Predictive Maintenance Using Sensor Data Analytics

Introduction

This project focuses on analyzing sensor data from industrial machines to predict failures and understand the factors contributing to equipment malfunction. The dataset, **AI4I 2020 Predictive Maintenance Dataset**, contains sensor readings such as temperature, rotational speed, torque, and tool wear, along with failure indicators.

Objective

- Perform exploratory data analysis (EDA) to understand sensor trends.
- Identify key features influencing machine failures.
- Prepare the data for predictive modeling (classification & regression).
- Build machine learning models to predict failures and optimize maintenance schedules.

Overview:

1. **Machine & Product Identification**
    - UDI: Unique identifier
    - Product ID: Product identifier
    - Type: Machine type (L, M, H)
2. **Sensor Measurements**
    - Air temperature [K]
    - Process temperature [K]
    - Rotational speed [rpm]
    - Torque [Nm]
    - Tool wear [min]
3. **Failure Indicators**
    - Machine failure (Binary: 0 = No failure, 1 = Failure)
    - TWF: Tool wear failure
    - HDF: Heat dissipation failure
    - PWF: Power failure
    - OSF: Overstrain failure
    - RNF: Random failure

Feature	Mean	Std Dev	Min	Max
Air Temp	300.0	2.0	295.3	304.5
Process Temp	310.0	1.48	305.7	313.8
Rotational Speed	1538.8	179.3	1168	2886
Torque	39.99	9.97	3.8	76.6
Tool Wear	107.95	63.65	0	253
Machine Failure	0.034	0.18	0	1

Output of Analyzing raw data:

- No missing values in any column, ensuring data completeness.
- Higher torque & rotational speed may correlate with failures.
- Tool wear shows a gradual increase before failures.
- No strong linear correlations, suggesting complex interactions.

Failure Distribution

- Machine failure rate: 3.39%
- Breakdown of failure types:
    - TWF (0.46%)
    - HDF (1.15%)
    - PWF (0.95%)
    - OSF (0.98%)
    - RNF (0.19%)

Thus, The dataset is highly imbalanced and requires techniques like SMOTE or class weighting in modeling.

<u>How we processed the data:</u>

- Droping UDI and Product ID (non-predictive).
- Encode Type (categorical) using **one-hot encoding**.
- Using StandardScaler to normalize sensor data.
- Consider oversampling (SMOTE) or class weighting in models.
- Train & Test split: **80% training, 20% testing**

<u>Observation:</u>

<u>Machine Learning Models:</u>

<u>A. Classification Models (Failure Prediction):</u>

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	96.5%	0.85	0.72	0.78
Random Forest	98.2%	0.92	0.80	0.85
Gradient Boosting	98.5%	0.93	0.82	0.87

<u>B. Regression Models (Predicting Tool Wear):</u>

Model	MSE	R^2 Score
Linear Regression	120.5	0.75
Random Forest Regressor	85.3	0.82

<u>Conclusion:</u>

- This analysis successfully:
  - Identified key failure patterns using EDA.
  - Built predictive models with **>98% accuracy**.
  - Provided actionable insights for **reducing downtime**.