# Formal Problem Proposal

## Research Question

The primary research question we aim to answer is: "How can we predict upcoming equipment failures using historical telemetry, error, maintenance, and failure data, in order to optimize maintenance schedules and reduce operational downtime and costs?"

## Expected Data Sources and Structure

The project will utilize several datasets to build the predictive model:

**1. Telemetry Data** (`PdM\_telemetry.csv`):

- Structure: Timestamped telemetry data including metrics such as voltage, rotation, pressure, and vibration collected from various machines.

- Frequency: Data collected at hourly intervals.

**2. Machine Information** (`PdM\_machines.csv`):

- Structure: Details about each machine including machine ID and model type.

- Frequency: Static information, one record per machine.

**3. Error Logs** (`PdM\_errors.csv`):

- Structure: Records of errors encountered by the machines, which may or may not precede a failure.

- Frequency: As they occur.

**4. Maintenance Records** (`PdM\_maint.csv`):

- Structure: Maintenance history including dates and types of maintenance activities performed.

- Frequency: As they occur.

**5. Failure History** (`PdM\_failures.csv`):

- Structure: Records of machine failures, classified by type of failure.

- Frequency: As failures occur.

## Expected Results

The expected outcome is a robust predictive model capable of forecasting imminent failures with high accuracy, thus enabling proactive maintenance interventions. Specific results include:

- A list of machines at high risk of failure within a defined future window (e.g., the next 24 hours).

- Identification of critical factors leading to equipment failures.

- A maintenance schedule optimized based on the predictive outcomes of the model.

## Expected Techniques

The following techniques and methodologies will be applied:

**1. Data Preprocessing:**

- Handling missing data, normalization, and data transformation.

- Time-based feature aggregation from telemetry data.

**2. Feature Engineering:**

- Creating new features from existing data to better capture the dynamics leading to failures.

- Label engineering to define the prediction window for failures.

**3. Predictive Modeling:**

- Machine Learning Models: Logistic Regression, Random Forest, Gradient Boosting Machines, and Neural Networks.

- Evaluation Metrics: Accuracy, Precision, Recall, F1 Score, and AUC-ROC.

**4. Model Validation and Testing:**

- Using techniques such as k-fold cross-validation and time-series specific train-test splits.

**5. Deployment:**

- Implementing the model in a real-world scenario to validate its effectiveness.

This proposal outlines a clear path from data acquisition through to actionable insights, leveraging advanced analytics to drive operational efficiency in equipment management.