A STUDY ON REMAINING USEFUL LIFE(RUL) PREDICTION OF AN AIRCRAFT ENGINE USING MACHINE LEARNING

[1]S. SOWMYA BHARATHI, [2]Dr.P.VIJAYLAKSHMI, [3] P. GAJALAKSHMI, [4]KRANTHI .E

[1]PG Student, [2,3]Assistant Professor, [4] Student

[1,2]Department of Mathematics, [3]Department of Software Engineering, [4]Mechanical Engineering

[1,2] Periyar Maniammai Institute of Science & Technology, Thanjavur-613 403, Tamilnadu, India

[3]CVRDE , AVADI

[4]Vellore Institute of Technology, Chennai Campus

Abstract

In this study we are mainly focused on predicting the Remaining useful life (RUL) of a Turbofan aircraft engine and the accuracy of the aircraft engine – Accuracy tells us the Performance of the engine. To ensure flight safety and reduce the cost of maintenance during aircraft engine Performance, a prognostics and health management scheme that focuses on flaw diagnosis, health assessment, and remaining life prediction is introduced to solve the problem. So, we are going to determine these parameters using Machine Learning with the help of Regression Analysis approach. We are going to use several Regression approaches and compare the Remaining Useful life and the accuracy with each of the approach and comment on the best Regression approach being used. We are going to work on Anaconda and Jupyter Notebook platform where we will implement the source code in Python. Finally, we would have a RUL value that would assign the health score to that engine

Keywords: Turbofan engine, Regression, Machine learning, Anaconda, Jupyter Notebook

1. Introduction

In an Aircraft Engine the failure of the engine is due to the the cause of accidents and lack of maintenance of the engine. Therefore, the safety and the reliability of aircraft engines (Turbofan engine in our case) are vital to the performance of aircraft. Basically, it is very difficult to ensure their safety and reliability due to their complicated structures and the engine failures has arisen inevitably due to effects of aging, environment, and variable loading as the working time increases. In the field of aircraft predictive maintenance, the traditional maintenance is either purely reactive (fixing or replacing an aircraft engine component after it fails) or it is blindly proactive (assuming that some certain level of superior performance degradation with not even a single input from the aircraft engine itself and maintaining the aircraft engine on a routine schedule whether maintenance is actually needed or not).

We are predicting the Remaining useful life using a turbofan engine, Turbofan Engine has the central engine core and it uses the 10 per cent of the intake air and 90 per cent of the intake air around the core is used to produce the thrust. Because the turbofan engine produces more thrust when the fuel burned in the core region and it mostly used for the passenger aircraft.

1.1 MACHINE LEARNING

In this Global World, the World is filled with data in fact a lot of data pictures, music, words, spread sheets, videos, and it doesn't look like it's going to slow down anytime soon. Machine Learning brings the promise of deriving meaning from all the data. The Value of Machine Learning is just beginning to show itself. There is lot of data in the world today; it is generated not only by people, but also by computers, phones and other devices. Traditionally, humans have analysed data and adapted systems to the changes in the data patterns. However, as the bulk of the data surpasses the ability for human beings to make a sense of it and it manually write those rules, we will turn gradually to automated systems that can learn from the given   data and importantly, the changes in data to adapt to shifting landscape.

1.1.1 HOW MACHINE LEARNING USED TODAY

We see Machine Learning all around us in the products we use today, while things like tagging objects and people inside of photos are clearly Machine Learning at play, In Youtube we see it automatically recommends are our next video to watch this is also powered

by Machine Learning. The biggest example of all is the Google Search, Every time you use Google search you are using a system that has many machine learning systems at its core,

From understanding the text of your required query or problems to adjusting the results based on your personal interests, such as knowing which results to show you first when searching. Today in this modern world Machine learning's immediate applications are already quite wide-ranging including image recognition, fraud detection and movie recommendation systems as well as text and speech systems too.
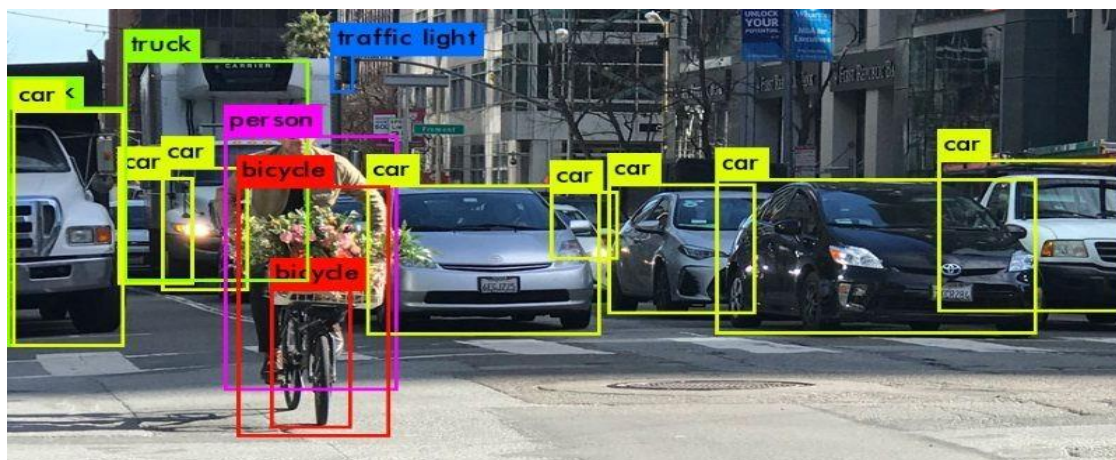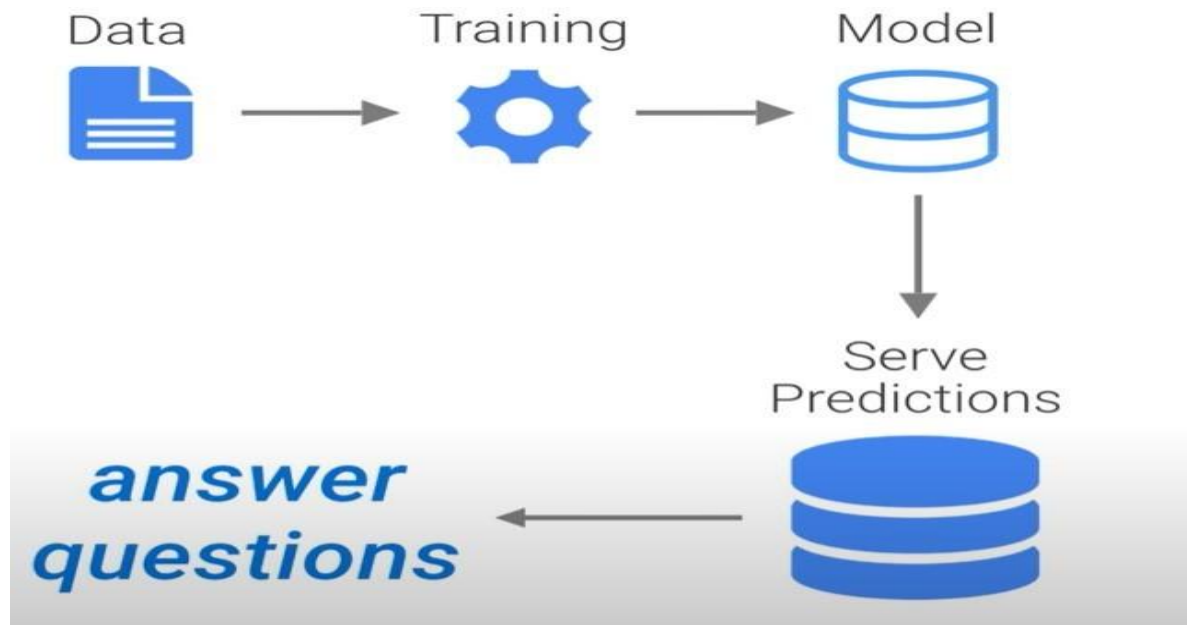

IMAGE RECOGNITION SYSTEM

Machine Learning can also be applied to wide range of fields, from diabetic retinopathy and skin cancer detection and even in transportation of self-parking and self-driving vehicles.

Above Flowchart shows how every Machine Learning model works

1.1.2 TYPES OF MACHINE LEARNING

1) SUPERVISED LEARNING:

Supervise Machine learning is a type of Machine Learning where we teach the machine using label data, so an input and your output is labelled data. Under supervised learning we have two main categories of problems, we have Regression and Classification Problems

Classification is basically about predicting a label or a class whereas Regression is about predicting a continuous quantity. Our Model and project is based on regression approach.

Types of Supervised Machine Learning Techniques

Classification   Regression
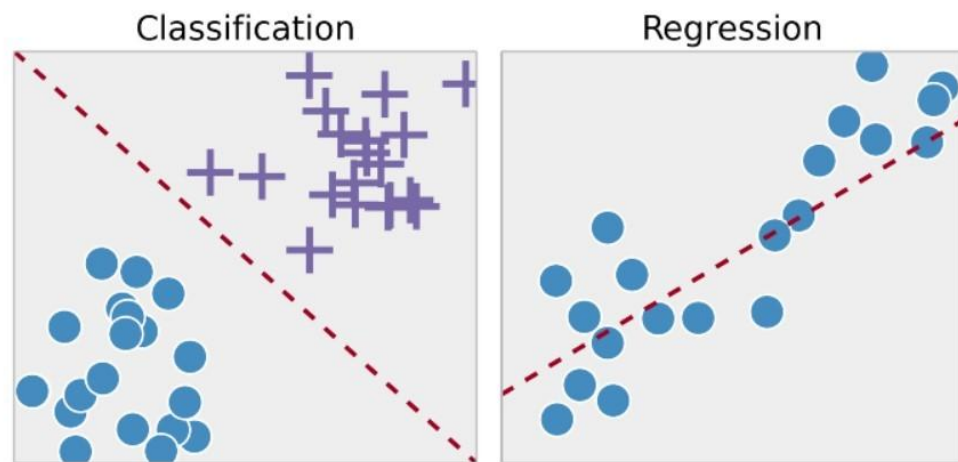
2) UNSUPERVISED LEARNING:

Unsupervised Learning is a machine learning method in which the users do not need to supervise the model. Instead, it lets the model to work on its own to discover the patterns and the data that was previously undetected. It mainly deals with the unlabelled data.

