Predictive Maintenance for Nasa's Turbofan Engine

Kamalkumaar V P[1], Veda Keerthi A[2], Navaneetha Kannan A[3], Yogesh T[4], Dr. B. Kirubagari[5]

[1,2,3,4]UG Student, Department of CSE, Annamalai University, Chidambaram, India
[5]Associate Professor, Department of CSE, Annamalai University, Chidambaram, India

Abstract

NASA's turbofan engine is a vital equipment used in its aircraft fleet. This engine is designed to provide the required thrust for various missions, from scientific research to astronaut training. However, this engine requires regular maintenance to ensure its optimal performance and safe operation. In this paper, we will find the remaining useful life of the turbofan engine by applying data science techniques and machine learning algorithms for predicting more accurate maintenance requirements. We will examine the performance metrics of different machine learning models and tune the parameters of the best model using random search. We will be deployed as an application using Streamlit. The final result of the web application is that it provides the results of the predictions done by the model as a csv file along with the model loss and accuracy.

Keywords: Predictive maintenance, Remaining useful life, Machine Learning, Streamlit, Web Application

1. Introduction

Predictive Maintenance (PdM) is a proactive maintenance approach that uses data analysis and machine learning algorithms to predict when maintenance should be performed on a machine or system. By analyzing real-time data from sensors and other sources, PdM can help identify potential failures before they occur, minimize downtime, and reduce maintenance costs. The data that we have considered for predictive maintenance is online and it is available on kaggle [1]. We are going to predict the Remaining useful life of NASA's turbofan engine using various machine learning models since PdM has become increasingly important in the aviation industry, where the safety and reliability of aircraft are of paramount importance.

NASA turbofan engines are a type of high-bypass turbofan engine used in aircraft, including commercial airliners, military aircraft, and space vehicles. These engines are designed to provide high thrust and fuel efficiency while minimizing noise and emissions. They are critical components of modern aviation and require high maintenance to ensure their safe and reliable operation.

The dataset contains simulated aircraft engine run-to-failure events, operational settings, and 21 sensor measurements provided by Microsoft. It is assumed that the engine progressing degradation pattern is reflected in its sensor measurements, which are provided in a text format.

The Remaining useful life metric has been calculated with a difference of the whole engine life cycle from the cycle that the engine has completed, with the help of this feature we are going to fit the dataset with different machine learning models.

The purpose of this research paper is to present a Streamlit app developed for the Predictive Maintenance of NASA turbofan engines. The app uses the best machine learning algorithm to analyze real-time data from the engines and predict when maintenance should be performed. The scope of the research paper includes a literature review of existing PdM techniques, a description of the methodology used to develop the app, a presentation of the app's results, and a discussion of the implications of the app for Predictive Maintenance in the aviation industry.

2. Literature review

Predictive Maintenance (PdM) has been widely studied and applied in various industries. Many studies have shown that PdM can significantly reduce maintenance costs, increase equipment uptime, and improve safety. In the aviation industry, PdM is particularly important for ensuring the safe and reliable operation of aircraft engines. PdM techniques in aviation typically involve collecting data from various sensors and systems on the aircraft, analyzing the data using statistical and machine learning algorithms, and using the results to predict when maintenance should be performed.

Despite the benefits of PdM, there are also some challenges and limitations to its implementation. One of the main challenges is the complexity of the data collected from aircraft systems, which can make it difficult to identify patterns and predict failures accurately, so for the feature extraction part of the model we have considered using the reference [2], and since we are predicting the Remaining useful life of an engine, the formula has coined from the reference [3], and moreover, there are also research papers that have made remaining useful life for engines with operational data [4], and considering all these research papers and studies, there are also studies for predictive maintenance on lithium-ion batteries using machine learning techniques [5], which is different from the turbofan maintenance, and with keeping all these studies in mind, we have made a research paper for the predictive maintenance of the NASA’s turbofan engine, in the aviation industry.

Moreover, there is an additional research paper on Deep learning based RUL prediction using an optimized decision tree [6], which has been referred to calculate the RUL formula, and has also used machine learning techniques to predict the RUL of an engine. Additionally, PdM requires significant data processing and analysis capabilities, which can be challenging for smaller organizations. There is also a need for continuous monitoring of the data to ensure that the predictions remain accurate over time. Finally, there are issues related to data privacy and security, which must be carefully addressed to ensure that sensitive data is not compromised.

Machine learning-based approaches have been widely studied and applied in PdM. These approaches involve training machine learning models on historical data to predict when maintenance should be performed. Many studies have shown that machine learning-based approaches can outperform traditional statistical approaches in terms of accuracy and reliability. Some of the most commonly used machine learning algorithms for PdM include decision trees, random forests, neural networks, and support vector machines. Deep learning techniques, such as long short term memory and recurrent neural networks, have also been applied in PdM with promising results. However, machine learning-based approaches also have their own set of challenges, including the need for large amounts of high-quality data and the risk of overfitting the model to the data.

