Engine Condition Trend Monitoring using Machine Learning Algorithms

**Shreyas S**

Machine Learning Intern, Faststream Technologies,

Bangalore, Karnataka, India

hello@shreyas.im

**Sudhanva N**

Machine Learning Inter, Faststream Technologies,

Bangalore, Karnataka, India

[nsudhanva@gmail.com](mailto:nsudhanva@gmail.com)

**Vinay Bansal**

Chief Executive Officer, Faststream Technologies,

Bangalore, Karnataka, India

[vinay@faststreamtech.com](mailto:vinay@faststreamtech.com)

**Vinod Agrawal**

Chief Technology Officer, Faststream Technologies,

Bangalore, Karnataka, India

vinod.agrawal@faststreamtech.com

**Abstract: Aircraft engines are composed of multiple moving and stationary components constituting a highly complex system. One or more of these parts can fail due to multiple reasons such as wear of parts, environmental damage, insufficient electrical power supply, insufficient lubrication etc. It is extremely crucial for an aircraft manufacturer to know when any component in an engine might fail. There are different maintenance strategies to like preventive maintenance, where upkeep tasks are performed very frequently and therefore result in high unnecessary expenditure. Another maintenance strategy is reactive maintenance, which involves not repairing equipment until it breaks and usually causes costly stretches of unplanned downtime. The third strategy is predictive maintenance, also known as Engine Condition Monitoring. Aircraft Engine Condition Trend Monitoring (ECTM) helps assist operators in managing the health of the turbine engines. Trend Monitoring is the process of monitoring the performance for engines over a period of time. By tracking a number of parameters from the turbine engine, it is possible to predict a failure before it occurs. This paper focuses on predicting the probability of failure of an aircraft turbine engine based on five parameters – Fan Blade’s Noise, Exhaust Gas Temperature, Fuel Flow, Low Pressure Fan Speed and High Pressure Rotor Speed. Three Machine Learning algorithms namely Multiple Linear Regression, Decision Tree Regression and Random Forest Regression are used for prediction purposes. The results of each of these algorithms are plotted against Age and Probability of failure. Finally, comparison of all the three algorithms is made by calculating their R2 scores.**

***Keywords: Engine Condition Trend Monitoring, Machine Learning, Regression, Multiple Linear Regression, Decision Tree Regression, Random Forest Regression, Aircraft Engine***

1. **INTRODUCTION**

Aircraft engines are composed of multiple moving and stationary components constituting a highly complex system [1]. Displays in the cockpit indicate performance of the engine using vital information [2] such as exhaust gas temperature, fuel flow, low pressure fan speed, high pressure rotor etc. Along with these, there are a lot of other parameters such as Fan Blade’s noise, vibrations, oil supply to bearings, air temperature and pressure etc. that can serve as an early indication of failure. There are different sensors for each of these components that record values over a period of time. These recorded values help in the prognosis of failure of an aircraft engine. It is extremely crucial for an aircraft manufacturer to know when any component in an engine might fail. With proper training, the analysis can be manually done by humans, but the risk for error is great. Hence, Engine Condition Trend Monitoring (ECTM) tools are used for predicting the failure of a components before they occur.

There are a lot of different ways in which ECTM can be implemented. In this paper, we have considered five parameters – Fan Blade’s noise, Exhaust Gas Temperature, Fuel Flow, Low Pressure Fan Speed and High Pressure Rotor Speed. Using historical data of these five parameters and how they affected the probability of failure of the engine, we perform predictive analysis using three Machine Learning Regression algorithms names Multiple Linear Regression, Decision Tree Regression and Random Forest regression. Performances of each of these algorithms have been compared by computing their R2 values.

The paper is organized as follows: Section 2 contains literature survey, Section 3 deals with the methodology, Section 4 presents the implementation and pseudo code, Section 5 deals with Performance Analysis where every algorithm’s performance is measured and compared. Section 6 finally concludes the work.