Unsupervised Learning Algorithms allow as much as users to perform more complex processing tasks when compared to supervised learning. Although, unsupervised machine learning model can be more and more unpredictable when it is compared with the other natural learning methods. An unsupervised learning algorithm mainly includes clustering, anomaly detection, neural networks, etc. Unsupervised Learning is basically clustering the model.

Clustering

sample   Cluster/group

3) REINFORCEMENT LEARNING:

Reinforcement Learning is defined as a Machine Learning technique that is always concerned with how the software agents should take their actions in an environment. Reinforcement Learning is a part of the deep learning technique that always helps you to maximize some portion of the cumulative reward.

This type of Neural Network learning method always helps you to learn how to attain a complex objective or to maximize a specific dimension over many steps.

Typical RL scenario
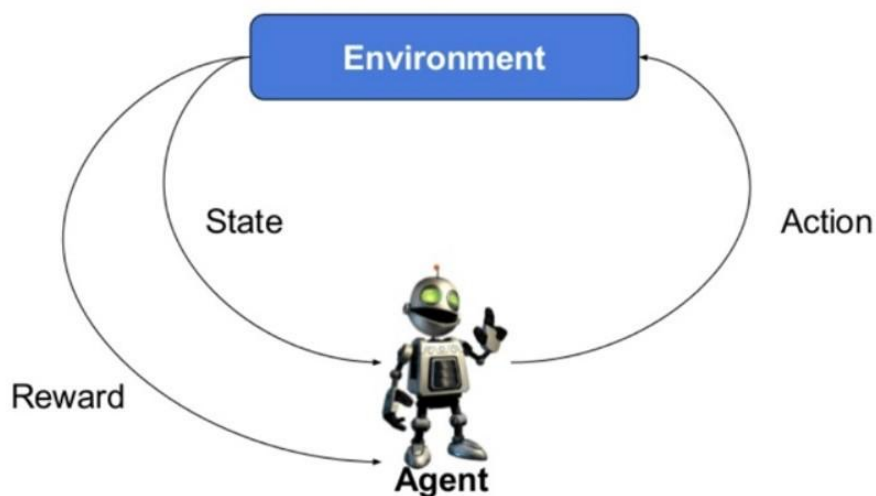
1.2 REGRESSION ANALYSIS

It is a predictive modelling technique to find the relationship among two or more variables so it is an predictive modelling technique to find out or to predict how is the relationship among the two variables or more variables that is where independent and which is the dependent variable which are the independent variable and how they are interrelated to each other. Regression is primarily used for prediction and casual inference in the sense to find out the relationship inference. There are Different types of Regression Namely:

1) LINEAR REGRESSION:

In Linear Regression the data is always modelled using a straight line and uses a straight-line equation. This type of regression is used in evaluating the trends and sales estimates, analysing the impact of the variable price changes, Assessment of risk in the financial services and in the insurance domain. In Linear Regression model the predicted output is always continuous and has a constant slope which is observed in the equation.

We can represent Linear Regression as:

$$y = a_0 + a_1X + \in$$

$y - dependent\ variable(Target\ variable)$

$X - Independent\ variable(predictor\ variable)$

$a_0 = intercept\ of\ the\ line\ (gives\ an\ additional\ degree\ of\ freedom)$

$a_1 = Linear\ regression\ coefficient(scale\ factor\ to\ each\ input\ value)$

$\in = random\ error$

2) LOGISTIC REGRESSION:

Logistic regression is the prospect of some obtained event is represented as a linear function of a various combination of predictor variables. Outcome of the logistic regression should be discrete/categorical such as:

Logistic Regression uses Categorical Variables and it mainly solves the classification problems, the curve of logistic regression is S-curve

Logistic regression equation can also be obtained from the linear regression equation as shown below.

- Equation of straight line can be written in the form as:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \cdots + b_n x_n$$

- In logistic regression the value of y can be between 0 and 1only

$$\frac{y}{1-y}; 0 \ for \ y = 0, and \ \infty \ for \ y = 1$$

- Taking logarithm of the function and we get the final equation

$$\log \frac{y}{1-y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \cdots + b_n x_n$$

3) RANDOM FOREST REGRESSION:

Random forest regression is one of the most prevailing supervised learning algorithms which is capable of performing regression as well as classification tasks. With the help of Random Forest regression, we can prevent the Over fitting in the model by creating random subsets of the dataset. The combination of these decision trees are called as base models and it can be represented in the form of:

$$g(x) = f_0(x) + f_1(x) + f_2(x) + \cdots$$

Random forest regression uses the Bagging or Bootstrap Aggregation method of ensemble learning in which aggregated decision trees runs in parallel and do not interact with each other.

4) POLYNOMIAL REGRESSION:

Polynomial Regression is very similar to that of multiple linear regression but at the same time instead of different varaibles like $x2, x3, x4, and\ so\ on \ldots x_n$ we have the same varaible x1 but it is different powers .In polynomial regression the equation varies with the degree(i.e degree which is more than one) .In the polynomial Regression the original model features are converted into the Polynomial Features of the requried degree and then it is modelled using the linear regression model.

Equation of polynomial Regression :

Linear regression Equation -

$$y = b_0 + b_1 X + \in$$

Multiple regression Equation (Logistic Regression) –

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \cdots + b_n x_n$$

Polynomial Regression Equation –

$$y = b_0 + b_1 x_1 + b_2 x_2^2 + b_3 x_3^3 + \cdots + b_n x^n$$

5) LASSO REGRESSION:

Lasso Regression is a method for automatically penalizing extra features, Lasso regression can set coefficient to a feature to zero. Lasso is expanded as Least absolute shrinkage and selection operator .This Regression helps us in reducing the overfitting but it not only helps in over-fitting it also helps us in feature selection . Feature selection using Lasso regression model technique can be well depicted by varying the regularization parameter.

Equation for Lasso Regression is:

$$L(X, Y) = Min\left(\sum_{i=1}^{n} (y_i - w_i x_i)^2 + \lambda \sum_{i=1}^{n} |w_i|\right)$$

6) RIDGE REGRESSION:

Ridge Regression is mainly to avoid the over-fitting. Ridge regression works by applying a penalizing term by reducing the weights and biases to overcome the over-fitting. This Over-fitting occurs when the trained model performs well on training data and performs poorly on the test datasets. This regression works by attempting at increasing the bias to improve the variance

Equation of Ridge Regression is:

$$L(X, Y) = Min(\sum_{i=1}^{n}(y_{i-}w_i x_i)^2 + \lambda \sum_{i=1}^{n}(w_i)^2)$$

1.3 ACQUIRING THE DATASET

For Predicting the Remaining useful life of the engine, we need the previous aircraft engine data set having 21 sensor values of each aircraft, **NASA** has created the prognostics and health Management PHM08 challenge data set has been NASA made publicly available, this data set is used to predict the failures and remaining useful life (RUL) of aircraft or jet engines over time**.**

Three different sets of datasets i.e. Train dataset, Test dataset, RUL dataset were simulated and analysed under different combinations of operational conditions and fault modes. Records of various and several sensor channels to characterize their fault evolution.

1. Data sets consist of multiple multivariate time series.

2. Each data set of the aircraft is further divided into training, test, RUL subsets.

3. Each time series is from a different set of engine i.e., the dataset can be considered to be from a set of engines of the same type.

4. There are three operational settings that have a substantial effect on engine performance. These operational settings are also included in the data. This aircraft data is contaminated with sensor noise.

5. In the training dataset, the fault develops in magnitude until system failure. In the test dataset, the time series ends in some time prior to the system failure.

6. The main objective is to predict the number of remaining operational cycles before failure in the test engine dataset, i.e., the number of operational cycles after the last cycle that the engine will continue to operate.

 7. Also provided a vector of true Remaining Useful Life (RUL) values for the test data.

8. The dataset are provided as a zip-compressed text file with 26 columns of numbers, separated by spaces.

The columns correspond to:

(a) Col 1 - Machine ID

(b) Col 2 - Current operational cycle

 (c) Col 3 - Operational setting 1

(d) Col 4 - Operational setting 2

(e) Col 5 - Operational setting 3

 (f) Col 6 - Sensor measurement 1

 (g) Col 7 - Sensor measurement 2 ……

(h) Col 26 - Sensor measurement 21

2.   METHODOLOGY

        This section introduces the relevant algorithms used in this research . As depicted in the below figure , the whole procedure for RUL prediction for an aircraft engine consists of two main steps : Data Processing and RUL Prediction using Regression analysis with and without Cross Validation .
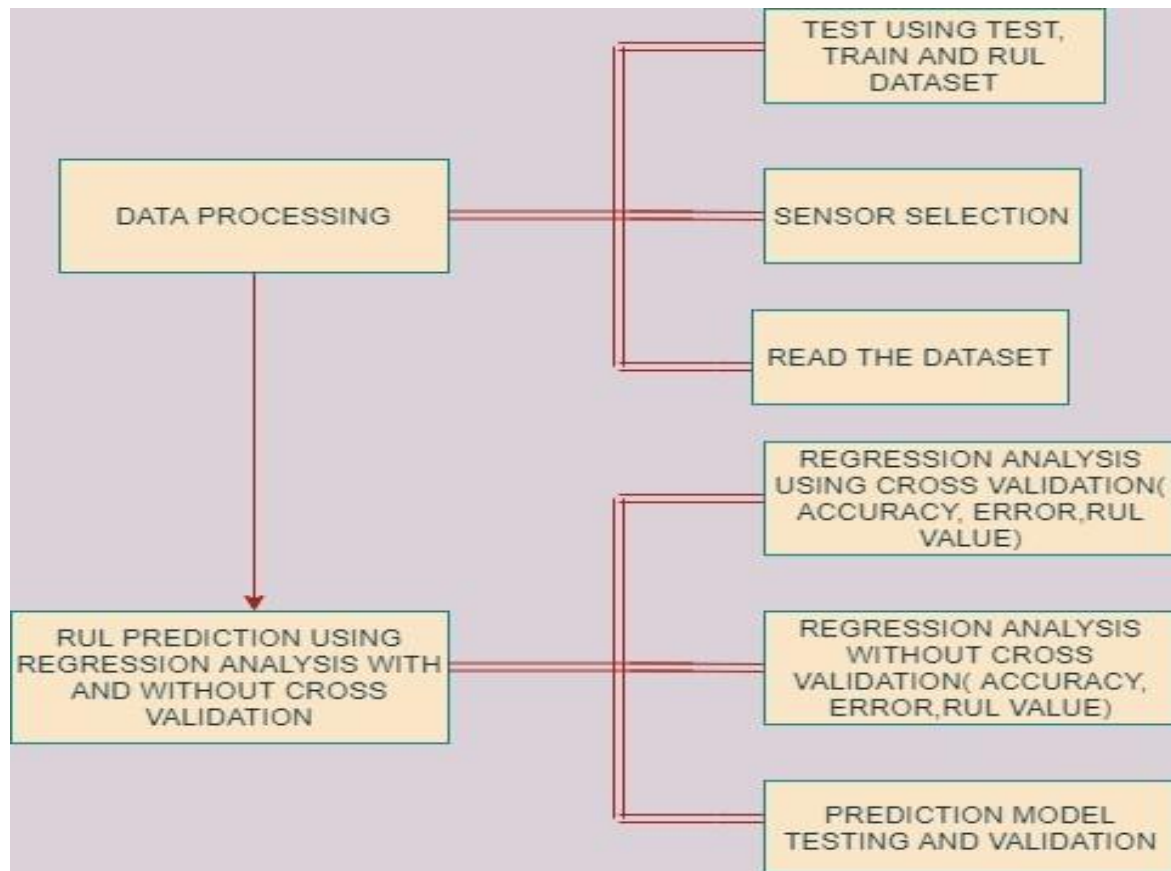
4.1 DATA PROCESSING

Initially Dataset is acquired. In our case the dataset is acquired from NASA data repository. In the NASA dataset, the dataset is divided into 3 datasets- Train Dataset, Test Dataset, RUL Dataset.

In the Train Dataset we have various parameters such as Machine I'd, Number of cycles, Operational setting 1(OP set1), OP set 2, OP set 3, Sensor 1, Sensor 2, Sensor 3,….. , Sensor 21. Training dataset is aircraft engine's run-to-failure dataset.

In the Test Dataset we have various parameters such as Machine I'd, Number of cycles, Operational setting 1(OP set1), OP set 2, OP set 3, Sensor 1, Sensor 2, Sensor 3,….. , Sensor 21. Testing dataset is aircraft engine's operating data without the failure events recorded.

We have almost 21000 values in the above datasets.

Index	Data Fields	Types	Descriptions
1	Id	Integer	Aircraft Engine Identifier
2	Cycle	Integer	Time, in cycles
3	Setting1	Double	Operational Setting 1
4	Setting2	Double	Operational Setting 2
5	Setting3	Double	Operational Setting 3
6	S1	Double	Sensor Measurement 1
7	S2	Double	Sensor Measurement 2
8	...	..	..
9	S21	Double	Sensor Measurement 21

In the above Dataset we have various sensor values but what we observed was in some sensors the values are same i.e. sensor 3 values is same as sensor 5 values and so on and for that we need to cut short the dataset so the Assortment of sensors that are sensitive to performance degradation and standardization of sensor data with varying dimensions are the primary tasks necessary to obtain a high RUL prediction accuracy. Three steps are needed process the data.

- So, Basically We need to test the dataset using the test, train, rul dataset.
- The next we need to select the required sensors to be used in the dataset . This step is required because when we choose the insensitive parameter data may reduce the RUL prediction accuracy.
- After the sensor selection, we will have the final dataset and after acquiring the final dataset we have to read the dataset and perform the required algorithms.

4.2 RUL PREDICTION USING REGRESSION ANALYSIS WITH CROSS-VALIDATION AND WITHOUT CROSS VALIDATION

Regression Analysis is a predictive modelling technique to find the relationship among two or more variables so it is a predictive modelling technique to find out or to predict how is the relationship among the two variables or more variables that is where independent and which is the dependent variable which are the independent variable and how they are interrelated to each other.

We have used Six Regression model techniques:

- LINEAR REGRESSION

- POLYNOMIAL REGRESSION
- LOGISTIC REGRESSION
- RANDOM FOREST REGRESSION
- LASSO REGRESSION
- RIDGE REGRESSION

4.2.1 CROSS VALIDATION:

Cross-validation technique is a resampling technique used to assess machine learning models on a limited data sample.

This technique has a single parameter called k that refers to the number of groups that a given data sample is to be split into. As such, the technique is often called as the k-fold cross-validation. When a specific value for k is chosen, it may be used in place of k in the place to the model, such as k=10 becoming 10-fold cross-validation.

The general procedure is as follows:

1.   Shuffle the following dataset randomly.
2.   Split the required dataset into k set of groups.
3.   For each of the unique or identity group:
    a) Take the group as a hold out or test data set
    b) Take the remaining or rest of the groups as a training data set
    c) Fit a model on the training set and evaluate it always on the test data set
    d) Retain the evaluation score and accuracy and discard the model
4.   Summarize the overall skill of the model using the sample of model evaluation scores.
        Importantly, each of the observation in the dataset sample is assigned to an individual group and stays in that group for the duration of the procedure. This means that each of the sample is given and used to train the model k-1 times.

After Performing the Regression Analysis and the cross-validation technique, we get the accuracy for each regression analysis and Error value for each regression analysis and the RUL value for each regression analysis.

3. RESULTS AND DISCUSSION

3.1 RESULTS:

Training and evaluation of the machine learning model was carried out using the dataset of the turbofan engine and the turbofan engine having 21 different sensor values. We tested using 6 algorithms i.e. linear regression, logistic regression, polynomial regression, Random forest regression, lasso regression, ridge regression. These six algorithms were carried out using the cross validation technique and without cross validation technique. So we have two sets of graph, one with cross validation and the one without cross validation.

First we take the one without cross validation – under this model we have to perform six regression techniques and each regression technique has two set of graphs. First graph is for validation of Prediction Vs Actual. First graph mainly tell us about the Validation Performance. Second set of graph is plotted against the Machine ID and RUL values. Both these plots were carried out for the six regression models. In similar way these two sets of graph was plotted even with the cross validation technique.

After Plotting the graph, we get the accuracy for each algorithm, we get the Mean squared error and mean absolute error and the final RUL values. All these values are mentioned in the below table.

