International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 02 | February - 2023 Impact Factor: 7.185 ISSN: 2582-3930

CONDITION BASED PREDICTIVE MAINTENANCE ON SHIP'S MAJOR EQUIPMENT USING AI

D Mohakul[1], CRS Kumar[2], Sweta Singh[3], Sudamman Katti[4]
[1]Defense Institute of Advanced Technology, Pune, [2]Defense Institute of Advanced Technology, Pune, [3]Defence Research and Development Organization, Pune, [4]Vishwakarma Institute of Technology, Pune
[1]mohakuldev@gmail.com, [2]suthikshnkumar@diat.ac.in, [3]shwetasingh.rde@gov.in, [4]sudaman.katti19@vit.edu

Abstract–This project mainly aims at diagnostics and prognostics approach for Main Engine fitted onboard Ship. Keeping in mind the complexity of equipment's fitted in ship and the frequency of defect, there is always a demands for prognostic maintenance approach rather than TBM (Time based Maintenance). Under these circumstances, the CbPM(Condition based predictive maintenance) helps in predicting the fault of a running machine and signals to perform maintenance on it only when required, thereby impacting operational readiness and logistic stability. Condition-based maintenance (CBM) is a program that banks upon three factor- data processing, maintenance decision making and data acquisition. The paper mainly focuses on implementing CBM on a mechanical system with emphasis on technology, algorithm and model for data processing. This project focuses on implementing CBM on machines rather than TBM by knowing the past defect history; maintenance schedule, periodic routine and then applying AI tool on the same for predicting the machines future condition. This will thereby help in improving the overall life of the equipment. An instance on the use of various classification algorithms is taken into picture and compared their accuracy based on the F1 score in this paper.

Keywords -Diagnostic*, Prognostics*, Condition-based predictive maintenance (CbPM), Main Engine, PPM (Planned preventive Maintenance)

I. Introduction

Engineering maintenance and prognostics are very crucial for the equipments fitted on-board ships. The main aim for prognostic analysis and diagnostic is solve the hindrance caused to the operational capability of the ship with non-availability of major equipment before any important exercise or operation. The condition-based predictive maintenance on ships engine will not only help in predicting faults in particular part of the engine, but also helps in logistic support beforehand. However, if equipment is not maintained correctly, then there are several effects such as high equipment downtime, impact on operational capability which further has security issues on stake. There are mainly three types of maintenance philosophies.

(a) <u>Reactive Maintenance</u> The maintenance team works on the defect post occurrence. It's time consuming and impact economically.

(b) <u>Preventive Maintenance</u> The periodic maintenance that is scheduled either by OEM or maintenance team in order to avoid regular breakdown in machinery. Preventive maintenance can be further classified into two types:-

(i) <u>Time-based-preventive maintenance</u>:- The time based maintenance depend on the OEM manual which mention the maintenance work to be carried out at stipulated time.

(ii) <u>Usage-based-preventive-maintenance</u>:-It's the maintenance schedule that that the team accept based on the running hours of the equipment, here again OEM manual is referred for the same.

(c) <u>Predictive maintenance: -</u> The aim of predictive maintenance is to work on prognostic maintenance approach for the equipment fitted onboard, thereby becoming better in predicting when the next malfunction will probably happen. The aim is to facilitate technician to be able to use tool-assisted analytical technique to answer definitively whether small maintenance today could prevent major repair down the line.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 02 | February - 2023 Impact Factor: 7.185 ISSN: 2582-3930

Marine Engine- Background

The marine diesel engines are mainly four stroke 20 cylinder or 16 cylinder heavy duty and robust engines. The engines are mainly V type. The engines are mainly sea water cooled and require the air pressure of 30 bars for cranking. The engine is divided into two banks as A bank and B bank where ignition and exhaust work simultaneously. The air intake inside the piston is controlled by the turbo-charger. The rpm of turbo-charged goes up to 1, 00,000 rpm. The maximum power output of 20 cylinder engine is approximately 9100kw. The power output from the engine is transferred to the shaft where it is connected with the gear box which controls the power and help in propeller movement resulting in ship's movement. Cut section of the Engine is shown in the fig 1.

fig 1

(a) **Operation basics**-: The marine engines fitted on-board Indian Coast Ship are mainly 20 cylinder or 16 cylinder 4 stroke diesel engine. The power is generated at the shaft of the main engine which is further connected to the gear box for power conversion. The overall operation provides torque to the ship. The basic parameter in picture while the operation of these heavy duty engines are-:

(i) Exhaust gas temperature
(ii) lub oil temperature
(iii) lub oil pressure
(iv) fuel pressure
(v) fuel temperature
(vi) Coolant temperature
(vii) Coolant pressure
(viii) Exhaust pressure
(ix) RPM

Apart from above parameter there are also other parameters that comes into picture while the operation of Main Engine. However, for the present paper we are going to take the aforesaid parameters and predict the engine condition and locate the exact fault.

