

Project Summary

This project focuses on developing a machine learning-based Remaining Useful Life (RUL) prediction system for turbofan engines using NASA CMAPSS dataset. The main objective is to enable predictive maintenance by estimating how many operational cycles remain before an engine is likely to fail. This foresight allows maintenance teams to make informed, timely decisions—balancing safety, cost, and operational efficiency.

Data Overview

- **Train Set:** Complete run-to-failure data for 100 engines
- **Test Set:** Partial life data up to a point before failure
- **RUL File:** Remaining Useful Life for each engine in the test set

Each record includes:

- Engine ID
- Cycle number
- 3 operational settings
- 21 sensor readings (columns 6–26)

Feature Engineering

- Extracted maximum cycle per engine to define the end-of-life point (EOL)
- Created the Remaining Useful Life (RUL) target as:
 $RUL = EOL - cycle_time$
- Selected sensors with known degradation trends
- Dropped low-variance or non-informative sensor columns after inspection
- Performed feature scaling using StandardScaler
- Ensured feature consistency between train and test sets before model inference

Model Selection and Justification

Multiple models were evaluated:

Model	RMSE (Initial Evaluation)
Support Vector Regressor	33.313
Random Forest Regressor	31.741
LightGBM	31.321

Selected Model: LightGBM

Reasons for selection:

- Best performance scores on experimentation
- High performance on tabular time-series data
- Built-in feature importance
- Fast training and evaluation time

Performance Summary

- **MAE:** 18.01
- **RMSE:** 24.31
- Performed best on engines with longer run histories
- Slight under-prediction for early failure cases

Maintenance Planning Use Case

The RUL prediction output can be directly integrated into a maintenance system. The RUL prediction model enables a data-driven maintenance strategy, transitioning from reactive or schedule-based approaches to predictive maintenance.

System Integration:

1. Daily or real-time ingestion of sensor data
2. Automatic risk classification based on RUL predictions

Predicted RUL	Recommended Maintenance Action
≤ 15 cycles	Immediate Repair
16–47 cycles	Schedule Inspection
> 47 cycles	Normal Operation Monitoring

3. Maintenance decision support dashboard
4. Scheduled intervention planning

Benefits:

- Failures are detected before they occur, minimizing unscheduled outages
- Eliminates unnecessary early replacements while avoiding catastrophic failures
- Supports prioritization of engine maintenance through data-driven risk ranking

Conclusion and Recommendations

The LightGBM-based RUL prediction system performs reliably, especially for engines with longer histories, enabling proactive, condition-based maintenance. By forecasting failures in advance, it can help to reduce downtime, extend asset life, and optimize resource planning.

The model does overestimate RUL in short-lived engines, which can be mitigated by adjusting thresholds (e.g., monitor at ≤ 47 cycles, repair at ≤ 15), allowing teams to balance early warnings with operational risk.

Limitations and Future Work

- Class Imbalance: Adjust training to better capture short-lived engine behavior
- Early Signals: Add features that reflect early degradation trends
- Temporal Learning: Explore LSTM or Transformer models for sequence data
- Sensor Noise: Use smoothing or filtering to address sensor instability
- Dynamic Thresholds: Create maintenance flags to context and cost-risk tradeoffs
- Optimization: Integrate cost-aware planning for greater operational impact