3.1.1) REMAINING USEFUL LIFE PREDICTION
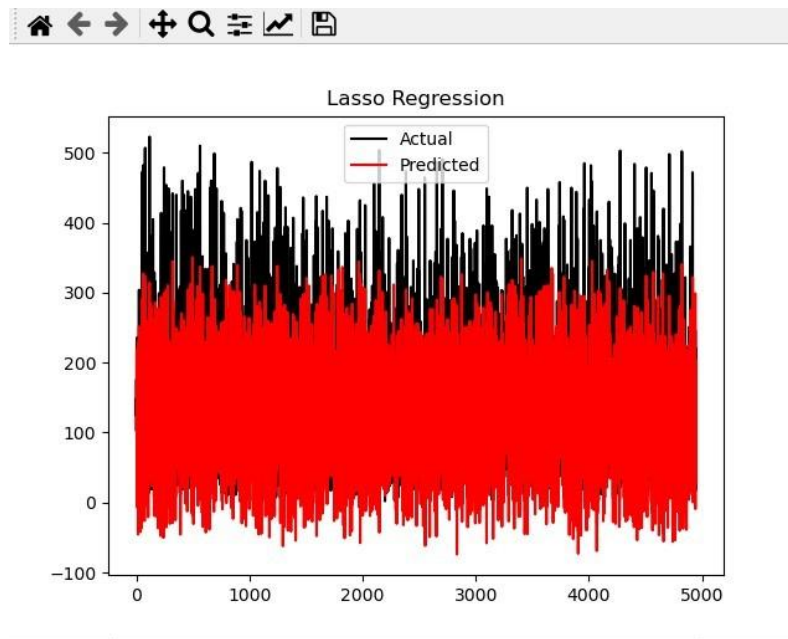
1) RANDOM FOREST REGRESSION
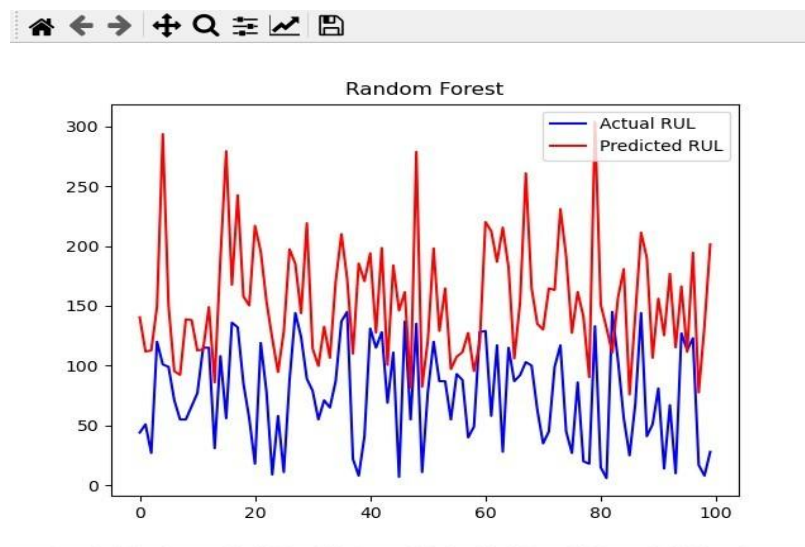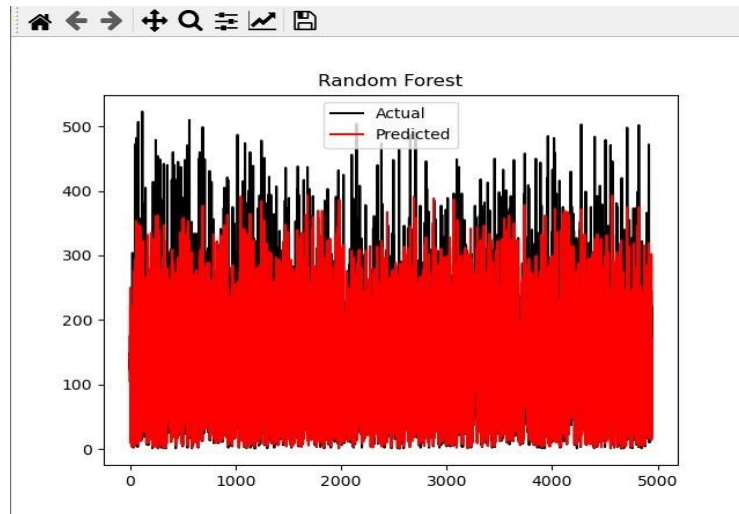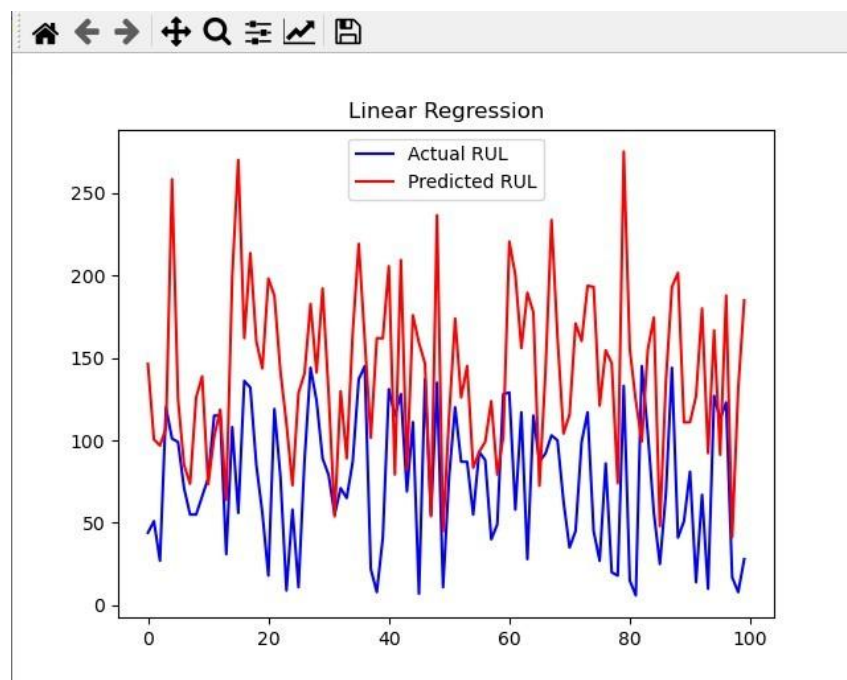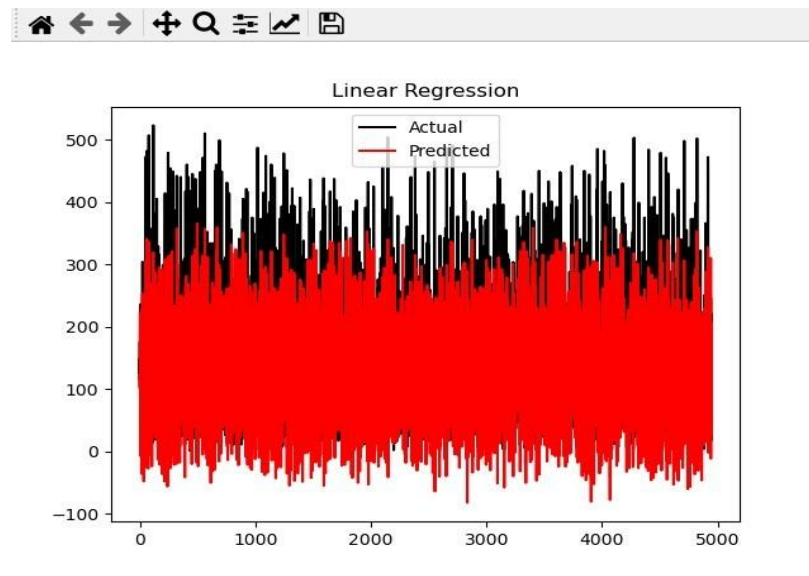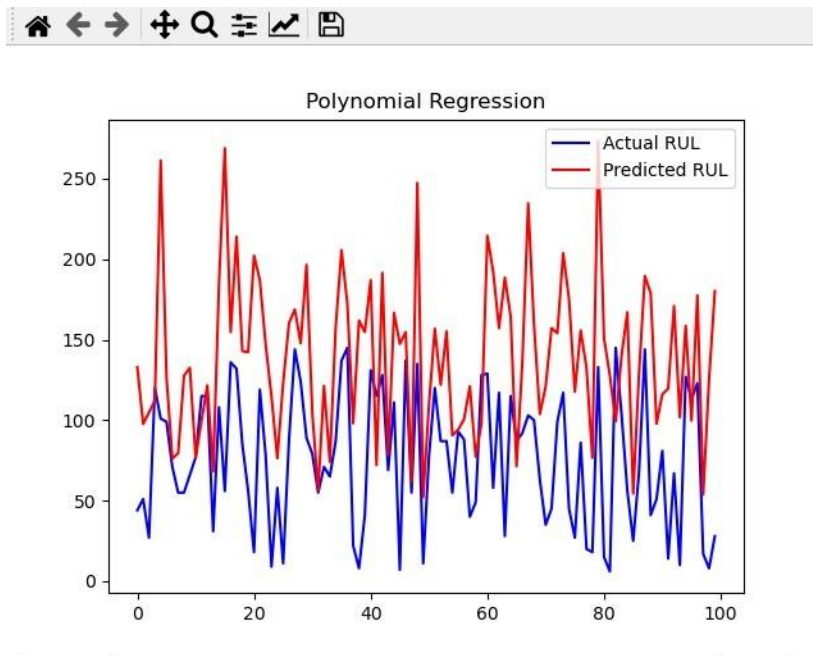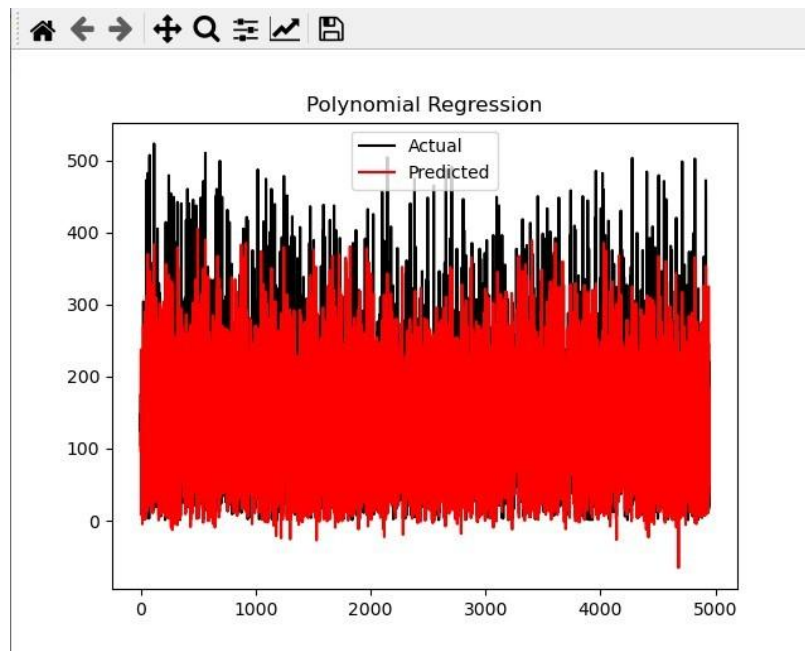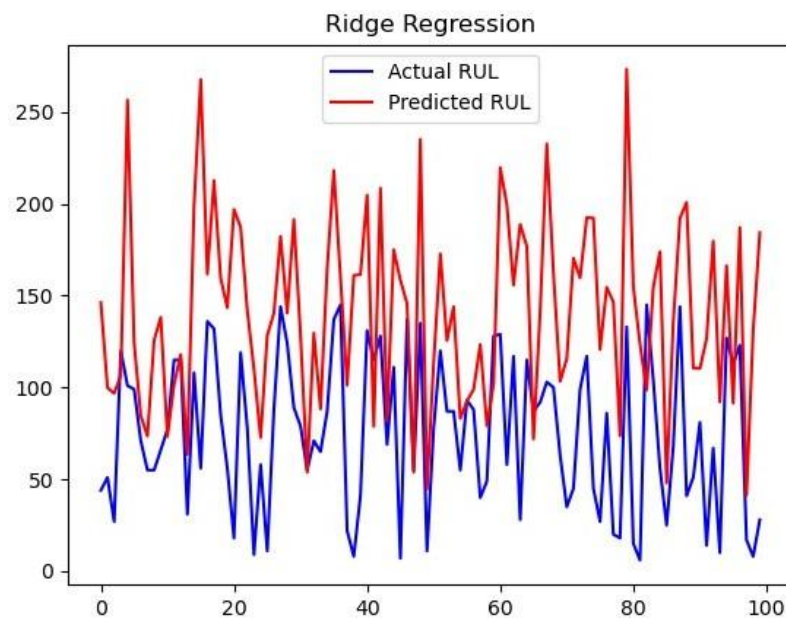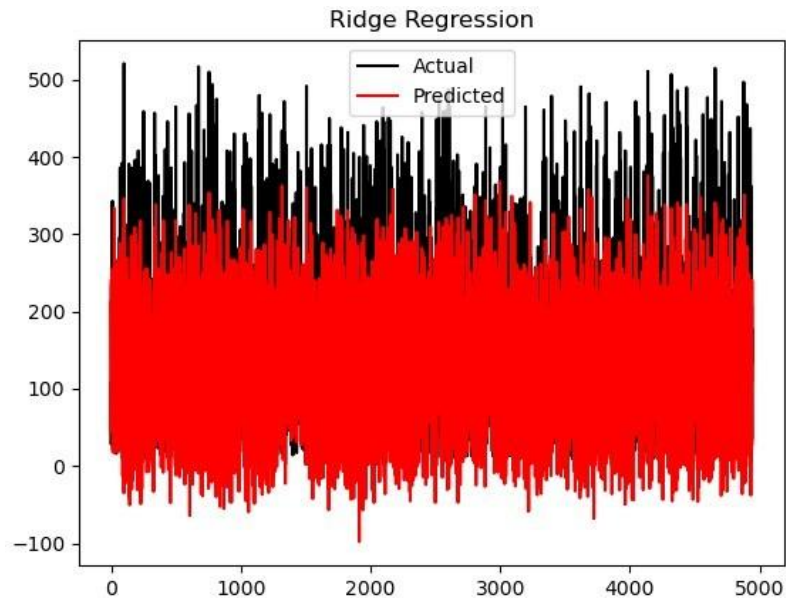
2) LINEAR REGRESSION

