NASA Turbofan Engine Degradation Simulation (CMAPSS)

MILESTONE 1 REPORT

Author: Chivukula Sai Rithivik

Data:30-09-2025

1.Objective

The objective of this project is to predict the **Remaining Useful Life (RUL)** of turbofan engines using the NASA CMAPSS dataset. Accurately predicting RUL can help in **preventive maintenance** and **failure avoidance** in real-world engine operations.

2.Dataset Description

The CMAPSS dataset contains run-to-failure data for multiple engines under different operating conditions and fault modes. It includes:

- **Engine ID (unit_number)**
- **Time step (time_in_cycles)**
- **Operational settings** (3 columns)
- **Sensor measurements** (typically 21 sensors)
- **RUL** (provided for training data)

There are four sub-datasets: **FD001, FD002, FD003, FD004**, each representing different operating conditions and fault complexities. Some sensors remain constant and provide no information for RUL prediction, while others show meaningful variations as engines degrade.

3.Methodology

- Feature **Selection:** Constant or low-variance sensors were removed to retain only informative features.
- **Data Scaling:** Sensor readings and operational settings were normalized using **Min-Max scaling** to bring all features to the same range.
- **Time-Series Sequences:** Engine data was converted into **rolling sequences** covering multiple time steps to capture temporal degradation patterns.

4. Results

- Low-variance sensors (e.g., sensor_14 in FD001) were excluded.
- **Min-Max scaling** applied to all features to standardize the data.
- **No null values** were detected in the dataset.
- **Rolling sequences** were successfully generated to preserve temporal dependencies for RUL prediction.
- Dataset is now ready for training temporal models such as LSTM or GRU.

5. Conclusion

The preprocessing of the CMAPSS dataset transformed raw engine sensor data into clean, normalized, and temporally structured sequences suitable for RUL prediction. By removing non-informative sensors, scaling features, and generating rolling sequences, the dataset is now optimized for predictive modelling, providing a strong foundation for accurate maintenance and failure prediction.