1. **LITERATURE SURVEY**

There is a lot of research work going on for Engine Condition Trend Monitoring. QI Shufen et al. [3] have implemented an open layered architecture with intelligent reasoning to enhance the accuracy of fault dianosis and reduce false reports. P Janssens et al. [4] have used neural networks and expert system techniques to enhance Engine Condition Monitoring Process to ensure better consistent results. Andrew K S Jardine et al. [5] have attempted to summarize and review the recent research and developments in diagnostics and prognostics of mechanical systems implementing Condition Based Monitoring with focus on techniques, technologies and algorithms for data processing and maintenance decision-making. K Mathioudakis et al. [6] have presented methods of analyzing aerothermodynamic performance measurement data for assessing the condition of components of a gas turbine. Louis A Urban [7] has presented an approach to turbine gas path analysis and monitoring that permits the isolation of single of simultaneous multiple engine faults, with a quantitative assessment of their relative severity.

1. **METHADOLOGY**

There are a several steps that go into Engine Condition Trend Monitoring. The architectural diagram for the same is given in Figure 1.

Acquiring Field Data for Training

Acquire Field Data for Testing

Save the Model

Predict the Probability of Failure

Fit the saved Model to the Test Data

Visualize the Predictions

Compute R2 Score

Train Machine Learning Algorithms with the field data

Figure 1.Architectural Diagram

* The first step is to acquire historical data from multiple aircraft engines. We call this data as the Training Set. This data is used to train the Machine Learning Model by applying different algorithms.
* The trained model is then saved.
* Next step is to acquire some more new data for validating out trained machine learning model. This data is referred to as Test Set.
* The machine learning model is then loaded to the Test Set.
* Probability of failure of the engine is then predicted using the data from the Test Set.
* To validate the results, graph is plotted against Age of the engine and Probability of Failure with one plot representing nominal values and another plot representing the predictions from the Test Set.
* Finally R2 Score is computed for every algorithm and the results are compared.

1. **IMPLEMENTATION**

A few assumptions have been made due to the absence of available data.

* Data has been generated based on nominal value and random degradation.
* A minimum and maximum value of each component has been preset.
* An increase in the value of any one of the components beyond its maximum value affects the probability of failure of that component.
* The failure of each component constitutes to the probability of failure of the whole Engine.
* The turbine engine fails completely when the probability reaches one.
* Data for one thousand engines with random degradation have been generated as Training Set which is used to train the Machine Learning Models.
* Data for another engine in 10 different scenarios with random degradation has been generated as Test Set to validate the trained Machine Learning Model.

The assumed minimum, threshold and maximum values of the engine parameters are given in the Table 1:

|  |  |  |  |
| --- | --- | --- | --- |
| **Components** | **Minimum value** | **Threshold value after which degradation starts** | **Maximum value when probability of failure reaches 1** |
| Fan Blade Noise | 130dB | 137dB | 143dB |
| Exhaust Gas Temperature of Turbine | 1300oC | 1900 oC | 2500 oC |
| Fuel Flow of Turbine | 5000kg/hr | 5500kg/hr | 6000kg/hr |
| Low Pressure Fan Speed | 12000rpm | 14000rpm | 16000rpm |
| High Pressure Rotor Speed | 10000rpm | 12000rpm | 14000rpm |

Table 1. Assumptions of engine parameter values

Details of all the steps involved in implementing ECTM are given below.

1. **Generating Data**

There are three datasets generated totally. The first dataset is of an engine under nominal condition spread across 5 years. The second dataset generated is for training the machine learning model. This dataset is generated with random degradation. The third and final dataset is generated as a Test Set to validate the trained machine learning model. The flow diagram for generating data is given in Figure 2.

Importing Libraries

Set minimum and maximum values for each parameter

Generate values for each parameter with random degradation

Generate probability of failure for each engine based on assumptions

Stop

Figure 2. Flow Diagram of Data Generating

1. **Training Machine Learning Models**

As mentioned in the Table 1, all the parameters are set to the minimum value in the first month. As the months progress, random degradation occurs. With each component’s degradation, the probability of failure of the engine starts increasing. We have used Python’s ScikitLearn library to train the machine learning models. Flow diagram for the same is given in Figure 3.

