Milestone 2 — Model Development & Training

Project: PrognosAI — AI-Driven Predictive Maintenance System

Dataset: NASA CMAPSS Multivariate Time-Series Sensor Data

■ Objective

The purpose of Milestone 2 is to design and train a deep learning model capable of estimating the Remaining Useful Life (RUL) of industrial machinery based on time-series sensor data. This step focuses on transforming raw, preprocessed data from Milestone 1 into a predictive model that can understand complex temporal relationships between sensor readings and degradation behavior. By leveraging sequential learning techniques like LSTM (Long Short-Term Memory), the system aims to accurately recognize early failure patterns, enabling predictive maintenance decisions that minimize unexpected breakdowns, reduce costs, and enhance equipment reliability.

■■ Process Summary

1. **Data Utilization** – The preprocessed CMAPSS dataset from Milestone 1 was used. The data was split into training, validation, and test sets to ensure unbiased model evaluation. 2. **Data Normalization** – Sensor readings were scaled to a standard numerical range, which improves the model's convergence and helps it interpret varying sensor magnitudes consistently. 3. **Sequence Preparation** – Time-windowed sequences were created to help the model analyze how sensor behavior evolves over time. Each sequence represented a portion of an engine’s operational history. 4. **Model Development** – A Long Short-Term Memory (LSTM) neural network was developed. LSTM networks are ideal for time-series data because they can retain information about previous cycles, allowing the model to learn degradation trends effectively. 5. **Training Strategy** – The model was trained across multiple epochs using an adaptive learning rate and early stopping mechanisms. This ensured stable learning, reduced overfitting, and retained the best-performing version through model checkpointing.

■ Results & Observations

- **Model Convergence:** The training and validation loss steadily decreased throughout training, showing that the model successfully captured useful temporal features from the sensor data. - **Prediction Behavior:** The predicted Remaining Useful Life values followed the actual RUL trends across multiple test engines. Although minor deviations were observed near end-of-life predictions, the overall accuracy and trend alignment were strong. - **Performance Analysis:** The model achieved a reasonable Root Mean Square Error (RMSE) on the validation dataset. This indicates the model can generalize well to unseen machinery, providing a strong baseline for predictive maintenance applications. - **Training Stability:** Early stopping prevented overfitting by halting training once validation performance stabilized, ensuring that the final model remained both robust and efficient.

■ Insights

- The LSTM architecture effectively captured long-term dependencies in multivariate time-series data, which is crucial for understanding machine degradation patterns. - Data normalization and sequence structuring played a vital role in stabilizing the learning process and improving model reliability. - Early stopping and dropout layers helped prevent overfitting, maintaining generalization even with complex sensor input data. - The results validate that the implemented model and data pipeline are functioning correctly, forming a solid foundation for further fine-tuning and evaluation in the next stages.

■ Deliverables Completed

- Successfully implemented and trained a deep learning model (LSTM) for RUL prediction. - Verified convergence through loss and accuracy visualization. - Validated model predictions against actual RUL data. - Saved trained model weights and preprocessing configurations for future evaluation. - Documented all experimental results and findings from the training phase.

■ References

- Dataset: NASA CMAPSS (Commercial Modular Aero-Propulsion System Simulation) - Frameworks: TensorFlow, Keras, Scikit-learn, Pandas, NumPy - Visualization: Matplotlib, Seaborn