Analysing NASA Turbofan Data For Remaining Useful Life

While there are thousands of road accidents worldwide every year, these accidents are usually discussed at a local level and have short-lived effects. However, when an aeroplane crash occurs, the situation is completely different; although accidents are rare, why does their impact have a global repercussion?

Aviation is an industry that safely transports billions of travellers every year and forms the backbone of the modern world. Rare aircraft accidents lead to questioning of both technology and safety measures. However, each accident is not just an individual event, but has the potential to create a turning point for aviation safety.

The aim of this paper is to examine methods for predicting the remaining life of aircraft engines and to evaluate the role of these predictions in preventing potential accidents.

When choosing this topic, I repeatedly asked myself the question 'What am I really interested in?' The answer to these questions was the 'Aviation' industry. Then, when I compared the injury rates in my research, I discovered the critical role of 'The Importance of Safe Travel of Passengers in Aviation'. The safety of travel was also related to 'Aircraft components'. For this reason, my research question became clear as 'How can I ensure the Safe Use of Aircraft Engines?' I asked questions to my teachers and friends who have experience in Data Science. For example, I asked one of my teachers whether working on such a subject would really benefit the society? My teacher told me that I chose a data set that directly affects human life and that I should try it. With such conversations, I have gained deeper knowledge on the subject.

In particular, with the 'Asking more questions' and 'Conversations with Effective Listening' that I learnt in the field of Data Science and Leadership, the issue of how to personalise a general topic became more understandable for me. In my first submitted work, the number of questions I asked myself and the people in front of me was more limited. The answers I got from here were guiding me, but they were not enough for me to advance an idea. By asking more questions, this subject became more understandable. I can carry out more comprehensive studies with both individual learning and what I have learnt from my peers. I started to question different questions. For example, 'How will I ensure my credibility as a researcher?' 'Which methodology will I apply when marketing a product?' 'What are the requirements to be a consultant?'. I learnt how to handle a topic, how to make it more grounded and how to be persuasive and reassuring. Learning all these issues affected my choice of topic in the project.

After choosing the topic, I had to determine my goal in the project and I wanted to develop a deep learning based model to predict the remaining useful life (RUL) of aircraft engines. Using sensor data from aircraft engines, this model will try to predict the remaining life of the engine. Thus, maintenance processes will be made more efficient, costs will be reduced and flight safety will be increased. By presenting multivariate data with complex temporal patterns, this dataset is critical not only for model development, but also for real-world adaptability of the results. Analysing NASA Turbofan Data will be an important step to improve flight safety by enabling early detection of aircraft engine failures.

Social Benefits and Impact on Human Life

This project is not only intended to be a technical achievement. It also has the potential to provide a tangible benefit to society by improving flight safety. Being able to predict aircraft engine failures before they occur directly contributes to preventing potential accidents and saving lives. Such technological advances play a critical role in preventing the loss of life each year as a result of aircraft accidents.

Another important impact of the project is to go beyond accident prevention and increase the confidence of individuals and communities. Adopting such an approach in the aviation sector not only ensures the safety of millions of travelling people, but also enables the thousands of technical experts working in this sector to do their jobs more safely and efficiently.

Strategic End Goals

Cycle 1 (Change Application):

- Development of a robust LSTM model to predict the Remaining Life of aircraft engines. This model will predict the remaining life of engines by analysing sensor data.

- Test and validate the performance of the developed model on NASA Turbofan Data Set. In the test phase, the prediction accuracy of the model will be evaluated by comparing with existing methods.

- Provide tangible and actionable insights into the potential of the resulting RUL predictions to optimise maintenance plans. The results of the model will guide decision makers in maintenance planning and improve the efficiency of maintenance processes.

- **Cycle 2 (Organisational Transformation):**

- Adoption and dissemination of a data-driven decision-making culture within the organisation. Data-driven decision making is a critical factor for the long-term success and sustainability of organisations.