3) POLYNOMIAL REGRESSION
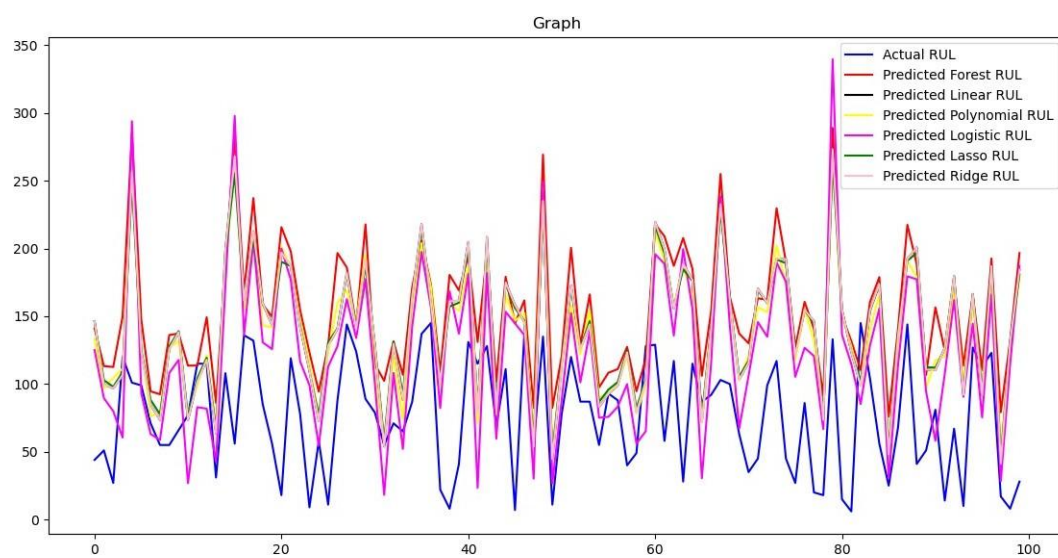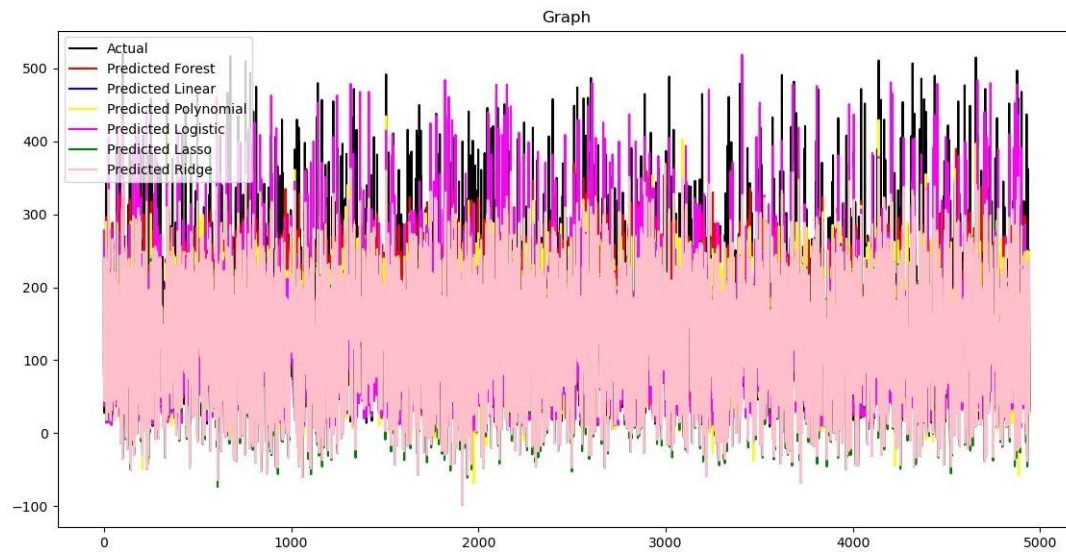
4) LOGISTIC REGRESSION
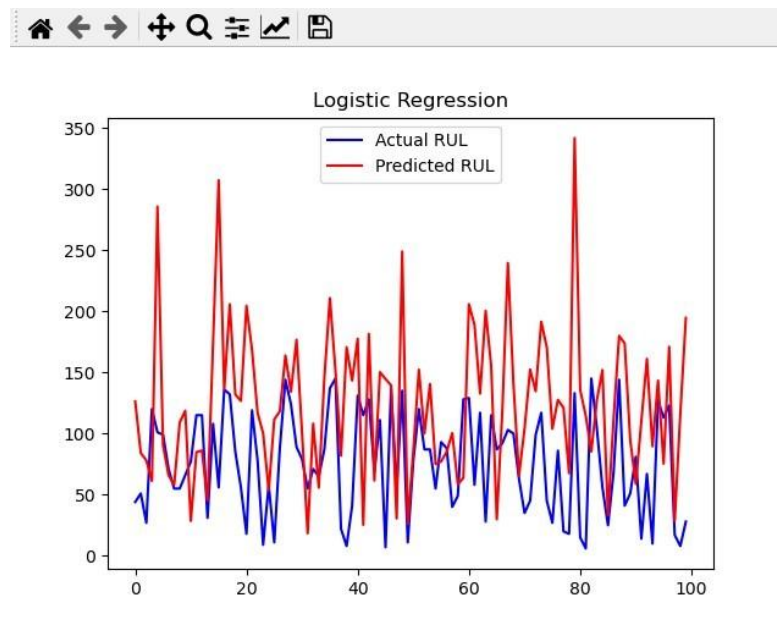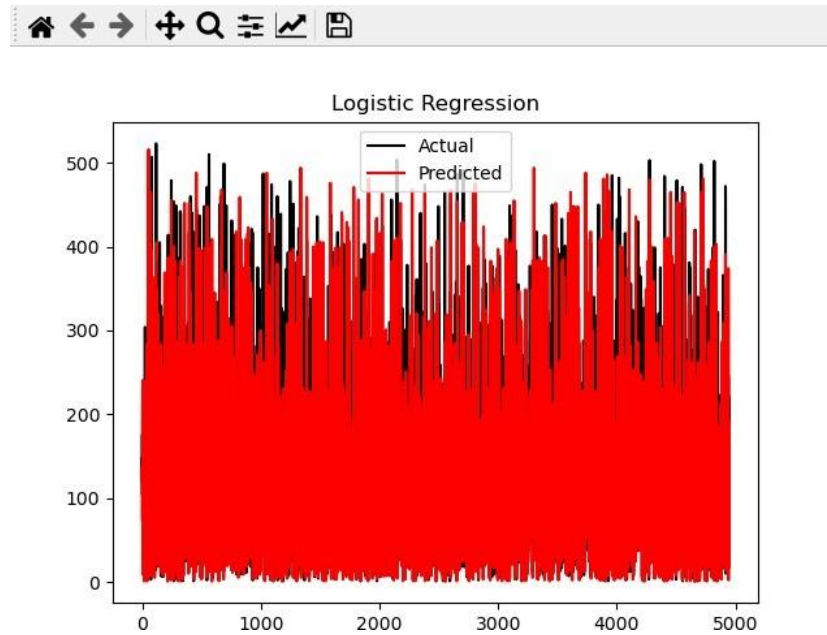
5) LASSO REGRESSION
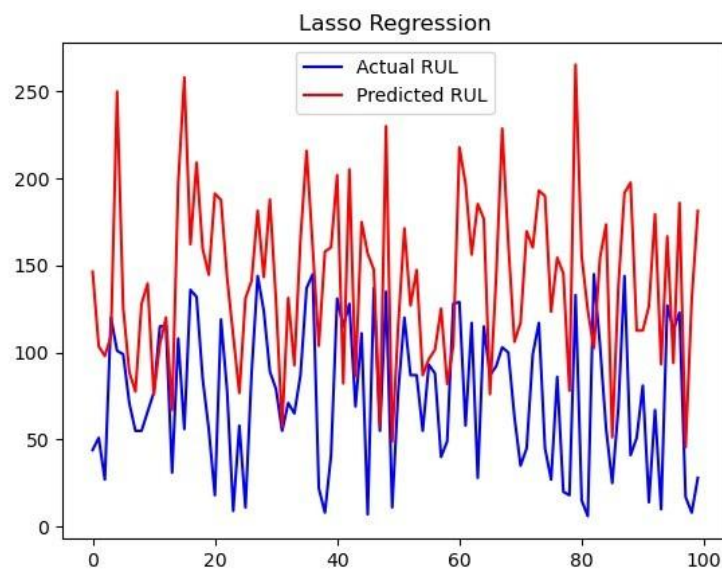
6) RIDGE REGRESSION

