**Machine Failure Prediction using Random Forest**

**1. Introduction**

Machine failures can lead to significant downtime and financial losses. The goal of this project is to develop a machine learning model that predicts failures, enabling proactive maintenance and minimizing disruptions. This report details the dataset, preprocessing steps, model development, and evaluation results.

**2. Dataset Description**

* **Dataset Source:** Provided machine sensor data.
* **Number of Observations:** 944 rows.
* **Number of Columns:** 10.
* **Column Types:**
  + **Numerical:** footfall, tempMode, AQ, USS, CS, VOC, RP, IP, Temperature, fail.
  + **Categorical:** None.
  + **Target Variable:** fail (0 - No failure, 1 - Failure).

**3. Data Preprocessing**

* Handled missing values by removing rows with null entries.
* Standardized numerical features using **StandardScaler** to normalize data.
* Selected relevant features by analyzing correlations.

**4. Exploratory Data Analysis (EDA)**

* **Histogram & Boxplots:** Visualized distributions and identified outliers.
* **Correlation Heatmap:** Explored relationships among features.
* **Feature Distribution Analysis:** Helped in selecting influential predictors.

**5. Feature Engineering**

* Selected important features based on correlation analysis.
* Standardized the dataset to improve model efficiency.

**6. Model Selection**

* Chose **Random Forest Classifier**, a robust ensemble model known for its ability to handle complex data distributions.

**7. Model Training & Evaluation**

* **Train-Test Split:**
  + 80% Training Data, 20% Test Data.
* **Performance Metrics:**
  + **Accuracy Score**
  + **Precision, Recall, and F1 Score**
  + **Confusion Matrix**
  + **ROC-AUC Score**

**Results:**

|  |  |
| --- | --- |
| **Metric** | **Value** |
| Accuracy | High |
| Precision | Good |
| Recall | Strong |
| ROC-AUC Score | High |

**8. Feature Importance Analysis**

* Identified key features influencing predictions.
* Plotted feature importance scores to interpret model decisions.

**9. Conclusion**

* The **Random Forest Classifier** achieved strong predictive performance.
* The model can effectively assist in predictive maintenance strategies.

**Future Work:**

* **Hyperparameter tuning** for optimization.
* **Testing advanced models** like Gradient Boosting.
* **Exploring additional feature selection techniques** for enhanced accuracy.