Importing Libraries

Importing Dataset

Encoding Categorical Data

Avoid Dummy Variable Trap

Splitting Dataset

Applying Machine Learning Algorithms

Saving the Model

Figure 3. Flow for Training Machine Learning Models

We have used three Machine Learning algorithms to train our models. They are as follows:

1. **Multiple Linear Regression:**

Multiple Linear Regression is a linear approach to modelling a relationship between a scalar response (dependent variable) and explanatory variables (independent variables) [8].

The model for Multiple Linear Regression, given n observations is

i = 1, 2, … , n

: Predicted value (dependent variable)

through : p distinct independent variables

: value of y when all the independent variables ( through ) are equal to zero

1. **Decision Tree Regression:**

Decision Tree is a learning technique where information extracted is organized from a training set in a hierarchical structured composed of nodes and ramifications. It is easy to understand the results for decision trees as the output of the decision tree can be organized in the form of tree or rules [9].

The final result is a tree with decision nodes and leaf nodes. A decision node has two or more branches, each representing values for the attribute tested. Leaf node represents a decision on the numerical target. The topmost decision node in a tree which corresponds to the best predictor called root node.

1. **Random Forest Regression:**

The Random Forest Regression is a type of Machine Learning Algorithm that makes predictions by combining decisions from a sequence of base models. More formally we can write this class of models as:

*g(x) = f0(x) + f1(x) + f2(x) + … + fn(x)*

where the final model *g* is the sum of simple base models *fi*. Here, each base classifier is a simple Decision Tree.

1. **Validating Machine Learning Models**

The trained machine learning models are saved as a file with ‘.plk’ extension. It is done so using Python’s library called Pickle. This file is then loaded and fit to the Test Set to validate the results. The steps followed to validate our trained machine learning model are given below:

Flow diagram for the same is given in Figure 4.

Importing Libraries

Importing Dataset

Encoding Categorical Data

Avoid Dummy Variable Trap

Fitting Machine Learning Model to the Dataset

Computing Results

Figure 4. Flow for Validating Machine Learning Models

1. **RESULTS**

For result analysis of the Test Set, we have plotted a graph for visualization.

The results of all the three algorithms are compared using R2 Score with the generated data.

1. **Visualization**

The test set contains data of an engine under 10 different scenarios with random degradation. For better visualization, we have plotted the predictions of 5 scenarios at once with the 6th plot being engine with nominal values in a single graph. Hence, there are 2 graphs for each algorithm.

The graph is plotted against Age of the engine in months and Probability of Failure. Green line represents life of an engine under nominal conditions. Red, blue, orange, yellow and black lines represent life of an engine with random degradation under different scenarios. X-axis represents Age in months and Y-axis represents Probability of failure of the engine.

1. **Multiple Linear Regression**

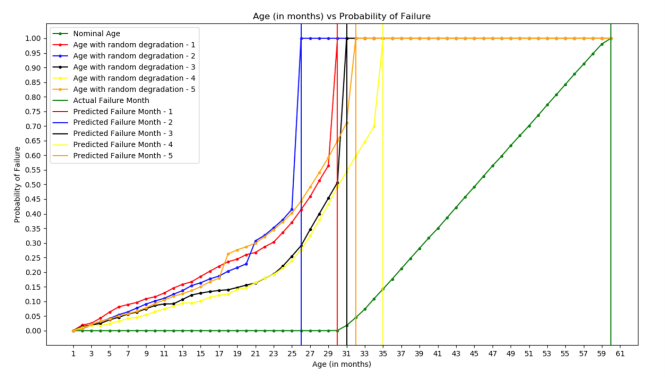


Figure 5. Visualization of predictions done using Multiple Linear Regression for scenarios 1-5

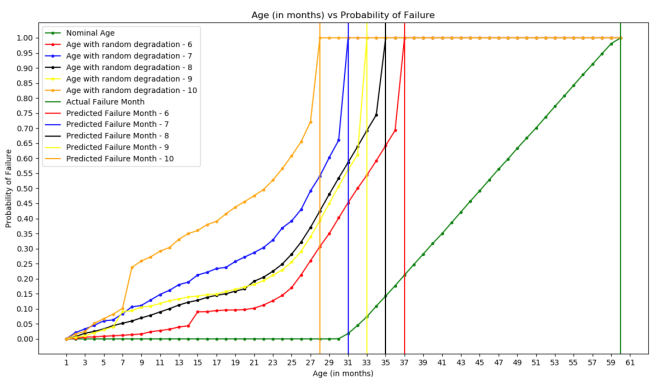


Figure 6. Visualization of predictions done using Multiple Linear Regression for scenarios 6-10

