# Milestone 3 Report: Model Evaluation & Performance Assessment

**Project:** PrognosAI: AI-Driven Predictive Maintenance System **Date:** October 30, 2025
**Submitted by:** Praveen Kumar S

## 1. Objective

The primary objective of this milestone was to rigorously evaluate the performance and predictive accuracy of the trained LSTM model. This was accomplished by testing the model against a separate, unseen dataset (`test_FD001.txt`) and comparing its predictions to the ground truth RUL values.

---

## 2. Methodology

The evaluation process was conducted in three distinct phases:

- **Test Data Preparation**: The `test_FD001.txt` dataset was loaded and subjected to the exact same preprocessing pipeline used for the training data. This included removing constant-value sensor columns and transforming the data into sequences of 50 cycles. A padding mechanism was implemented to handle engines with shorter operational histories, ensuring a uniform input shape for the model.
- **Prediction**: The prepared test sequences (`X_test`) were fed into the trained LSTM model to generate a Remaining Useful Life (RUL) prediction for each of the 100 engines in the test set.
- **Performance Measurement**: The model's accuracy was quantified using the **Root Mean Squared Error (RMSE)** metric. Additionally, a visual analysis was performed by plotting the model's predicted RUL against the actual RUL values.

---

## 3. Results and Analysis

The evaluation yielded both a quantitative score and qualitative insights into the model's behavior.

**Quantitative Results: RMSE Score**

The model's performance on the test set was calculated, resulting in the following score:

- **Root Mean Squared Error (RMSE): 29.98 cycles**

This score indicates that, on average, the model's RUL predictions deviate from the actual values by approximately 30 cycles.

**Qualitative Results: Prediction Analysis**

The scatter plot comparing the actual RUL to the predicted RUL provided critical insights into the model's predictive patterns.

- **Key Observation**: The model demonstrates a strong tendency to predict RUL values within a narrow range, centered around **80 cycles**.
- **Model Strengths**: For engines whose actual RUL is in the mid-range (approx. 70-100 cycles), the model's predictions are reasonably accurate. This shows it has successfully learned to identify the general characteristics of a mid-life engine.
- **Model Limitations**: The model currently struggles with predicting extreme values. It tends to be overly pessimistic for healthy engines (predicting ~80 when the actual RUL is high) and overly optimistic for engines near failure (predicting ~80 when the actual RUL is low).

---

## 4. Conclusion

**Milestone 3 is complete.** The model has been successfully evaluated, and its performance has been quantified and analyzed.

The results show that the model has learned a valid baseline for engine degradation and can make reasonable predictions, achieving an RMSE of approximately 30 cycles. However, the analysis also reveals a clear area for future improvement: enhancing the model's ability to predict RUL values across the entire spectrum of an engine's lifecycle, not just the middle range.

The current model provides a solid foundation for the next phase of the project: **Milestone 4: Risk Thresholding & Alert System**.