(b) Monitoring and Maintenance -: Monitoring the big data from the engine and analyzing the data include the engine maintenance program. It must include following areas for maintenance-:

(i) Engine performance monitoring based on parameters mentioned in 2.1(a).
(ii) Vibration monitoring
(iii) Exhaust gas temperature
(iv) Frequent visual inspection.

Individually, each of these is important indicators and collectively they provide a complete picture as to the actual condition of the engine.

Scope

The scope of this paper is to develop a set of AI model to cover the following-:

(a) Probability Failure Model: This model predicts the probability of failure based on past failures and behavioral patterns of the machinery parameters. It keeps on mapping the parameter deviation and learn the co-relation between different parameter that lead to failure.
(b) Failure Cause Analysis Model: These are used to obtain the cause of past failure and thus predict the future failure causes. At the initial-instances, it will deduct the failure cause based on the pre-defined algorithm and accept manual input from user to confirm the exact cause of failure and apply reinforcement learning for predicting future causes of failure. This helps the probability failure model to predict the failures.
(c) Anomaly Detection Model: This model is an analytic model defining the deviation from the normal behavior and analyses the deviation pattern.
(d) Maintenance Model.**:** The above discussed three models provide input to this model to predict the maintenance and give suggestion to the operator for the required maintenance.
(e) Health Score Model.: The health score metric will intelligently correlate and display the health dashboard that will aid the user in decision making. This serves as a background for operator to minimize the maintenance cycles.
(f) Optimization Model.: This model analyses and provides operator the action to be performed to use the ship resources optimally. for e.g-:

- Fuel consumption analysis: This will understand the consumption pattern of fuel to indicate the health status of the main engine and generator. Consequently, providing suggestion for optimum speed requirement based on available fuel.
- Analysis of ship load consumption pattern and provide suggestion to improve /optimize the consumption and use of the generators.
- Analysis of the tank liquids to intimate the total available liquid and expected time of consumption and to focus on the critical requirement of ship for long survivability.
- The operating hour of ship loads including the auxiliary pumps etc. will provide a better automation to use the pump wisely.

II. RELATED WORK

(a) Giuseppe Aiello, Antonio Giallanza, Salvatore Vacante, Stefano Fasoli, Giuseppe Mascarella,"Propulsion Monitoring System for Digitized Ship Management: Preliminary result from a case study", ISM 2019

The paper state that the condition of the engine can easily be diagnosed, if the EGT (Exhaust gas temperature) is only taken into consideration. Since the exhaust gas temperature reading defines the health of the piston, hence higher or lower cylinder temperature defines the "Faulty behavior of the engine". The author demonstrates the procedure and utility of this method with numerical representation.

(b) Rosmaini Ahmad, Shahrul Kamaruddin,"An overview of time-based and condition-based maintenance in industrial application", Computer and Industrial engineering, Elsevier, August 2012.

A brief comparison between CBM and TBM technique is discussed in this paper. The author tries to state that TBM being a traditional technique is time consuming and not economical viable. On the other hand the CBM technique helps in predicting the future condition of the equipment and rectifying small defect today which thereby can prevent major repair down the line.

(c) N Balakrishnan, Angello I.Devasigamani, K.R.Anupama, Nitin Sharma, "Aero-Engine Health Monitoring with Real Flight data using Whale Optimization Algorithm Based Artificial Neural Network Technique", BITS Pilani, Dec 2020.

This paper has proposed that Whale Optimization Algorithm based artificial intelligence, is one of the best algorithms in training any neural network incorporating with wide range of optimization problem. The paper also states that aircraft engine health depend on engine pressure ratio, fuel flow, RPM and EGT (Exhaust gas temperature).

