■ PrognosAI - Milestone 1 Report

■■■ Contributor

Nithin G J – Project Developer

■ Objective

Prepare the NASA CMAPSS dataset for model training by: Loading and preprocessing data, performing feature engineering, generating rolling window sequences, computing Remaining Useful Life (RUL) targets, and ensuring data integrity.

■ Dataset Overview

Subset	Training Sequences	Test Sequences	Features per Sequence
FD001	17,731	10,196	24
FD002	46,219	26,505	24
FD003	21,820	13,696	24
FD004	54,028	34,081	24

■■ Preprocessing Steps

1. Loaded CSV files for train and test sets
2. Checked and cleaned data (no missing values found)
3. Normalized sensor readings
4. Computed RUL targets for each engine cycle

■ Feature Engineering

• Created rolling window sequences with window size = 30 cycles
• Each input sample shape: (30, 24)
• RUL target corresponds to the last cycle in each window

■ Data Integrity Check

• Verified shapes of training and testing arrays
• No missing values detected
• Sensor values in reasonable range (~ -0.04 to ~100)

Example:
X_train: (17731, 30, 24)
y_train: (17731,)
X_test: (10196, 30, 24)
y_test: (10196,)

■ Milestone 1 Completion

✔ Cleaned & preprocessed CMAPSS data
✔ Generated rolling window sequences
✔ Computed RUL targets
✔ Verified data integrity & absence of missing values
✔ Documented feature engineering steps