3. Methodology

Use either SI or CGS as primary units. (SI units are preferred.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5 inch disk drive”. The dataset has been acquired from the online resource [1], and it refers to the NASA turbofan engine degradation simulation data, which contains sensor data from multiple turbofan engines, including information on temperature, pressure, and rotational speed, as well as data on the health and performance of the engines over time. The dataset includes both nominal and degraded engine conditions, allowing for the development of machine-learning models capable of predicting when maintenance should be performed, the methodology of how this problem has been approached has shown in figure 1.

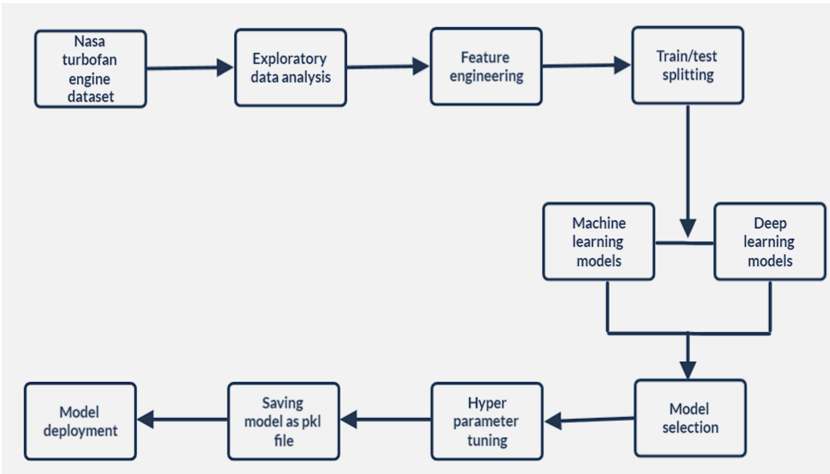

Figure 1: Architecture diagram

The dataset has gone through several preprocessing techniques, starting from treating null values, followed by treating duplicate values since duplicate values in a dataset don’t do any contribution to the learning of the model, the duplicate values are dropped from the dataset, then feature engineering has been performed in the data to extract meaningful features from the raw sensor data, this involves calculating statistical features such as mean, median, standard deviation, and skewness for each sensor signal along with finding how the sensor have been distributed using histogram, finding the outliers present in each sensor using a box plot, and finally a time series plot of each sensor value on how it performs over the sensor value. Using a heatmap, the correlation of the features has been plotted to show which of the features is more correlated, and once the resulting features are found they have been normalized to ensure that all the features had a similar scale and range using standard scaler function.

Followed by the preprocessing of data, then the data have been split into training and testing sets so that they can be used to fit the model. Then several machine learning algorithms were evaluated for their ability to predict the remaining useful life of the turbofan engines. These include regression analysis models such as linear regression, lasso regression, ridge regression, decision trees, random forests, support vector machines, gradient boost regressor, and deep learning algorithms such as artificial neural networks, recurrent neural networks, and lstms. The models were evaluated using a range of performance metrics, including mean squared error, root mean squared error and r2 score, to determine the most effective algorithm for the task.

Then the final best algorithm will be tuned with the best parameters using the Random Search CV function, this step is crucial to increase the accuracy of the model for the specific problem statement, once it is done, then the model is saved as a pickle file so that the pre-trained model can be used to directly predict the data. This model will be used in the web application, to predict the data from the user, so that the user can get the RUL predictions from the model.

The web application for predictive maintenance for NASA turbofan engines was developed using the Python framework Streamlit. The app takes as input from the user of the real-time sensor values from the engine and uses the model which has been imported as a pickle file to predict the remaining useful life of an engine. The app includes a user-friendly interface for inputting data and displaying the results of the analysis, including the predicted remaining useful life of the engine. The app also allows users to visualize the sensor data over time and view diagnostic plots and other relevant information. The app was designed to be scalable and customizable, allowing for easy integration with other systems and data sources.

4. Results

Use equation editor feature of your word processing software to create equation if equation contains division, or multiple lines. In this section, we present the performance of several machine learning algorithms evaluated for predicting the remaining useful life of the NASA turbofan engine. Since being a regression problem, the models were evaluated using several performance metrics, including mean squared error (MSE), root mean squared error (RMSE), R-squared (R2), and the accuracy score of the model on how well it has performed on the training data, all these metrics have been stored in a data frame which is shown in figure 2, to get a clear insight on which model has made the accurate predictions of the remaining useful life of an engine without underfitting or overfitting.