(d) Nicolas Nadai, Arthur Melani, Gilberto Souza, Silvio Nabeta,"Equipment Failure Prediction Based on Neural Network Analysis Incorporating Maintainers Inspection Finding", IEEE,2017.

The author has described about "Radial basis function Neural Network" which is commonly applied for classification and regression has proved to be a technique in monitoring the operating condition of a hydro generator and its result can be used for predictive maintenance as experimented by the author.

(e) Longlong Zhang, Zhiliang Liu, Dashuang Luo, Jing Li and Hong-Zhong Huang "Review of Remaining useful life prediction using support vector machine for engineering assets", Quality Reliability risk Maintenance and safety Engineering (QR2MSE) 2013 International Conference on,pp.1793-1799,2013.

The author has given importance of RUL of machinery so in order to predict its life cycle and reduce the maintenance cost. The author has also used SVM algorithm on small set of dataset and also discussed the importance of using other algorithm for improvement so in order to predict the RUL of a machine. The author has also stated the field on predicting the RUL is open and more problems can be solved in future.

(f) Value addition to above Review

The above research papers forecast prediction approach on various machineries that are fitted on static platform or on aircraft engine, however research on Marine engine fault prediction is yet to be carried out which become a challenging task view the system being fitted on a dynamic platform and the parameters impacting the engine condition can become highly uneven because of external environment (bad sea state).

III. PROPOSED METHODLOGY

(a) Data Collection and Methodology

In order to analysis the project based on the parameter mentioned above, it become important to collect relevant data. The data collection was very hectic task keeping the data confidentiality intact. Low volume of data was collected from a marine ship in Pdf format (175 data point) and the same was manually typed into .CSV format. Synthetic data was produced as the data set was very less for analysis of different set of algorithm. The detail of the data generation is enumerated below-:

Step 1-Synthetic Data Generation (Gaussian Copula and Oversampling)

The data consisted of six features, namely, lubricating oil pressure, lubricating oil temperature, coolant temperature, coolant pressure, and fuel pressure and engine rpm. Based on these features, the target variable by the name of engine condition is to be predicted, if engine condition=0, the engine will fail and if engine condition=1, no failure will take place. The data consisted of 174 points, since this amount of data is insufficient for Implementation of algorithms, generation of synthetic data using Gaussian copula was done. The Gaussian (or normal) copula is the copula of the multivariate normal distribution which is defined by the following:

$$C_{\text{Gaussian}}(u_1, u_2; \rho) = \Phi_\rho\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right)$$

where $ø_p$ is a joint distribution of a multi-dimensional standard normal distribution, with linear correlation coefficient ρ, ø being the standard normal distribution function. The marginal probability distribution of each variable is uniform on the interval [0,1]. Copulas are utilized when you need to model the dependency structure of a multivariate distribution. For example, they are used frequently in finance to capture the dependency structure between multiple time series. However, copulas are not well-defined on non-stationary data. Copulas are used to model the dependence between random variables. The foundation of Copulas is centered around Sklar's theorem which states that any normal multivariate distribution can be represented through univariate distributions and a copula. Synthetic data is then generated along the multivariate distribution. The python library called synthetic data vault was used for this synthetic data generation. However, before performing this step, oversampling was required as the original data was unbalanced. The original data consisted of only 3 data points with engine condition=0, so to tackle this issue, random oversampling was done which basically created copies of the original 0 engine condition points and ramped up the number to 51 from 3. Gaussian copula was then applied to this new data to generate two lakh original data points. The distribution of this new data as per engine condition was as follows-

	Engine rpm	Lub oil pressure	Fuel pressure	Coolant pressure	lub oil temp	Coolant temp
Engine Condition						
0	68685	68685	68685	68685	68685	68685
1	131315	131315	131315	131315	131315	131315

Table 1. Data distribution based on classes

After generation of synthetic data, target variable 'y' was defined to be engine condition while the rest of the features were denoted by variable 'x'. After this, train test split was done and 30 percent of the data was used for testing of models. Data was also generated using Generative Adversarial Networks (GANs) for comparison.

