Remaining Useful Life (RUL) Prediction Using Machine Learning

Summary

This project focuses on predicting the Remaining Useful Life (RUL) of aircraft engines using sensor data. Using machine learning techniques such as Linear Regression and Random Forests, I developed a model that can estimate how many operational cycles an engine has before failure. This has critical applications in predictive maintenance and operational safety.

Objective

As an aspiring ML/AI engineer, my objective was to:

- Apply supervised learning to a real-world problem
- Preprocess time-series sensor data
- Train, evaluate, and interpret machine learning models
- Deploy the model for future prediction on unseen data

Dataset Description

Source: NASA's CMAPSS dataset (via Kaggle)

Files Used: DS01-005.h5

Features:

- Engine sensor readings (temperatures, pressures, flow rates)
- Operational settings
- RUL (target variable)

We focused on one engine variation file to build and validate the model. Each row represents an engine cycle.

Methodology

- **Data Loading** using h5py
- **Exploratory Analysis**: Understanding sensor patterns
- **Preprocessing**:
  - Feature selection
  - Train/test split by engine units
  - Feature scaling using StandardScaler
- **Modeling**:
  - Linear Regression (baseline)
  - Random Forest Regressor (final model)
- **Evaluation**:
  - Metrics: MAE, RMSE, R^2
  - Plots: Actual vs Predicted, Residuals
- **Deployment**:
  - Saved model and scaler using joblib

- Created test pipeline in new notebook

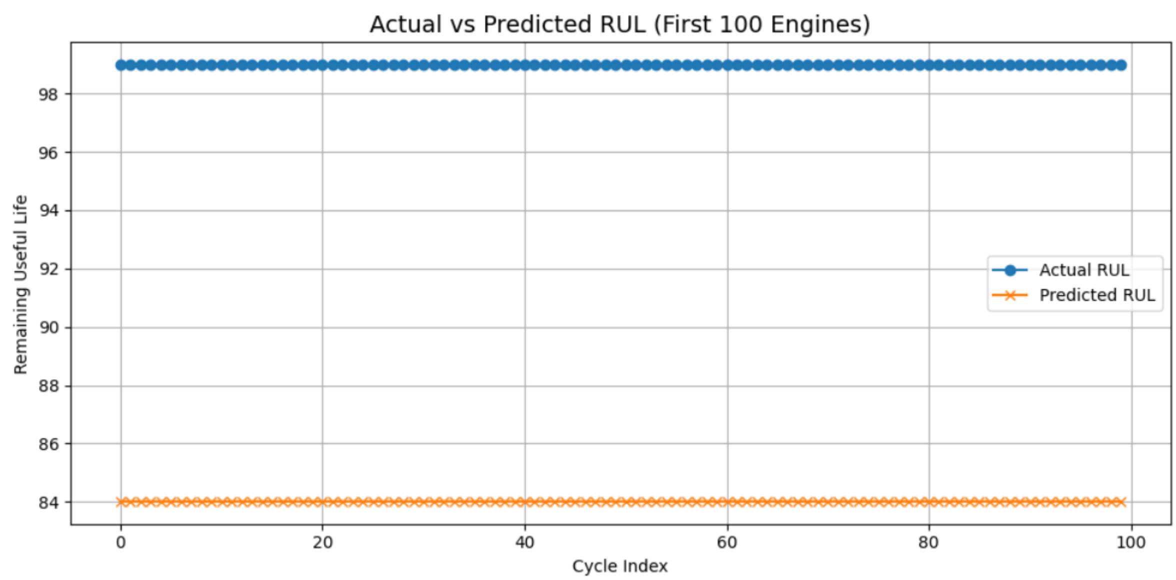
Model Results

Model	MAE	RMSE	R²
Linear Regression	10.22	12.01	0.76
Random Forest (Top 5 Features)	9.43	11.82	0.77