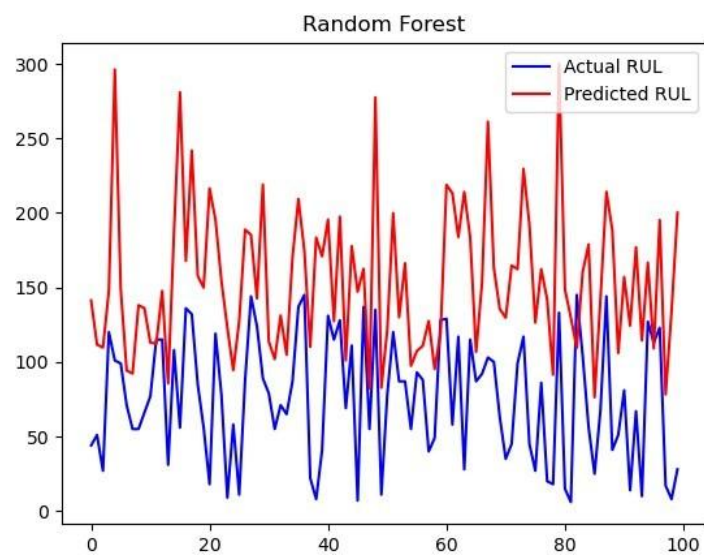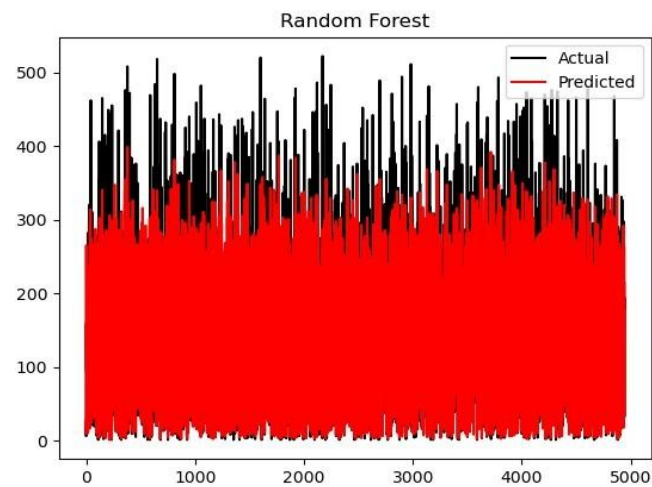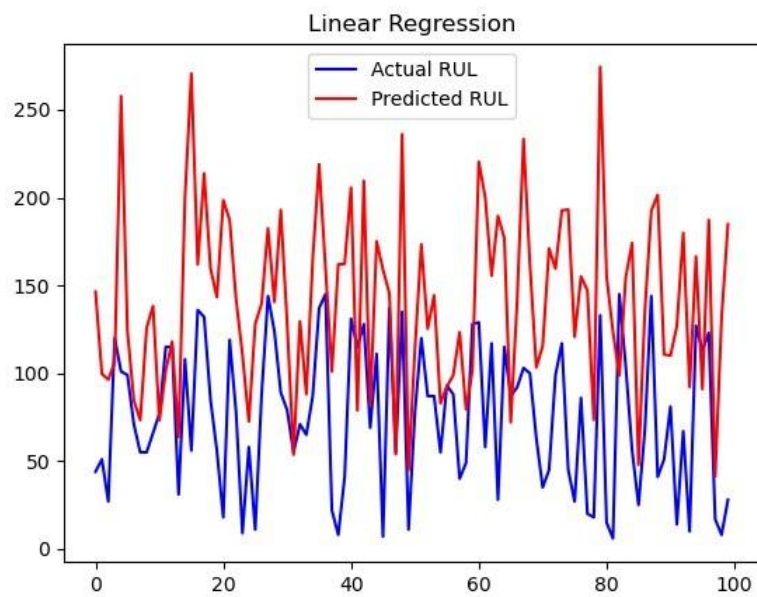**7) GRAPHICAL REPRENSATION OF THE ABOVE REGRESSION TYPES IN A SINGLE GRAPH (REMAINING USEFUL LIFE)**
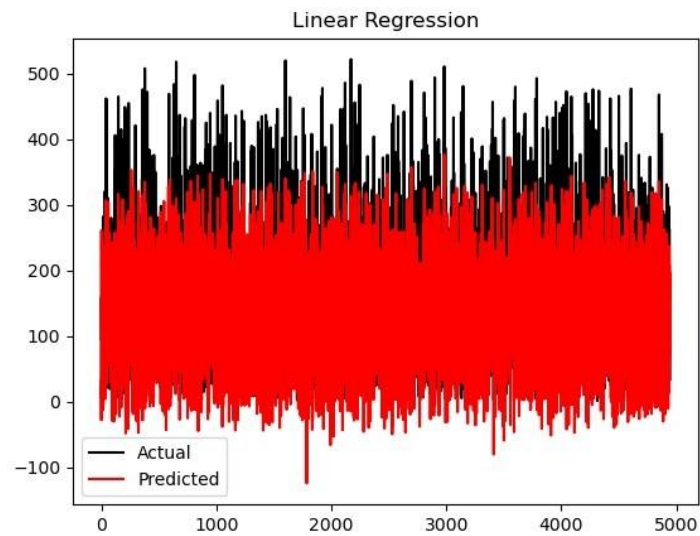
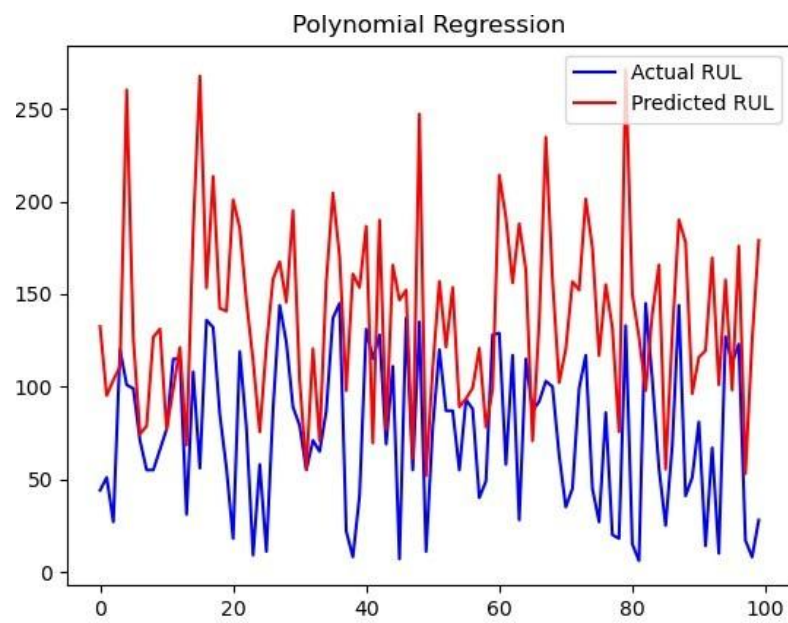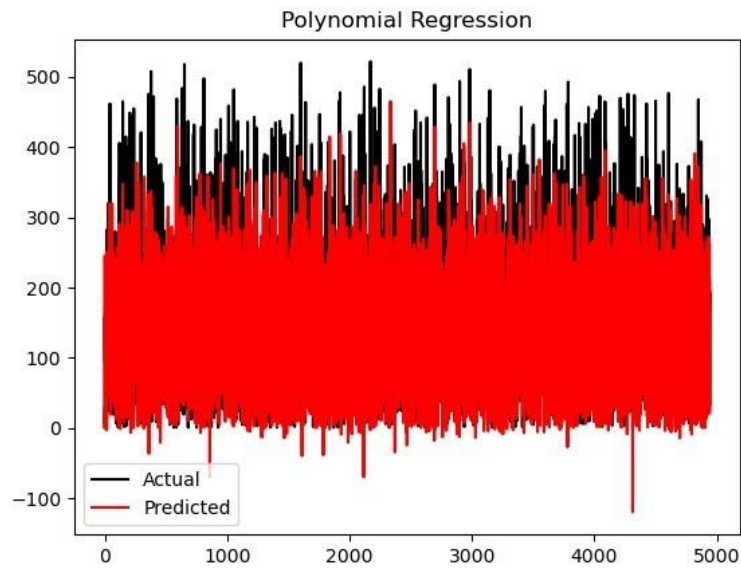**3.1.2) REMAINING USEFUL LIFE PREDICTION USING CROSS VALIDATION**
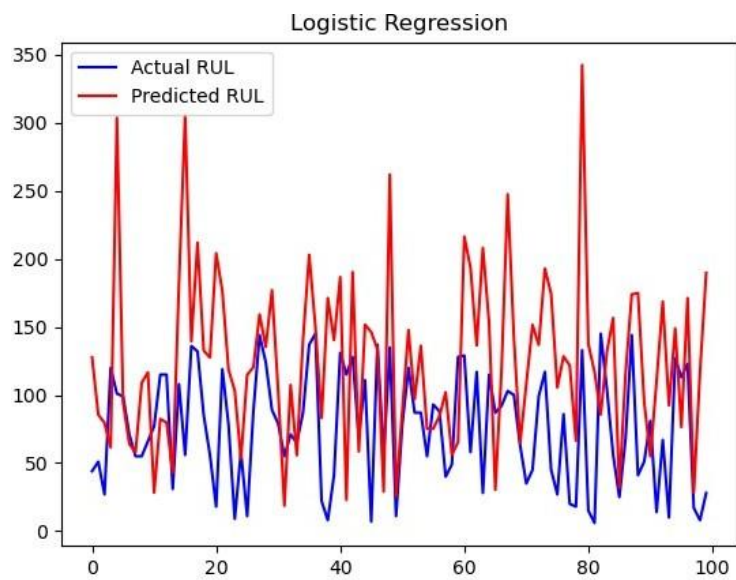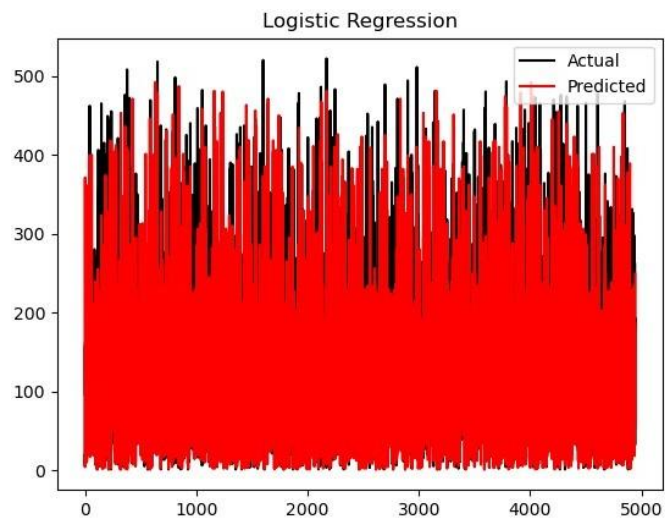
1) RANDOM FOREST REGRESSION
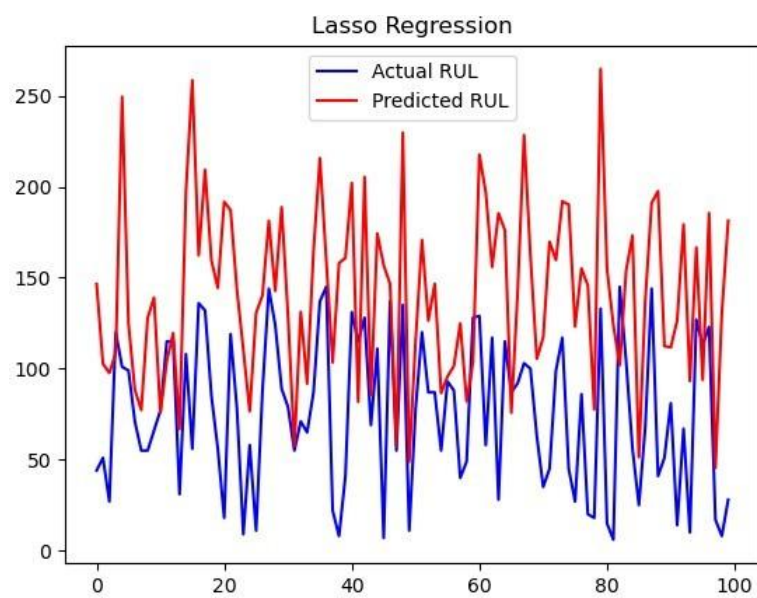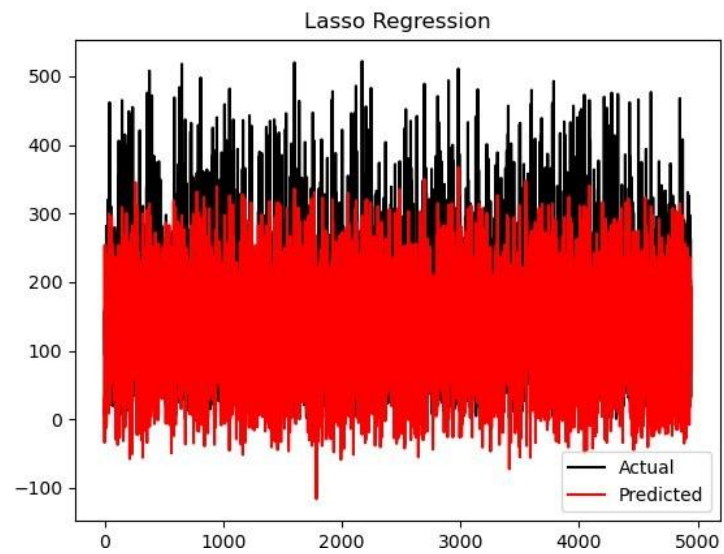

