PrognosAI: Predictive Maintenance System

Milestone 2 Report: Model Development, Optimization, and Evaluation

By Thilak

Executive Summary

This report details the development of a predictive maintenance model for turbofan engines using the NASA CMAPSS FD001 dataset. The primary objective was to predict the Remaining Useful Life (RUL) of an engine based on its operational settings and sensor data. A Long Short-Term Memory (LSTM) neural network was designed, trained, and evaluated for this time-series forecasting task. The model successfully learned patterns from the historical data, and the training was regularized using Dropout and Early Stopping to prevent overfitting. The final trained model was saved for potential future deployment in predictive maintenance applications.

1. Project Overview

Objective: To build a deep learning model capable of accurately predicting the Remaining Useful Life (RUL) of a turbofan engine, which is a critical component of predictive maintenance strategies designed to minimize operational downtime and prevent catastrophic failures. **Dataset:** The NASA CMAPSS (Commercial Modular Aero-Propulsion System Simulation) FD001 dataset was used. This dataset contains simulated time-series data for a fleet of turbofan engines, capturing various sensor readings and operational settings until failure.

2. Data Preparation and Preprocessing

The initial phase focused on preparing the data for the time-series model. **Data Loading:** The project utilized three key files — train_FD001.csv, test_FD001.csv, and RUL_FD001.csv — loaded into pandas DataFrames with columns for engine ID, cycle number, operational settings, and sensor measurements. **Feature Engineering:** Remaining Useful Life (RUL) was calculated as the number of cycles remaining before engine failure. **Data Scaling:** Features were normalized using MinMaxScaler to a range of [0,1] for improved neural network performance.

3. Model Development

Sequence Generation: Data was converted into overlapping sequences using a sliding window of 30 time steps, producing an input shape of (samples, 30, 24). **Data Splitting:** 80% of the data was used for training and 20% for validation. **LSTM Model Architecture:** 1. LSTM Layer (128 units, return_sequences=True) 2. Dropout (0.2) 3. LSTM Layer (64 units) 4. Dropout (0.2) 5. Dense (32 units, ReLU activation) 6. Output Layer (1 unit for RUL prediction) Compiled with Adam optimizer and MSE loss function.

4. Training and Evaluation

The model was trained for up to 100 epochs with batch size 64, using EarlyStopping to prevent overfitting (patience=10). Training stopped after 20 epochs once validation loss stabilized. Training and validation MSE and MAE curves showed convergence, indicating successful learning and generalization.

5. Conclusion and Model Persistence

The LSTM model effectively predicted Remaining Useful Life (RUL) of turbofan engines using NASA CMAPSS data. Regularization prevented overfitting, and the final trained model was saved as 'lstm_rul_model.h5' for future deployment in predictive maintenance systems.