**Milestone 1: Data Preparation & Feature Engineering**

**Project Name:** PrognosAI: AI-Driven Predictive Maintenance System Using Time-Series Sensor Data

**Dataset:** NASA Turbofan Jet Engine Data Set

**Prepared by:** Durga Veera Prasad V

**Objective:**

Prepare the NASA CMAPSS dataset for training a predictive maintenance model by cleaning data, calculating RUL, normalizing features, and generating sequences for model input.

---

**Modules Used and Their Purpose:**

**pandas:** Load and manipulate tabular datasets (`train`, `test`, `RUL` files).

**numpy:** Numerical operations, array handling, and RUL calculations.

**sklearn.preprocessing.MinMaxScaler**: Normalize sensor and operational features between 0 and 1.

os: Handle file paths and check for dataset files.

(**TensorFlow/Keras** is not used in Milestone 1; **it will be used in Milestone 2** for modeling.)

---

**Steps Implemented:**

1. Data Loading

- Loaded training, test, and RUL files for FD001–FD004 datasets.

- Assigned column names and set RUL indices for easy lookup.

2. RUL Calculation

- Training data: `RUL = max\_cycle - current\_cycle`, capped at 125 for stability.

- Test data: Adjusted RUL using `RUL\_FD00X.txt` plus remaining cycles.

3. Data Normalization

- Normalized sensor and operational settings using `MinMaxScaler`.

- Applied same scaler to training and test datasets to maintain consistency.

4. Feature Engineering & Sequence Generation

- Created rolling window sequences (`sequence\_length = 50`) for each engine.

- Generated RUL targets for sequences.

- Prepared `X\_train, y\_train` and `X\_test\_dict, y\_test\_dict` arrays for modeling.

---

**Deliverables:**

- Preprocessed and normalized datasets.

- RUL targets for all engines.

- Rolling window sequences ready for LSTM input.

**Evaluation:**

- Checked dataset shapes, missing values, and sequence generation.

- Verified normalization consistency across datasets.