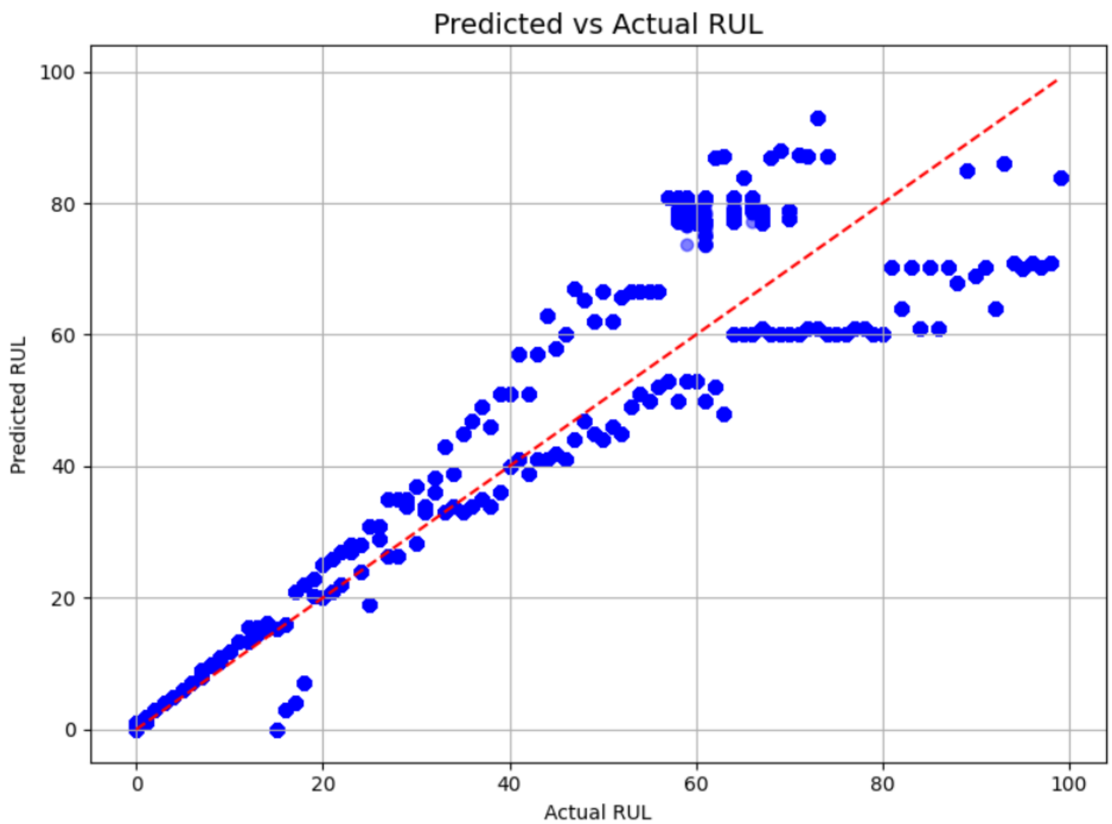
Despite using fewer features, Random Forest outperformed Linear Regression and showed good generalization on unseen data.

Visualizations

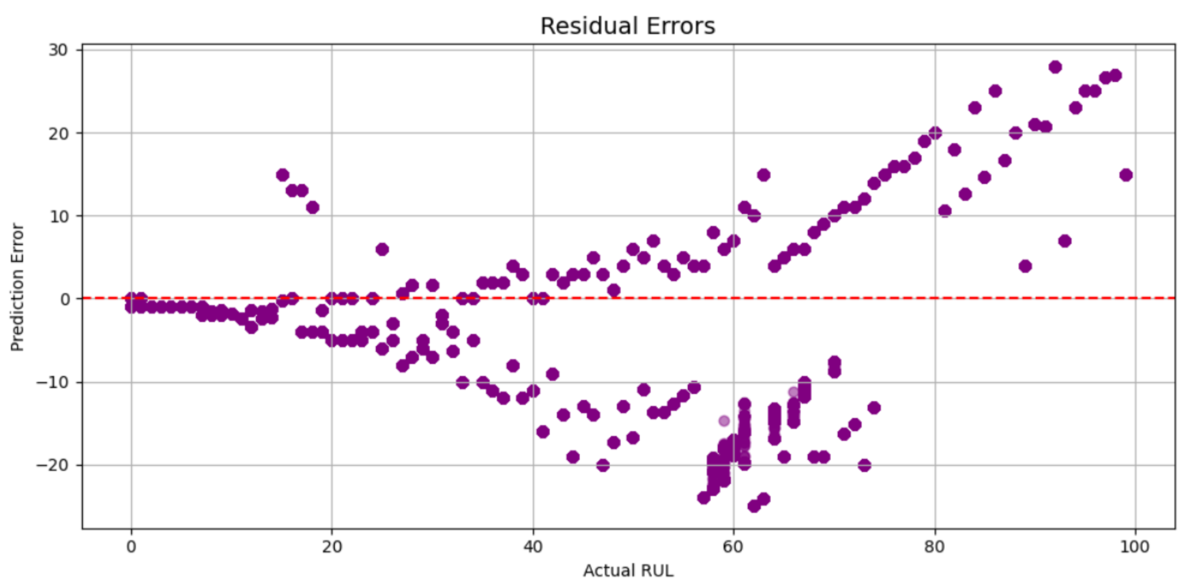
- **Actual vs Predicted RUL** (First 100 samples)



