Project 10

Predictive Maintenance for Industrial Equipment

Priyal Jain[1], Ayush Singh[2], and Aditya Thomas[3]

[1]2021MT60949
[2]2021MT10899
[3]2021MT60944

November 24, 2024

Abstract

Predictive maintenance leverages data-driven techniques to foresee equipment failures, enabling proactive interventions that enhance reliability, minimize downtime, and reduce operational costs. This project aims to develop a machine learning model for predicting equipment failures using sensor data from industrial equipment. Employing the NASA C-MAPSS dataset of turbofan engines, the study explores advanced methods such as feature engineering, noise removal and utilizing deep learning techniques like LSTM networks for sequence prediction. This research underscores the effectiveness of predictive maintenance in optimizing industrial operations while addressing challenges such as feature selection and model interpretability. The proposed model's outcomes will be evaluated using performance metrics such as root mean square error, mean absolute error and s-score.

1 Introduction

Predictive maintenance, an essential component of Prognostics and Health Management (PHM), enables proactive equipment management by predicting failures before they occur. Traditional maintenance strategies, including reactive and preventive maintenance, are often inefficient and costly, as they either address failures after they occur or involve unnecessary repairs. In contrast, predictive maintenance harnesses data-driven techniques to anticipate failures based on real-time data, ensuring timely interventions and reducing unplanned downtime.

The application of machine learning to predictive maintenance has demonstrated significant potential, particularly in leveraging multivariate sensor data to predict equipment failures. In this project, the focus is on developing and evaluating a machine learning model capable of predicting failures in turbofan engines using the NASA C-MAPSS dataset. This dataset offers a rich source of multivariate time-series data, capturing sensor readings for parameters such as temperature, pressure, and vibration.

This study involves a comprehensive analysis, including a literature review of predictive maintenance methodologies, exploratory data analysis, feature engineering, and the implementation of machine learning models. Challenges such as the interpretability of complex models and noisy data are also addressed. The ultimate goal is to identify the most effective predictive strategies and assess their potential impact on industrial applications.

2 Problem Statement

Suppose that N equipment of the same type are monitored within a pre-specified time range T. For every equipment i ($i \in N$), we have sensor readings from R sensors at M_i time points, $(T_{i,1}, \ldots, T_{i,M_i})$ such that $T_{i,j} \in T$ for $j = 1, \ldots, M_i$. The number of observations and the observation times can be different across equipment, i.e., $M_i \neq M_{i'}$ for $i \neq i'$. The Remaining Useful Life (RUL) for the ith equipment since the last observation in X_i is denoted as Y_i. The goal of the RUL estimation problem is to learn a mathematical mapping from X_i to Y_i.

When modeling from the Functional Data Analysis (FDA) point of view, for every equipment i ($i \in N$), the M_i sensor readings from a given sensor r are regarded as discretized realizations of the underlying continuous curve $X_{i,r}(t)$ with $t \in T$ and $r = 1, \ldots, R$. The RUL estimation problem is essentially about learning the following mapping from continuous sensor observations to the RUL label Y_i,

$$Y_i = F\big(X_{i,1}(t), \ldots, X_{i,R}(t)\big). \qquad (1)$$

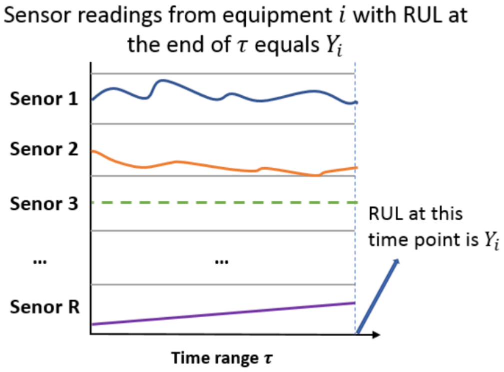


Figure 1: Illustration of the problem setting for RUL estimation.

The RUL estimation problem setting from the functional data perspective is demonstrated in Fig. 10.

3 Data Desctiption

Data Set	Train Trajectories	Test Trajectories	Conditions	Fault Modes
FD001	100	100	One (Sea Level)	One (HPC Degradation)
FD002	260	259	Six	One (HPC Degradation)
FD003	100	100	One (Sea Level)	Two (HPC and Fan Degradation)
FD004	248	249	Six	Two (HPC and Fan Degradation)

Figure 2: Dataset

- **Column 1:** Corresponds to engine number (This column is indexed 0 above because of Python's numbering convention).
- **Column 2:** Corresponds to cycle number. If engine 1 fails after 192 cycles, the entries of the second column for engine 1 will go from 1 to 192. Similarly for other engines.
- **Columns 3, 4, 5:** 3 operational settings.
- **Columns 6–26:** 21 sensor measurements.