2) LINEAR REGRESSION

3) POLYNOMIAL REGRESSION

Polynomial Regression

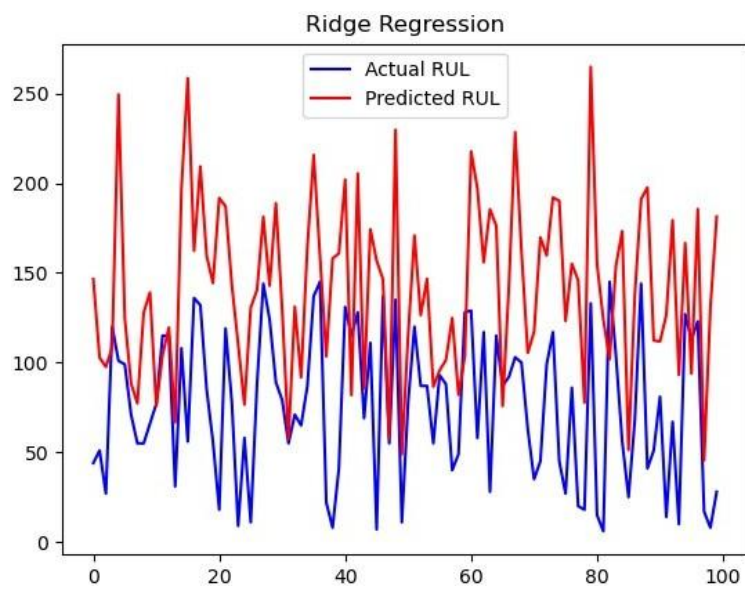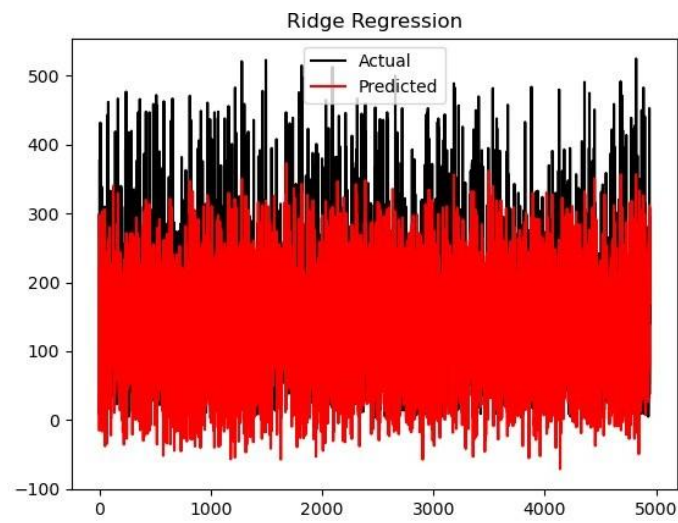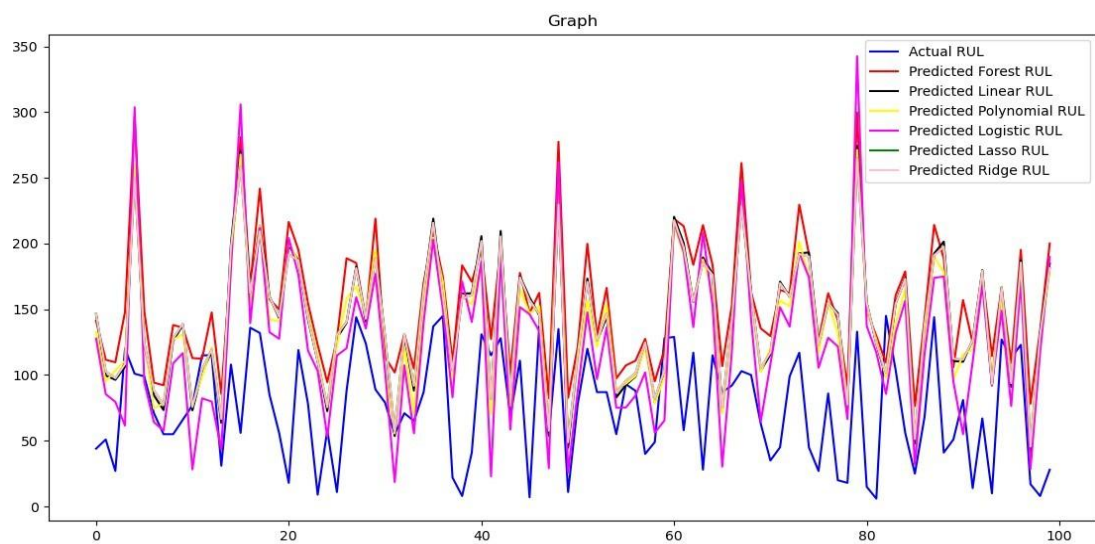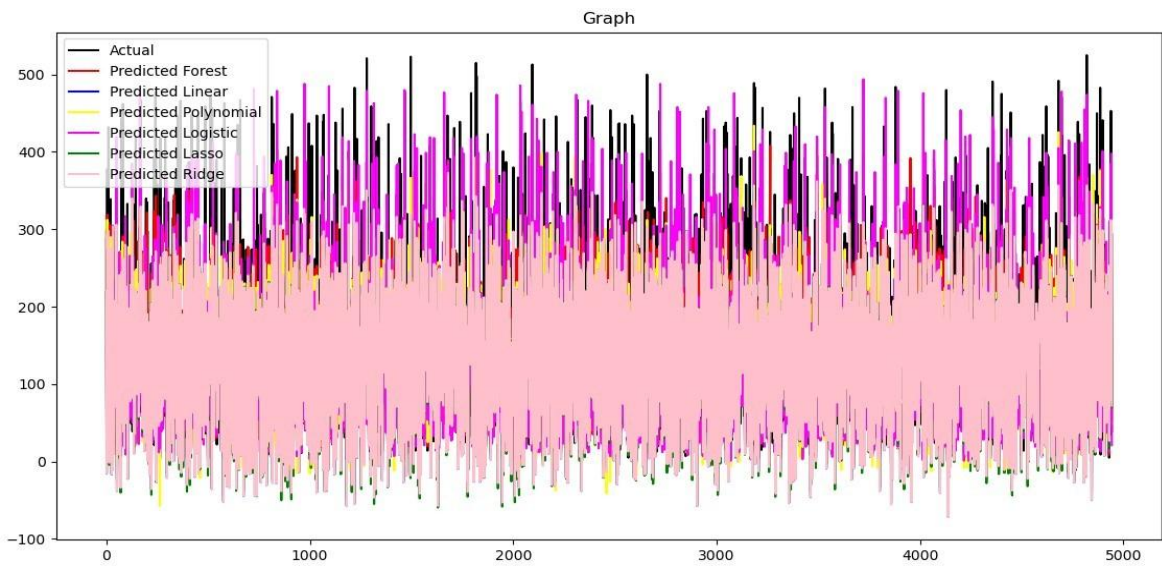

