**Industrial Equipment Anomaly Detection & Predictive Maintenance System - Project Documentation**

**1. Introduction**

Industrial equipment downtime results in significant operational costs and safety risks. This project implements an anomaly detection and predictive maintenance pipeline that leverages real-time sensor data to detect machine failures as they occur and forecast future maintenance needs. By classifying failure types, the system enables targeted interventions to reduce unplanned downtime and extend equipment life.

**2. Problem Statements**

1. Can sensor data be used to reliably detect anomalies and machine failures in industrial equipment?
2. How accurately can the system classify the type of failure (e.g., heat dissipation, power issues, tool wear, overstrain)?
3. Can predictive models forecast maintenance requirements ahead of time to minimize unplanned downtime and maintenance costs?

**3. Data Sources**

The primary dataset predictive\_maintenance.csv contains 10,000 records with the following fields:

* **UID**: Unique identifier for each record (1–10,000)
* **Product ID**: Quality code (L = Low, M = Medium, H = High)
* **Air Temperature [K]**: Sensor readings generated around 300 K (σ ≈ 2 K)
* **Process Temperature [K]**: Air temperature + ~10 K noise (σ ≈ 1 K)
* **Rotational Speed [rpm]**: Derived from a 2860 W power input with noise
* **Torque [Nm]**: Centered at 40 Nm with ±10 Nm variation
* **Tool Wear [min]**: Cumulative measure influenced by product quality
* **Machine Failure**: Binary label (0 = No Failure, 1 = Failure)
* **Failure Type**: Categorical label (0 = Heat Dissipation, 1 = Power, 2 = Tool Wear, 4 = Overstrain)

Data integrity checks confirmed no missing values or duplicates.

**4. Methodology**

1. **Data Preprocessing**:
   * Outlier removal for noise-prone sensor readings
   * Label encoding of categorical fields (Product ID, Failure Type)
   * Feature scaling using StandardScaler
2. **Exploratory Data Analysis (EDA)**:
   * Visualization of distributions (box, scatter, heatmap)
   * Correlation analysis to identify key predictors
3. **Feature Engineering**:
   * Combining or transforming sensor features
   * Dimensionality reduction (if applicable)
4. **Model Development**:
   * **Failure Detection**: Binary classification using Random Forest (and comparison with Logistic Regression, SVM, K‑NN, Naive Bayes)
   * **Failure Type Classification**: Multi-class classification for records flagged as failures
   * Hyperparameter tuning with GridSearchCV
5. **Evaluation**:
   * Metrics: Accuracy, Precision, Recall, F1‑Score
   * Confusion matrices and classification reports
6. **Deployment**:
   * Flask application for real-time monitoring dashboard
   * Model serialization with joblib

**5. Results**

* **Failure Detection Accuracy**: *XX.XX%*
* **Failure Type Classification Accuracy**: *YY.YY%*
* **Confusion Matrix** and **Classification Reports** generated for both detection and classification tasks.

*Note: Replace* ***XX.XX%*** *and* ***YY.YY%*** *with the actual performance metrics obtained from the final model.*

**6. Insights**

1. **Temperature Variance**: Large fluctuations in air and process temperatures correlate strongly with Overstrain failures.
2. **Tool Wear Trends**: Gradual increases in tool wear serve as early indicators, forecasting failures up to several operational cycles ahead.
3. **Rotational Speed & Torque**: Combined analysis of speed and torque improves detection sensitivity, especially for Power‑related failures.

**7. Conclusion**

This project demonstrates a robust workflow for detecting anomalies and forecasting maintenance needs in industrial machinery. By integrating EDA, machine learning, and a real‑time dashboard, it offers a scalable solution to reduce downtime, optimize maintenance schedules, and lower operational risks. Future enhancements could include incorporating additional sensor modalities and implementing advanced deep learning models for further performance gains.

*End of Documentation*