(b) Model Framing

Different algorithm was used on the set of generated data to check the most efficient algorithm that can be applied on real data-:

Step 1-Application of Logistic Regression

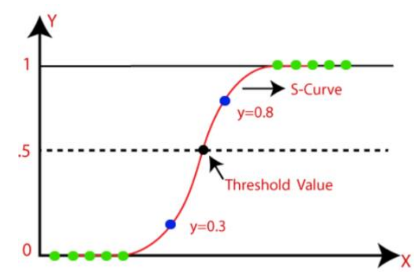

Fig 2:Logistic Regression

Logistic regression is considered to be one of the best Machine Learning algorithms that fall under Supervised Learning technique. It predicts the value of dependent variable based on the set of independent variable. The outcome is moreover categorical or in other word discrete which can be 0 and1 or true or false. Though logistic and linear regression both are used to solve the classification problem, however their approach of solving varies. The cost function used in this case for minimization is called binary cross entropy

Sigmoid Function-

$$\frac{1}{1+e^{-X_i \cdot \theta}}$$

Binary cross entropy cost function-

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1-y_i)\cdot log(1-p(y_i))$$

Step 2-Support Vector Classifier

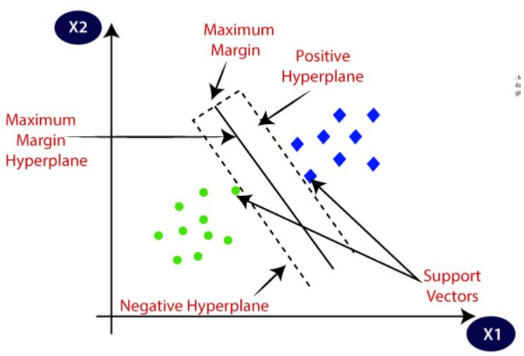

Fig 3:Support Vector Classifier

Support Vector Machine or SVM is considered to be most popular Supervised Learning algorithms, which is used for solving classification and regression problems. The SVM algorithm creates the line or decision boundary that can segregate n-dimensional space into classes. This decision boundary is called a hyper plane.. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.

SVM can be of two types:

Linear SVM: It mainly separate data linearly, which means the dataset are classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier used is called as Linear SVM classifier.

Non-linear SVM: When the data are not linearly separable, hence Non-linear SVM is used. It uses the classifier to convert the low dimensional data to high dimensional data which help in separating the data easily using hyper-plane.

Step 3- K neighbors Classifier-

K-NN algorithm checks the similarity of new data from the clusters of available data and put the new data into the category that most suits its similarity. The categorical classification of new data can be carried out using K- NN algorithm.

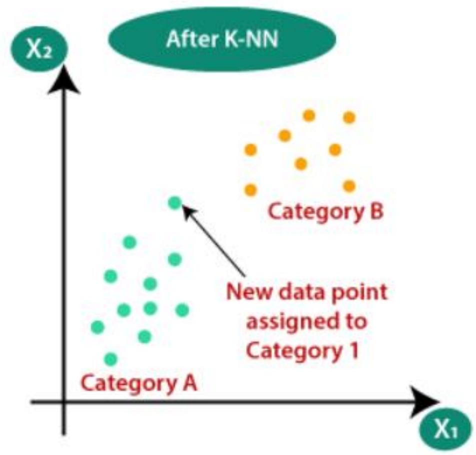

Fig 4:K nearest nieghbors Classifier

Let us consider the example in above figure with k=7. If by calculating the Euclidean distance we got the nearest neighbors, as four nearest neighbors in category A and three nearest neighbors in category B, the category for the new data point will be A.

Step 4 – Checking for Outlier

Before validation on original data checking for outliers was done, along with that, verification to make sure that the two lakh data points did not include any of the original 174 data points was also done using the pandas library in python. For outlier identification, box plots for all 6 features were done and number of values outside the interquartile ranges was calculated.

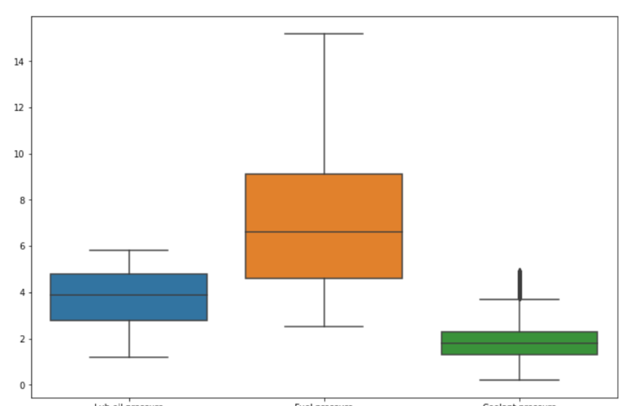

Fig 5:Box plots for predicting features

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 02 | February - 2023 Impact Factor: 7.185 ISSN: 2582-3930

The amount of outliers for two lakh data points was found to be around ten thousand points. This is an acceptable number. Also, considering the fact that the model can learn better with more variety of data, these points were not excluded.