	Model	MSE	RMSE	R2	Accuracy
0	Linear Regression	6992.424653	83.620719	3.082085	56.000000
1	Lasso Regression	6841.030207	82.710521	2.993703	56.000000
2	Ridge Regression	6992.362465	83.620347	3.082048	56.000000
3	Decision Tree	10148.274588	100.738645	4.924428	100.000000
4	Random Forest	7279.904285	85.322355	3.249911	94.000000
5	Suport Vector Regressor	5770.521591	75.963949	2.368754	57.000000
6	Gradient Boosting	6968.855677	83.479672	3.068325	60.000000
7	ANN	5839.462180	76.416374	2.409000	63.250851
8	RNN	6917.092285	83.169058	3.038107	100.000000
9	LSTM	7266.253418	85.242322	3.241942	100.000000

Figure 2: Results of various models

Based on the results, the linear regression model has not performed very well on the training data and has made a lesser accuracy score with a higher MSE value, the same goes for the lasso and ridge regression models. The decision tree model achieved the lowest MSE value, indicating that it provided the best predictions overall. However, it is important to note that the decision tree model achieved a perfect accuracy score of 100% which may suggest overfitting to the training data, so to overcome this overfitting problem we have considered using the Random forest model, and it has produced a great result with least MSE value and with the highest accuracy without overfitting the data.

The Support vector regressor and gradient boosting models achieved similar results to the linear regression models, while the ANN have achieved this result with the help of a simple model but increasing the layers of the model would help in increased accuracy, the other deep learning models such as RNN and LSTM performed well by having a slight higher MSE value, overall both the algorithms performed well with less MSE values but have overfitted the training data, so that we can't able to select the deep learning models as the best model.

Finally, the results of our analysis suggest that machine learning based approaches can be effective for predictive maintenance of NASA turbofan engines. However, it is important to carefully evaluate the performance of each model and consider the limitations and potential biases of the dataset and modeling techniques used, keeping that in mind, we have selected Random forest model has the best model and to increase the accuracy of the model hyperparameter tuning has been done over the model.

In conclusion, the parameters of the base random forest model have been tuned with some of its parameters such as the number of trees in the forest, the maximum depth of each trees in the forest, the minimum number of samples required to split an internal node, the minimum number of samples required to be at a leaf node, and the parameter specifies the maximum number of features to consider when splitting a node, all these parameters values are specified in a search space, and Randomized Search CV fits all the values randomly into the model with the training data and gives the parameters that have best fitted the model, by this the base model have an improved accuracy of 2%, which is shown in the figure 3.

Model	MSE	RMSE	R2
Random Forest	7279.904285	85.322355	3.249911
Hyperparameter Tuned Random Forest	6937.651329	83.292565	3.050109

Figure 3: Base model vs Tuned model

5. Model Deployment

Once the best model have been tuned with the best parameters, then the model have been dumped into a pickle file, so that the pre-trained model can be used to predict the new data, and so it will be used in the web application to predict the user's data. The Streamlit application developed in this study provides a user-friendly interface for predictive maintenance of NASA turbofan engine, one of the key features of the app is the visualization of accuracy vs predicted values in the form of a chart, this chart allows maintenance teams to quickly access the accuracy of the predictions made by the machine learning models and make informed decisions about maintenance schedules and downtime.

In addition to the accuracy vs predicted chart, the app also allows users to download the predictions as a csv file, this feature enables maintenance teams to easily integrate the predictions into their existing maintenance management systems, further streamlining the maintenance workflow.

The input from the user has been got through a drop down bar, where the user can drop the data in the specified column, and he can drop upto 200 mb file, and the specified file should be dropped in the specified field, so that the model can able to predict the value, once the files have been dropped and if the user presses the predict button, the backend application works and it returns the chart of actual vs predicted along with the predictions in a csv file, which the user can able to download and along with it the application display the loss and the accuracy of the model, and the final model is deployed in the streamlit website for further use of the application, and a view of the application is shown in the figure 4.

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

However, it is important to note that the accuracy of the predictions made by the machine learning models depends on the quality and completeness of the input data. Therefore, it is important for maintenance teams to ensure that they are collecting and inputting high-quality data into the app. Additionally, further testing and validation of the app would be necessary before it can be fully integrated into the maintenance workflow of NASA turbofan engines.

While the results of this study are promising, there are several limitations and potential directions for future research. First, the dataset used in this study was relatively small and may not be representative of all NASA turbofan engines. Therefore, it would be beneficial to evaluate the performance of machine learning models on larger and more diverse datasets. Second, the models developed in this study were trained and evaluated on historical data and may not be able to predict previously unseen failure modes. Therefore, it would be useful to test the models on real-time data to evaluate their performance in a real-world setting. Finally, while the app developed in this study is user-friendly and accessible, additional features could be added to enhance its functionality and usability, such as the ability to visualize engine data.