Trend between sensor readings and RUL was observed for all the sensors. Trajectories FD001 and FD003 both have redundant sensors which can be dropped as they are constant or rarely change. FD004 and FD002 both have a redundant sensor 16.

Figure 3: Sensor 15

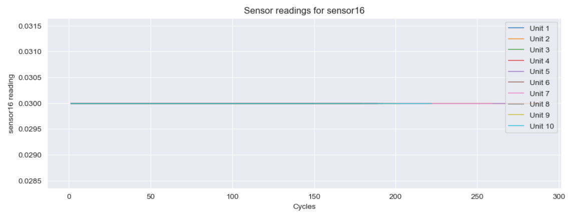


Figure 4: Sensor 16

Figure 5: Sensor 20

4 Data Pre-Processing

4.1 Removing Columns with low variance

Columns with a very low standard deviation were ignored as it shows the values are roughly constant. Constant values do not have much bearing on machine learning methods, especially here since we are dealing with time series data.

4.2 Removing highly co-related columns

By utilizing a heatmap, we observed that certain sensors exhibited high correlation with one another. To reduce redundancy and improve data quality, we removed one of the highly correlated sensors.

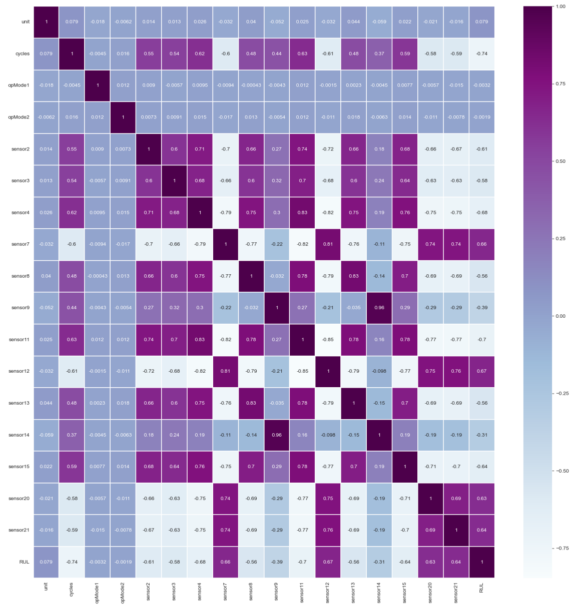


Figure 6: Heat Map for FD001

4.3 Filtering Noise using Savitzky-Golay

Since the dataset is extremely noisy, the Savitzky-Golay filter was employed to enhance data clarity. This is a digital filtering method that smooths noisy data while preserving important features like trends and peaks. A quadratic polynomial was fitted to a moving window of data points and the window length was tailored to the number of cycles, respecting the data's periodic nature.

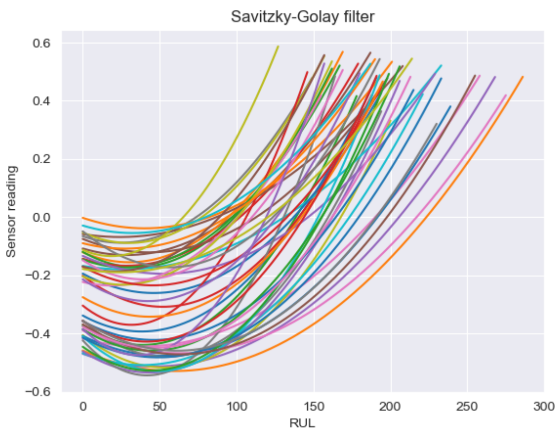


Figure 7: Plot for sensor15

5 Evaluation Metrics

RMSE

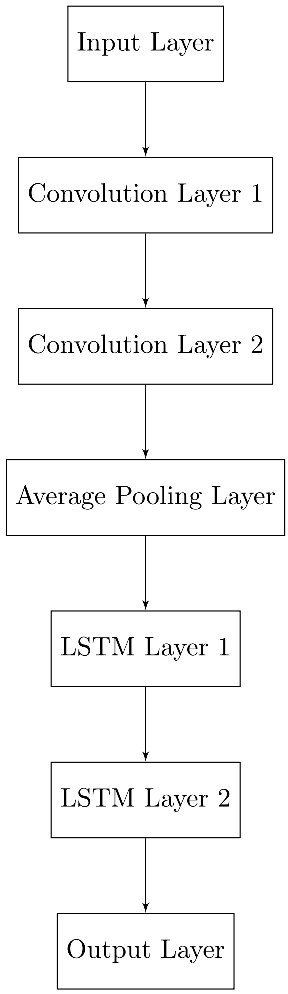
$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} h_i^2}$$

S-SCORE

$$S = \begin{cases} \sum_{i=1}^{n} \left(e^{\frac{h_i}{3}} - 1\right), & \text{if } h_i < 0, \\ \sum_{i=1}^{n} \left(e^{\frac{h_i}{10}} - 1\right), & \text{if } h_i \geq 0. \end{cases}$$

6 Advanced Model

The model presented here employs a combination of convolutional and Long Short-Term Memory (LSTM) layers to predict the Remaining Useful Life (RUL) of industrial equipment. The convolutional layers are used to capture local temporal patterns and extract meaningful features from the input sequence. Subsequently, the LSTM layers learn long-term dependencies across the time steps. The input sequence consists of sensor readings over several cycles, which are processed through convolution, pooling, and LSTM layers to generate the RUL prediction.