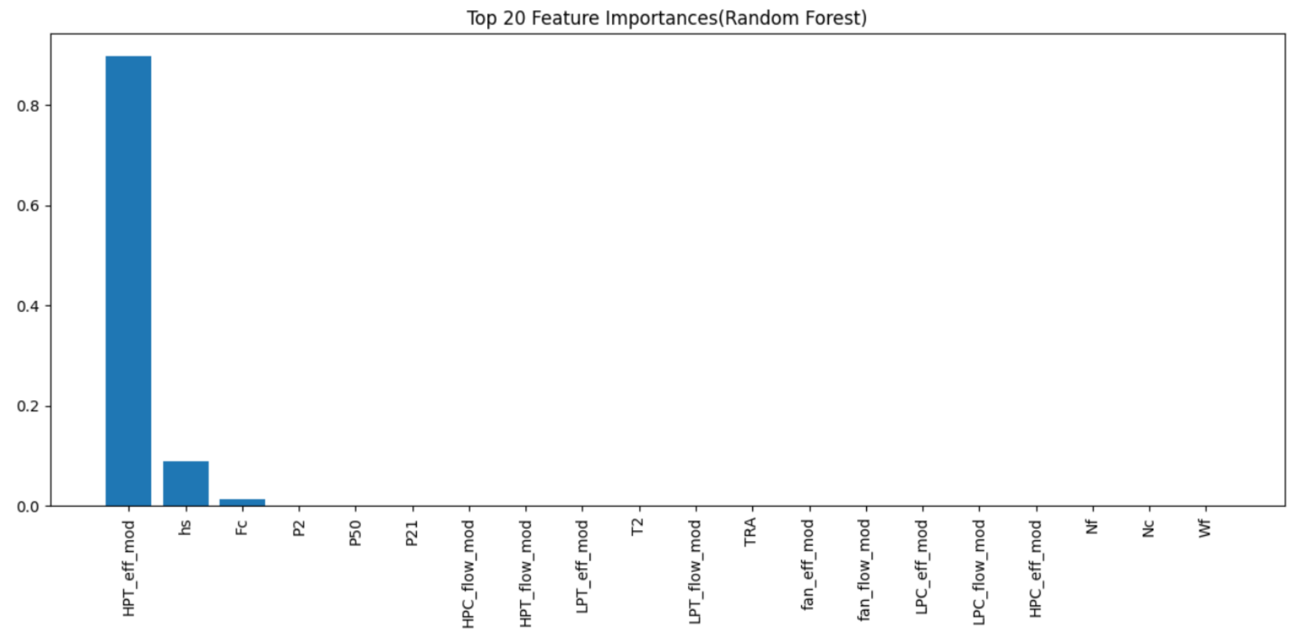
- **Predicted vs Actual Scatter Plot**



- **Residual Error Distribution**



- **Feature Importance Ranking**



Visualizations helped diagnose model bias and understand feature influence.

Key Learnings

- Handling time-series data for supervised learning
- Importance of data scaling and avoiding data leakage
- Interpreting regression model results
- Saving & deploying ML models using joblib
- Communicating results through visuals

Future Work / Research

This project can be improved by:

- Incorporating all .h5 engine variations
- Trying advanced models like XGBoost or LSTM (for time-series modeling)
- Hyperparameter tuning for Random Forest
- Adding rolling features or feature engineering for better accuracy
- Building a dashboard for RUL visualization