Step 5- Validation

To verify correlation, heat map of the two lakh data points was plotted.

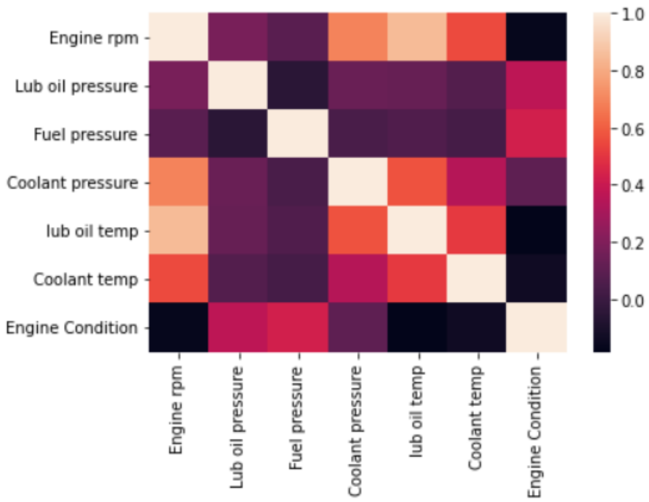

Fig 6: Heat map depicting correlation

After all of the above procedure, the 3 trained models were tested on the original data for validation. All the models gave an accuracy of around 90 percent, this provides validation for the synthetic data. For further validation, the data of 5000 points was generated and testing was for all 3 models on this new data, the accuracy was again observed to be around 90 percent. As a result, the validity of the synthetic data generated by Gaussian copula was verified

IV. RESULT AND DISCUSSION

Logistic Regression- After application of logistic regression on synthetic data, the results were as follows,

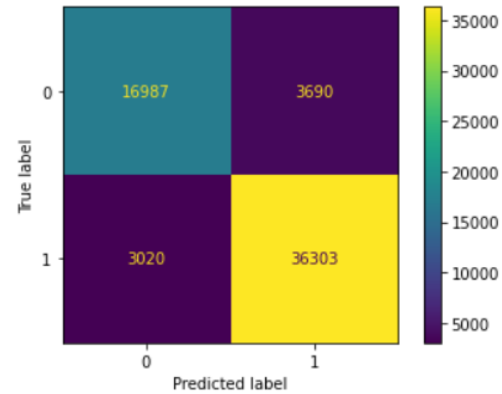

Fig 7: Logistic Regression confusion matrix

The classification report was as follows-

	precision	recall	f1-score	support
0	0.85	0.82	0.84	20677
1	0.91	0.92	0.92	39323
accuracy			0.89	60000
macro avg	0.88	0.87	0.88	60000
weighted avg	0.89	0.89	0.89	60000

Fig 8: Logistic Regression classification report

The overall accuracy on training data and testing data was found to be around 0.887 and 0.888 respectively.

Support Vector Classifier (Linear Kernel)- After application of support vector classifier with linear kernel on synthetic data, the results were as follows,

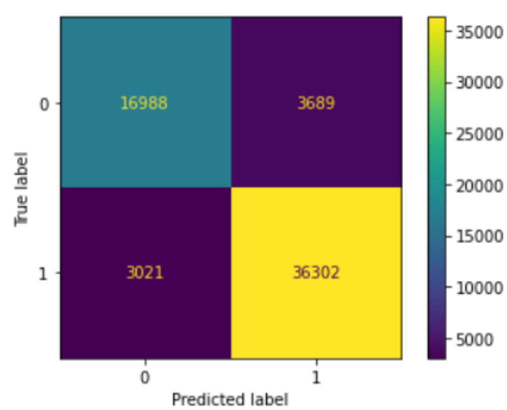

Fig 9: SVC (Linear) confusion matrix

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 02 | February - 2023 Impact Factor: 7.185 ISSN: 2582-3930

The classification report was as follows-

	precision	recall	f1-score	support
0	0.85	0.82	0.84	20677
1	0.91	0.92	0.92	39323
accuracy			0.89	60000
macro avg	0.88	0.87	0.88	60000
weighted avg	0.89	0.89	0.89	60000

Figure 10: SVC (Linear) classification report.