7 Results

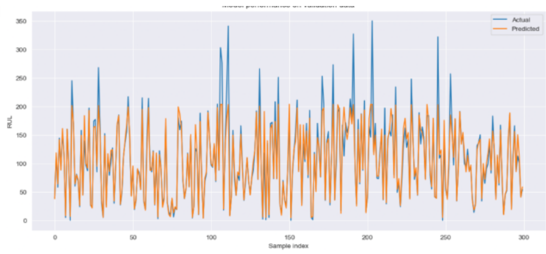


Figure 8: Model performance on Validation Data

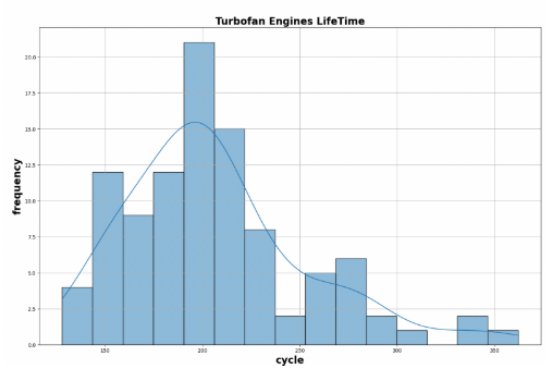


Figure 9: TurboFan Engines Lifetime

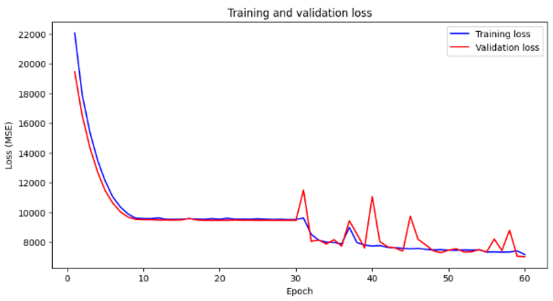


Figure 10: Training and Validation loss

Table 1: RMSE for Different Models

Model	FD001	FD002	FD003	FD004
SVM	22.33	46.54	28.76	29.12
Random Forest	23.21	45.70	31.15	34.39
LSTM	19.78	29.44	21.17	33.01

Table 2: s-Score for Different Models

Model	FD001	FD002	FD003	FD004
SVM	1901.34	614000	2344.02	593000
Random Forest	1871.13	823000	1992.98	734000
LSTM	440.12	4900	911.87	6700

8 Conclusion

Why LSTM is Better Than SVM and RF for RUL Prediction :

- Captures Temporal Dependencies: LSTM models handle sequential data effectively by capturing time-dependent patterns in engine degradation.
- Handles Long-Term Dependencies: LSTMs can remember information over long sequences, crucial for modeling gradual degradation.
- Processes Variable Sequence Lengths: They accept input sequences of varying lengths without needing fixed-size inputs.
- Reduces Feature Engineering: LSTMs learn features directly from raw time-series data, minimizing manual preprocessing.
- Adapts to Dynamic Patterns: Capable of adjusting to changes in data patterns over time, unlike static models like SVM and RF.

9 References

1. S. Zheng, K. Ristovski, A. Farahat, and C. Gupta, "Long Short-Term Memory Network for Remaining Useful Life Estimation," 2017 IEEE International Conference on Prognostics and Health Management (ICPHM), 2017, pp. 88-95 https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7998311.
2. Q. Wang, S. Zheng, A. Farahat, S. Serita, and C. Gupta, "Remaining Useful Life Estimation Using Functional Data Analysis," arXiv:1904.06442 [cs.LG], 2019. https://arxiv.org/pdf/1904.06442
3. O. Asif, S. A. Haider, S. R. Naqvi, J. F. W. Zaki, K.-S. Kwak, and S. M. R. Islam, "A Deep Learning Model for Remaining Useful Life Prediction of Aircraft Turbofan Engine on C-MAPSS Dataset," IEEE Access, vol. XX, no. XX, 2022. https://ntrs.nasa.gov/api/citations/20090029214/downloads/20090029214.pdf.
4. A. Saxena and K. Goebel, "Turbofan Engine Degradation Simulation Data Set," NASA Ames Research Center, Moffett Field, CA, Tech. Rep. NASA TM-2009-215988, 2008. https://data.nasa.gov/Aerospace/CMAPSS-Jet-Engine-Simulated-Data/ff5v-kuh6