# Models Explored:

* + **Logistic Regression**
  + **Decision Tree Classifier**
  + **Random Forest Classifier**
  + **Gradient Boosting Classifier**

# Model Performance Summary:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** | **AUC-ROC** |
| **Logistic Regression** | 0.89 | 0.83 | 0.85 | 0.84 | 0.90 |
| **Decision Tree** | 0.81 | 0.75 | 0.79 | 0.77 | 0.82 |
| **Random Forest** | 0.91 | 0.85 | 0.88 | 0.86 | 0.92 |
| **Gradient Boosting** | 0.92 | 0.87 | 0.89 | 0.88 | 0.94 |

**Graphs:**



* + **Top Performer**: The **Gradient Boosting Classifier** showed the best overall performance across all metrics, indicating it is well-suited for this dataset.
  + **Random Forest**: Provided competitive results and demonstrated robustness in feature importance evaluation.
  + **Logistic Regression**: Performed well and can serve as a simple, interpretable baseline model.
  + **Decision Tree**: Underperformed compared to the other models, particularly in precision and AUC-ROC.

# Strategic Recommendations:

1. **Focus on Usage Hours and Age**: Equipment with higher usage hours and greater age has a significantly higher likelihood of failure, suggesting these factors should be closely monitored.
2. **Monitor Vibration and Temperature**: Regular monitoring and controlling of vibration and temperature levels can help prevent failures.
3. **Optimize Preventive Maintenance**: Equipment with higher maintenance history also shows a relationship with failures, suggesting preventive maintenance strategies should be re-evaluated.
4. **Model Selection**: Based on the performance, **Gradient Boosting** is recommended for predicting equipment failures.

# Conclusion:

This project demonstrates the use of machine learning to predict equipment failures, providing valuable insights for optimizing maintenance strategies and minimizing plant downtime. Gradient Boosting was identified as the most effective model, and feature importance analysis highlighted key factors that influence equipment failure.

# Future Work:

1. **Hyperparameter Tuning**: Conduct a more extensive search using Grid Search or Random Search to further improve model performance.
2. **Exploring Other Algorithms**: Investigate advanced models such as Neural Networks or ensemble techniques like LightGBM to enhance predictions.
3. **Enhanced Feature Engineering**: Explore additional features like real-time sensor data (e.g., noise levels, humidity) for more accurate predictions.
4. **Long-Term Analysis**: Consider analyzing historical trends over time to predict future failures better.
5. **Interactive Dashboard**: Develop a real-time dashboard for plant operators to monitor equipment health and predict failures on the fly.

# Code Repository:

The complete implementation is available in the accompanying Jupyter Notebook or Python script, accessible via the project's GitHub repository.