Polynomial Regression

4) LOGISTIC REGRESSION

5) LASSO REGRESSION

6) RIDGE REGRESSION

7) GRAPHICAL REPRENTATION OF THE ABOVE REGRESSION TYPES IN A SINGLE GRAPH (REMAINING USEFUL LIFE WITH CROSS VALIDATION)

3.1.3) TABLE VALUES

TABLE 1: ACCURACY

	ACC. OF RANDOM FOREST REGRESSION	ACC. OF LINEAR REGRESSION	ACC OF POLYNOMIAL REGRESSION	ACC OF LOGISTIC REGRESSION	ACC OF LASSO REGRESSION	ACC OF RIDGE REGRESSION
REMAINING USEFUL LIFE	0.7125941	0.64569419	0.68056245	0.0147851	0.6440175	0.6456946
REMAINING USEFUL LIFE USING CROSS VALIDATION	0.7145289	0.6480614	0.68942190	0.01646403	0.6461373	0.6480619

TABLE 2: RUL VALUE

	RUL OF RANDOM FOREST REGRESSION	RUL OF LINEAR REGRESSION	RUL OF POLYNOMIAL REGRESSION	RUL OF LOGISTIC REGRESSION	RUL OF LASSO REGRESSION	RUL OF RIDGE REGRESSION
REMAINING USEFUL LIFE	141.179887	146.354670	132.742205	125.894420	146.143091	146.254782
REMAINING USEFUL LIFE USING CROSS VALIDATION	140.530832	146.268523	132.473058	125.443347	146.093730	146.253181

TABLE 3: ERROR VALUE WITHOUT CROSS VALIDATION

	RANDOM FOREST REGRESSION	LINEAR REGRESSION	POLYNOMIAL REGRESSION	LOGISTIC REGRESSION	LASSO REGRESSION	RIDGE REGRESSION
MEAN SQUARED ERROR	2752.1623	3415.5712	3021.1920	4506.0220	3400.5380	3415.5585
MEAN ABSOLUTE ERROR	35.3137	43.292068	38.3822	44.0406	43.2418	43.2919

TABLE 4: ERROR VALUE WITH CROSS VALIDATION

	RANDOM FOREST REGRESSION	LINEAR REGRESSION	POLYNOMIAL REGRESSION	LOGISTIC REGRESSION	LASSO REGRESSION	RIDGE REGRESSION
MEAN SQUARED ERROR	2885.5945	3455.0862	3132.2615	4501.9514	3473.7391	3455.0798
MEAN ABSOLUTE ERROR	36.1754	43.1533	38.9259	43.9902	43.2929	43.1532

3.2 DISCUSSION:

The different sets of algorithms were evaluated in the same way on the same data for consistent comparison of results. For each of the dataset, the test results were compared with the actual values of the RUL available in the data set. The mean squares of the errors were observed for each regression technique and it is observed that the best results were obtained by random forest algorithm. Random Forest algorithm gives us the best accuracy, the better error values compared to others.. It was observed that the performance of all the algorithms were consistent in the datasets, generating proportional accuracy for the different algorithms tested.

4.   CONCLUSION AND FUTURE SCOPE

The main objective of predictive maintenance is to predict the remaining useful life of the engine. The Remaining Useful Lifetime prediction (RUL) has been carried out so as to plan the maintenance requirements of the turbo fan engine. By doing this kind of predictive maintenance, failures can be predicted and maintenance can be scheduled well in advance to

prevent the failure. This will reduce the cost and effort for doing maintenance. It increases safety of employees and reduces lost in production time. In our work, we have studied the performance of some set of machine learning algorithms. We have used regression techniques and cross validation technique. In future, the algorithms can be tested for more and more algorithms, more data and always be one step ahead in predicting the maintenance requirements.

5.   REFERENCES

[1] NASA DATASET - https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/

[2] Pecht M G. Prognostics and Health Management of Electronics, Encyclopedia of Structural Health Monitoring. Hoboken, New Jersey: John Wiley & Sons, Ltd, 2009.

[3] Haile M A, Riddick J C, Assefa A H. "Robust particle filters for fatigue crack growth estimation in rotorcraft structures," IEEE Transactions on Reliability, vol. 65, no. 3, pp. 1438–1448, 2016.

[4] Zhao F Q, Tian Z G, Zeng Y. "Uncertainty quantification in gear remaining useful life prediction through an integrated prognostics method," IEEE Transactions on Reliability, vol. 62, no. 1, pp. 146 –159, 2013.

[5] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112): Springer.

[6] Koenig, F., Found, P. A., & Kumar, M. (2019). Innovative airport 4.0 condition-based maintenance system for baggage handling DCV systems. International Journal of Productivity and Performance Management, 68(3), 561-577

[7] Lu, F., Wu, J., Huang, J., & Qiu, X. (2019). Aircraft engine degradation prognostics based on logistic regression and novel OS-ELM algorithm. Aerospace Science and Technology, 84, 661 - 671.

[8] Yan, W. (2006, Oct). Application of Random Forest to Aircraft Engine Fault Diagnosis. Paper presented at the The Proceedings of the Multiconference on "Computational Engineering in Systems Applications".

[9] Zhao, Z., Liang, B., Wang, X., & Lu, W. (2017). Remaining useful life prediction of aircraft engine based on degradation pattern learning. Reliability Engineering & System Safety, 164, 74 - 83.

[10] Liu, L., Wang, S., Liu, D., Zhang, Y., & Peng, Y. (2015). Entropy-based sensor selection for condition monitoring and prognostics of aircraft engine. Microelectronics Reliability, 55(9), 2092 - 2096.

[11] https://www.javatpoint.com/regression-analysis-in-machine-learning

[12] https://www.geeksforgeeks.org/machine-learning/

[13] http://www.sthda.com/english/wiki/regression-analysis-essentials-for-machine-learning#:~:text=Regression%20analysis%20consists%20of%20a,function%20of%20the%20x%20variables.

[14] Byington, C. S., Roemer, M. J., & Galie, T. (2002, March). Prognostic enhancements to diagnostic systems for improved condition-based maintenance [military aircraft]. Paper presented at the Proceedings, IEEE Aerospace Conference.

[15] Wang, T.; Yu, J.; Siegel, D.; Lee, J. A similarity-based prognostics approach for remaining useful life estimation of engineered systems. In Proceedings of the International Conference on Prognostics and Health Management, Denver, CO, USA, 6–9 October 2008; pp. 1–6

[16] Wang, D.; Tsui, K.L. Brownian motion with adaptive drift for remaining useful life prediction: Revisited. Mech. Syst. Signal Process. 2018, 99, 691–701

[17] Zhu, S.P.; Huang, H.Z.; Peng, W.; Wang, H.; Mahadevan, S. Probabilistic physics of failure-based framework for fatigue life prediction of aircraft gas turbine discs under uncertainty. Reliab. Eng. Syst. Saf. 2016, 146, 1–12.

[18] Si, X.S.; Wang, W.B.; Hu, C.H.; Zhou, D.H. Remaining useful life estimation: A review on the statistical data driven approaches. Eur. J. Oper. Res. 2011, 213, 1–14.

[19] Saxena A., Goebel K., Simon D., Eklund N. Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation; Proceedings of the 1st International Conference on Prognostics and Health Management (PHM08); Denver, CO, USA. 6–9 October 2008.

[20] Xiang Li, Qian Ding, Jian-Qiao Sun , "Remaining Useful Life Estimation in Prognostics Using Deep Convolution Neural Networks" in the journal "RELIABILITY ENGINEERING AND SYSTEM SAFETY", Apr 2018.