**Predictive Maintenance for Industrial Machines**

**Problem Statement:**

Industries need to optimize their maintenance schedules in the era of digitalization to ensure minimal downtime and enhanced production efficiency. Predictive maintenance leverages data-driven insights to predict machine failures before they occur, allowing businesses to proactively service equipment and reduce unplanned breakdowns.

Using the given dataset, our objective is to build a machine learning model to predict when a machine is likely to fail based on various operational parameters. This predictive approach will help industries transition from reactive to proactive maintenance strategies, ultimately improving productivity and reducing costs.

**Business Objective:**

* Develop a predictive model to identify potential machine failures before they happen.
* Reduce unplanned downtime and improve operational efficiency.
* Minimize maintenance costs by scheduling timely interventions.
* Enhance overall equipment effectiveness (OEE) and production reliability.

**Features in the dataset:**

1. **UID:** Unique identifier for each observation.
2. **Product ID:** Comprising quality variants **L (Low), M (Medium), H (High)** along with a variant-specific serial number.
3. **Air Temperature (K):** Randomly generated values normalized around **300K** with a standard deviation of **2K**.
4. **Process Temperature (K):** Derived from air temperature with an additional **10K** and a standard deviation of **1K**.
5. **Rotational Speed (rpm):** Calculated from a power of **2860W** with normally distributed noise.
6. **Torque (Nm):** Normally distributed around **40 Nm** with a standard deviation of **10 Nm**.
7. **Tool Wear (min):** Represents the accumulated wear of the tool during operation. Quality variants affect tool wear rates.
8. **Machine Failure (Target):** A binary label indicating whether the machine failed (1) or not (0).

**Failure Modes (Causes of Machine Failure):**

1. **Tool Wear Failure (TWF):** Occurs when the tool reaches a wear limit between **200–240 minutes**.
2. **Heat Dissipation Failure (HDF):** Happens if the difference between air and process temperature is **< 8.6K** and rotational speed is **< 1380 rpm**.
3. **Power Failure (PWF):** Occurs when the power (torque × rotational speed) is **< 3500W** or **> 9000W**.
4. **Overstrain Failure (OSF):** Happens when the product of tool wear and torque exceeds predefined thresholds for different quality variants.
5. **Random Failures (RNF):** Each process has a **0.1%** chance of failing due to unpredictable reasons.

**Approach:**

1. **Exploratory Data Analysis (EDA):**
   * Examine the distribution of machine operating parameters.
   * Identify correlations between features and machine failures.
   * Analyze failure occurrences based on product quality (L, M, H).
2. **Feature Engineering:**
   * Create interaction variables between operational parameters.
   * Derive new features that capture failure conditions more explicitly.
   * Encode categorical variables (Product ID).
3. **Model Building & Training:**
   * Implement various classification models (Logistic Regression, Random Forest, XGBoost, etc.).
   * Compare model performance using **accuracy, precision, recall, and F1-score**.
   * Handle class imbalance if machine failures are rare in the dataset.
4. **Model Evaluation & Optimization:**
   * Use cross-validation and hyperparameter tuning for the best-performing model.
   * Assess feature importance to understand key drivers of failure.
5. **Deployment:**
   * Deploy the trained model using **Flask or Streamlit** for real-time failure prediction.
   * Develop an interactive dashboard for machine monitoring.

**Expected Outcomes:**

* A reliable predictive model that accurately forecasts machine failures.
* Early warning system for maintenance teams to take proactive actions.
* Reduction in unplanned downtime and associated financial losses.
* Data-driven insights into machine performance and failure causes.