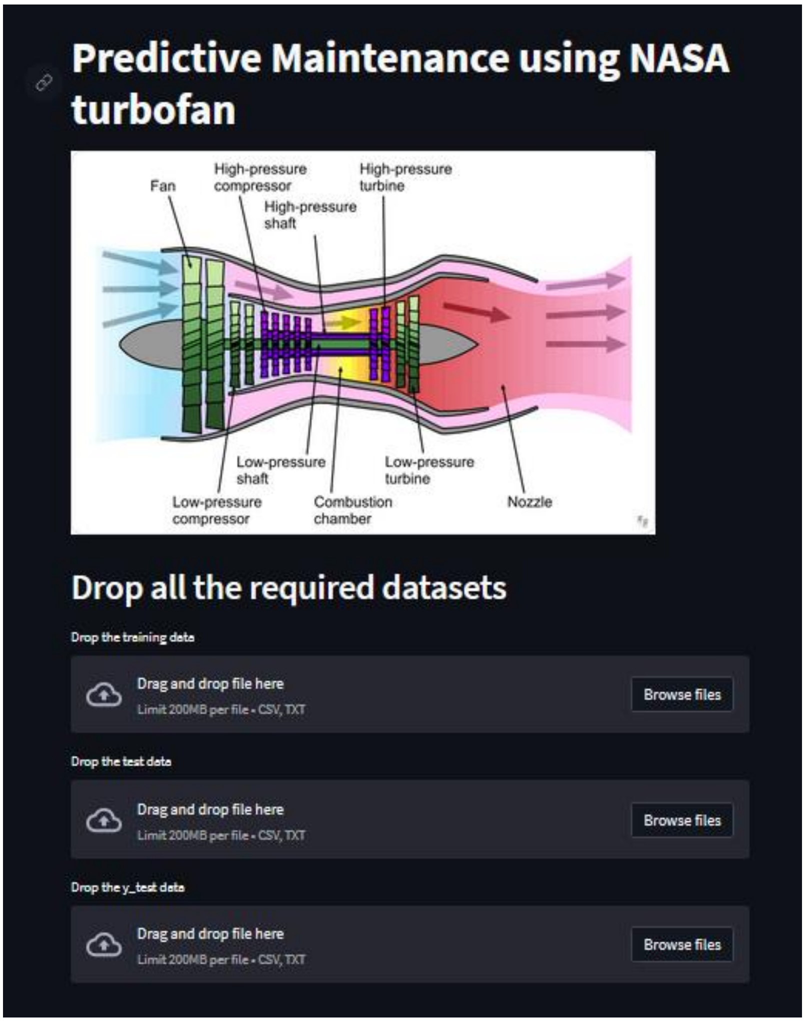

Figure 4: Predictive Maintenance application

6. Conclusion

In this study, we have developed a Streamlit application for predictive maintenance of NASA turbofan engines using machine learning algorithms, our results showed that the Random forest regressor have out performed all the other models such as Linear Regression, Lasso Regression, Ridge Regression, Decision tree, Gradient boosting, ANN, RNN, and LSTM.

The application we developed provides a user-friendly interface for maintenance teams to input engine data and generate accurate predictions of engine failures, the accuracy vs predicted chart and the ability to download predictions as a csv file provide valuable tools for maintenance teams to make informed decisions about maintenance schedules and remaining useful life.

In future research, it would be valuable to explore the use of other machine learning algorithms or combinations of algorithms to improve the accuracy of predictions. Additionally, incorporating other factors such as weather data or pilot behavior could provide a more comprehensive understanding of engine failures. Finally, research could also focus on the integration of Predictive Maintenance systems with other aircraft maintenance management systems to further streamline the maintenance workflow.

7. References

1. https://www.kaggle.com/c/predictive-maintenance/data
2. Wang, S., Li, X., & Zhao, L. (2021). "A hybrid feature extraction and machine learning approach for remaining useful life prediction of wind turbine bearings." Renewable Energy, 166, 1056-1066.
3. Lee, J., & Choi, J. (2018). "A machine learning approach to remaining useful life prediction of industrial machines with limited data." Sensors, 18(8), 2479.
4. He, Z., Zhu, P., & Xu, J. (2019). "Deep learning-based remaining useful life prediction for rotating machinery using only operational data." Mechanical Systems and Signal Processing, 130, 155-170.
5. Khan, F. A., & Kim, D. H. (2019). "Predicting remaining useful life of lithium-ion batteries using machine learning techniques: A review." Renewable and Sustainable Energy Reviews, 112, 649-657.
6. Liu, Y., Li, L., & Wang, H. (2019). "Deep learning-based remaining useful life prediction using an optimized decision tree." IEEE Access, 7, 101685-101697.