The overall accuracy on training data and testing data was found to be around 0.887 and 0.888 respectively.

Support Vector Classifier (Polynomial Kernel)- After application of support vector classifier with polynomial kernel on synthetic data, the results were as follows,

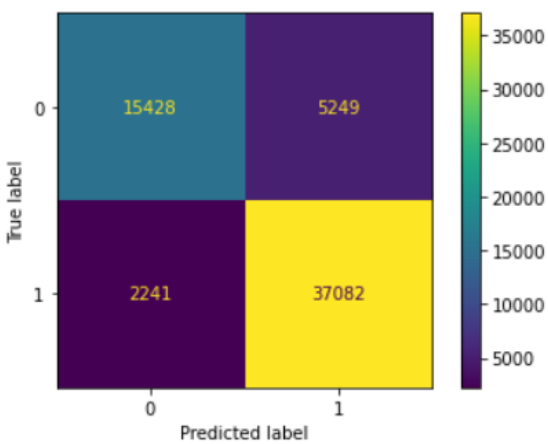

Fig 11: SVC (Polynomial) confusion matrix

The classification report was as follows-

	precision	recall	f1-score	support
0	0.87	0.75	0.80	20677
1	0.88	0.94	0.91	39323
accuracy			0.88	60000
macro avg	0.87	0.84	0.86	60000
weighted avg	0.88	0.88	0.87	60000

Fig 12: SVC (Polynomial) classification report.

The overall accuracy on training data and testing data was found to be around 0.8757 and 0.8751 respectively.

KNN Classifier (Nearest Neighbors=15)- After application of k neighbors classifier with number of neighbors as 15 on synthetic data, the results were as follows,

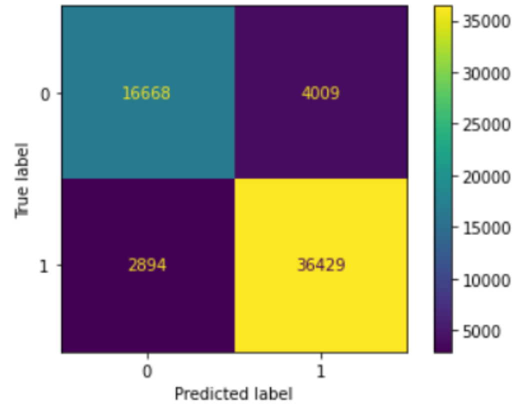

Fig 13: K nearest neighbor classifier confusion matrix

The classification report was as follows-

	precision	recall	f1-score	support
0	0.85	0.81	0.83	20677
1	0.90	0.93	0.91	39323
accuracy			0.88	60000
macro avg	0.88	0.87	0.87	60000
weighted avg	0.88	0.88	0.88	60000

Figure 14: K neighbors classifier classification report.

The overall accuracy on training data and testing data was found to be around 0.89 and 0.888 respectively.

V.CONCLUSION AND FUTURE WORK

The paper has shown the importance of using CBM technique as compared to TBM technique which impacts the maintenance team on cost and time. Further, the paper has used only one algorithm for model analysis, however the same cannot be said to be best or appropriate one view constraints in dataset. The parameters used were different from as used by the researchers in their paper. During generation of synthetic data, generative adversarial networks were also used for comparison. The data generated had correlation values with engine condition of the order 10-3 while the data generated with Gaussian copula has correlation values of the order 10-1. Furthermore, the correlation values were fairly close to correlation values of the original data. Logistic Regression and Support vector classifier had similar accuracy values and f1 scores, however when tested on original data, misclassifications were observed to be more in case of support vector classifier (linear kernel). Slight underfitting was observed for both these models. For Support vector classifiers with Polynomial Kernel, the accuracy values were the lowest. However, the problem of underfitting was not observed, this can be attributed to increased complexity of transformation. .For K neighbor's classifier, the accuracy values were observed to be highest for both training and testing on original data. This is due the fact that k neighbors are a non parametric algorithm, meaning it does not make assumptions on underlying data. It does not learn from the training set immediately, instead it stores the dataset and at the time of classification, it performs an action on the dataset. For all of the above models, the original 174 data points were completely unseen data. The fact that all models trained on synthetic data gave good accuracy and it proves that these models can be deployed in real conditions for predictive maintenance.