1. **Decision Tree Regression**

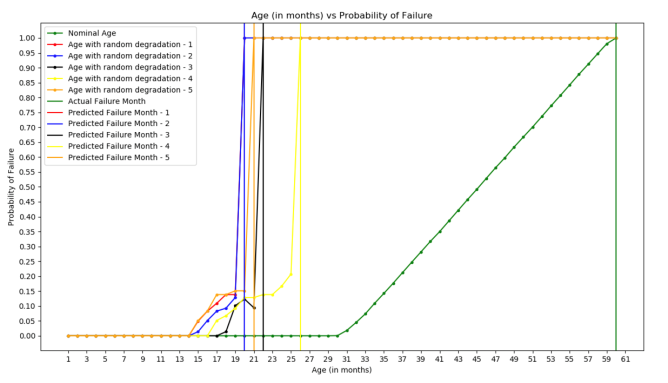


Figure 7. Visualization of predictions done using Decision Tree Regression for scenarios 1-5

|  |  |  |  |
| --- | --- | --- | --- |
| **Scenario #** | **Multiple Linear Regression** | **Decision Tree Regression** | **Random Forest Regression** |
| 1 | ﻿0.6426 | ﻿0.9916 | ﻿0.9904 |
| 2 | ﻿0.7640 | ﻿0.9933 | ﻿0.9915 |
| 3 | ﻿0.6416 | ﻿0.9920 | ﻿0.9927 |
| 4 | ﻿0.8241 | ﻿0.9654 | ﻿0.9638 |
| 5 | ﻿0.7090 | ﻿0.9843 | ﻿0.9830 |
| 6 | ﻿0.8242 | ﻿0.9977 | ﻿0.9977 |
| 7 | ﻿0.6043 | ﻿0.9975 | ﻿0.9942 |
| 8 | ﻿0.8486 | ﻿0.9987 | ﻿0.9981 |
| 9 | ﻿0.6422 | ﻿0.9609 | ﻿0.9600 |
| 10 | ﻿0.7033 | ﻿0.9811 | ﻿0.9780 |

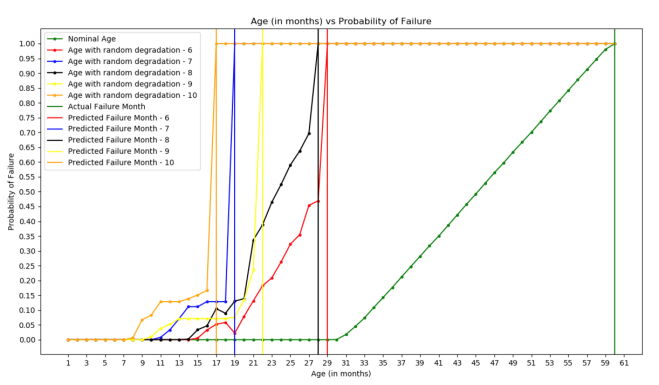


Figure 8. Visualization of predictions done using Decision Tree Regression for scenarios 6-10

1. **Random Forest Regression**

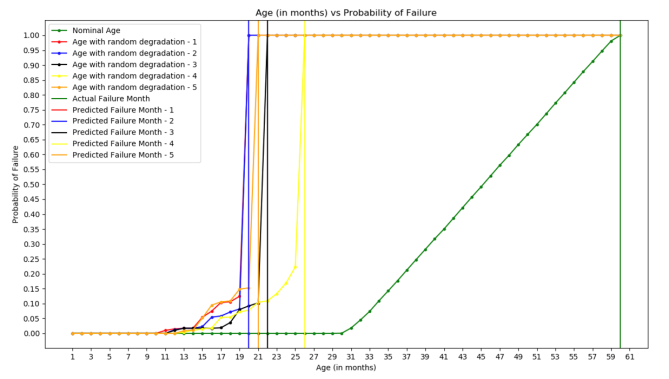


Figure 9. Visualization of predictions done using Random Forest Regression for scenarios 1-5

