Turbofan Engine Degradation Prediction using NASA C-MAPSS Dataset

Project Overview

This project uses machine learning and deep learning models to predict the **Remaining Useful Life (RUL)** of aircraft engines using the NASA C-MAPSS FD001 dataset. Predicting RUL is crucial for preventative maintenance, reducing unexpected failures, and improving operational efficiency in aviation and other critical systems.

- **Objective**: Predict the RUL for each engine unit based on sensor readings and operational settings.
- **Tools**: Python, Pandas, scikit-learn, XGBoost, PyTorch, TensorFlow (Keras), Matplotlib, Seaborn.

Dataset Summary

Source: NASA Prognostics Center of Excellence (C-MAPSS FD001)

- **Train Engines**: 100
- **Test Engines**: 100
- **Total Features**: 3 Operational Settings + 21 Sensor Readings
- **Target**: Remaining Useful Life (RUL)

Each row corresponds to a single time step (or cycle) for an engine.

Data Preprocessing

- **RUL Calculation**: Computed from the max cycle of each engine in training data.
- **Feature Selection**: Removed flat/noisy sensors using variance threshold.
  - Retained: sensor_2, sensor_3, sensor_4, sensor_7, sensor_8, sensor_9, sensor_11–sensor_15, sensor_17, sensor_20, sensor_21.
- **Normalization**: StandardScaler applied to all features.
- **Sequence Generation**: Created sliding time windows of 30 cycles for LSTM inputs.

Modeling Approaches

We compared 4 regression models:

- **Random Forest** – Baseline ensemble model using decision trees.
- **XGBoost (Tuned)** – Gradient boosting with hyperparameter tuning (RandomizedSearchCV).

- **LSTM (PyTorch)** – Recurrent model trained on temporal sequences.
- **LSTM (Keras Tuned)** – Deep LSTM network tuned with KerasTuner.

Evaluation Metrics

- **MAE (Mean Absolute Error)**: Measures average absolute error.
- **RMSE (Root Mean Squared Error)**: Penalizes large errors more than MAE.

Final Results

Model	MAE	RMSE
LSTM (Keras Tuned)	28.91	39.53
LSTM (PyTorch)	30.68	42.05
XGBoost (Tuned)	34.28	45.64
Random Forest	34.51	45.95

Visualizations

- Scatter plots: True vs. Predicted RUL
- Feature importance plots (RF, XGB)
- LSTM loss curves
- Final model comparison (MAE & RMSE bar plot)

Key Takeaways

- LSTM models performed best, benefiting from sequential modeling.
- Feature selection significantly reduced noise and improved accuracy.
- Hyperparameter tuning notably improved XGBoost and Keras performance.
- The final pipeline is robust and can be deployed in real-time monitoring tools.

Future Work

- Use remaining datasets (FD002–FD004)
- Try CNN-LSTM hybrids or transformers
- Deploy the best model in a real-time interface (Streamlit/Flask)

Project Structure

turbofan-rul-prediction/ data/ raw/ processed/ notebooks/ 3_Modeling.ipynb 4_Model_Report.ipynb reports/ model_comparison.md model_comparison.pdf

Created by: [Muhamed Ghoz]
Last Updated: 2025-07-31