- It will fulfil not only a technical achievement, but also an ethical responsibility to protect human life. In a field such as aviation, where safety is paramount, predicting and preventing accidents is expected to make a significant difference on an individual and societal level.

Methodology

1. Data Preprocessing:

- Load and clean the NASA Turbofan Data Set to make it ready for analysis. Missing data, noisy data and inconsistencies in the dataset should be removed. Data cleaning is a critical step to improve the accuracy of the model (NASA, 2021).

- Since this dataset represents sensor data from real aircraft engines, it plays a critical role in the modelling process and is crucial for the reliability of the results.

- Start the analysis process by normalising the sensor data and creating appropriate time series sequences for the LSTM model. Preparing the data in the appropriate format is crucial for the success of the model.

- Evaluate the model performance by dividing the data into training, validation and test sets. Proper separation of training and test sets will increase the generalisation capacity of the model (Muller, 2017).

2. Feature Engineering:

- Analyse the importance level of different sensors in the dataset and derive relevant features that will improve the performance of the model. This allows the model to make faster and more accurate predictions.

- Apply sliding window techniques to better capture time dependencies. This technique helps to detect sequential dependencies in time series data (Bergstra et al., 2013).

3. Model Development:

- Build a model capable of time series forecasting using the LSTM algorithm. LSTM is very good at learning long-term dependencies and is well suited for this project.

- Try to maximise the accuracy of the model through hyperparameter optimisation (e.g., learning rate, number of layers, number of units per layer). Hyperparameter settings can significantly affect the success of the model (Chollet, 2017).

4. Model Evaluation:

- Measure model performance using common evaluation metrics such as Mean Square Error (MSE) and Root Mean Square Error (RMSE). These metrics will show how accurately the model makes predictions (Hastie et al., 2009).

- Demonstrate the superiority of LSTM by comparing LSTM results with baseline methods such as linear regression and random forests. These comparisons will show how effective the model is compared to other methods (Srivastava et al., 2015).

5. Deployment Plan:

- Develop a prototype dashboard to visualise the resulting RUL estimates. This dashboard will facilitate the decision-making processes of the maintenance teams.

- Propose strategies to facilitate the integration of the model into the organisation's existing maintenance workflows. Integrating the model into the workflow will increase organisational efficiency (Wang et al., 2020).

This methodology provides an organised way to understand and solve complex problems. It brings effective solutions to problems through steps such as data processing, model development and integration of results. It also provides a useful approach for leadership by improving organisational efficiency.

While the chosen methodology is powerful, it has limitations. For example, gaps and noise in real sensor data can negatively affect the performance of the model. Also, while the LSTM model is successful in learning complex time series relationships, the training time can be long and require high computational power. Finally, the accuracy and applicability of the model depends on the quality of the data used and its suitability for real-world scenarios. Therefore, I need to utilise both my technical and leadership skills efficiently.

In this process, I want to explore how I can make the right decisions as a leader. I believe that change is not only about achieving goals, but also about creating a cultural transformation. I want to create a more open and inclusive environment, and the impact I will have will be to create an environment where everyone has opportunities for development. I aim to fulfil my ethical and moral responsibilities in the right way in this process. I think that my teachers, who I can consult in such intensive projects, who help me learn new information every week, who support me to look at the world more critically and question-based, and my friends who motivate me will play a role.

References

Bergstra, J., Yamins, D., & Cox, D. (2013). Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, 115-123.

Chollet, F. (2017). *Deep learning with Python*. Manning Publications Co.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Muller, A. (2017). *Introduction to machine learning with Python: A guide for data scientists*. O'Reilly Media.

NASA. (2021). *NASA Turbofan Engine Degradation Simulation Data Set*. Retrieved from https://www.nasa.gov/

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2015). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1), 1929-1958.

Wang, Y., Xu, L., & Zhang, Y. (2020). Predictive maintenance for aviation systems: A case study using deep learning techniques. *Journal of Aerospace Engineering*, 33(1), 1-9.