Milestone 1: Data Preparation & Feature Engineering

Project Name: PrognosAI: AI-Driven Predictive Maintenance System Using Time-Series Sensor Data
Dataset: train_FD001 – NASA Turbofan Jet Engine Data Set
Prepared by: Durga Veera Prasad V
Date: 30th September 2025

1. Objective:
- Load, preprocess, and prepare the CMAPSS dataset for model training.
- Create rolling window sequences.
- Compute RUL targets.
- Save processed data for model training.

2. Dataset Description:
- Source: train_FD001.txt (converted to CSV: train_FD001.csv)
- Columns: unit_nr, time_in_cycles, op_settings 1-3, sensors s1-s21, RUL
- Selected 10 sensors + 3 operational settings for modeling.

3. Methodology:
Step 1: Load Data - Read CSV using pandas.
Step 2: Compute RUL - RUL = max(cycle) - current cycle per engine.
Step 3: Feature Selection & Normalization - Select features, fill missing values, standardize.
Step 4: Generate Rolling Window Sequences - Window length 30, sequence labeled with last cycle RUL.
Step 5: Data Verification - Check missing values, shapes, sample RULs.
Step 6: Save Processed Data - Save X and y as .npz for model training.

4. Results:
- Number of sequences: 17,731
- Sequence length: 30 cycles
- Number of features per sequence: 13
- Missing values in features used: 0
- Sample RUL values: [162, 161, 160, 159, 158, 157, 156, 155, 154, 153]

5. Conclusion:
- Data preparation pipeline successfully completed.
- X (rolling windows) and y (RUL labels) ready for model training.
- Data integrity verified; pipeline is reproducible and robust.