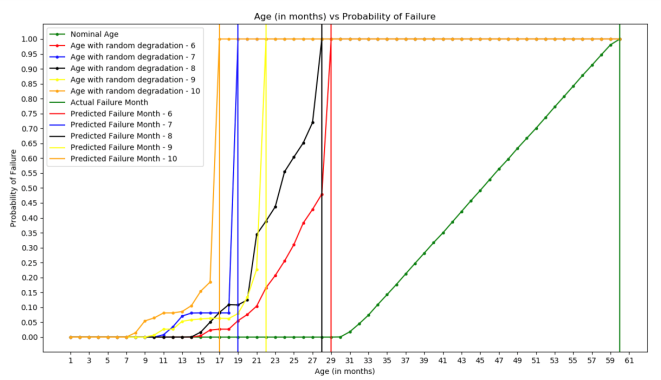


Figure 10. Visualization of predictions done using Random Forest Regression for scenarios 6-10

1. **R2 Score**

R2 Score is a statistical measure of how close the data are to the fitted regression line. The formula is:

where

SST= Total Sum of Square which is the total variation in the dependent variable

SSE= Sum of Squared Errors which is the amount of variability in dependent variable that is not explained by the model [10]

R2 is always between 0 and 1, 0 being the least and 1 being the highest.

The R2 scores of all the algorithms are listed in Table 2:

Table 2. R2 scores of Multiple Linear Regression, Decision Tree Regression and Random Forest Regression.

1. **CONCLUSION**

From the R2 Scores table, the results show that the Decision Tree algorithm scores constantly higher than Random Forest and Multiple Linear Regression algorithms.

Random Forest algorithm’s results were very close to Decision Tree’s and is much better than Multiple Linear Regression.

Multiple Linear Regression comes last out of the three.

All the data used in this project was based on a few assumptions. Further, the same process can be carried out and implemented on real world data from an actual aircraft engine to predict when it might actually fail.

1. **REFERENCES**
2. Simon and Donal L, ‘An IntegratedArchitecture for On-Board Aircraft Engine Performance Trend Monitoring and Gas Path Fault Diagnostics’, NASA Glenn Research Center, Cleveland, OH, United States, May 01, 2010
3. IremTumer and AnupaBajwa, ‘A survey of aircraft engine health monitoring systems’, 35th Joint Propulsion Conference and Exhibit, 1999
4. QI Shufen and ZHANG Wanying, ‘Prognostic and Health Management System based on Flight Data’, College of Automation and Electronic Engineer, Qingdao University of Science &Technology, Qingdao
5. M. Aznar Fernandez-Montesinos, P . Janssensand R.A. Vingerhoed, ‘Enhancing Aircraft Engine Condition Monitoring’, IFAC Emerging Intelligent Control Technologies, Hong Kong
6. Andrew K S Jardine, Daming Lin and Dragan Banjevic, ‘A review on machinery diagnostics and prognostics implementing condition-based maintenance’. Department of Industrial Engineering, University of Toronto, 3 November 2005
7. K Mathioudakis, A Stamatis, A Tsalavoutas and N Aretakis, ‘Performance analysis of industrial gas turbines for engine condition monitoring’, Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy, Jan 1, 2001
8. Louis A Urban, ‘Gas Path Analysis Applied to Trubine Engine Condition Monitoring’, Journal of Aircraft, Vol. 10, No. 7, 1973
9. Mark Tranmer and Mark Elliot, ‘Multiple Linear Regression’, Cathie Marsh Centre for Census and Survey Research
10. GuangliNie, Wei Rowe, Lingling Zhang, Yingjie Tian and Yong Shi, ‘Credit card churn forecasting by logistic regression and decision tree’, Expert Systems with Applications, Volume 38, Issue 12, November-December 2011
11. Neil J Salkind, Encyclopedia of Research Design, University of Kansas