International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2019): 7.583

Predictive Reliability Modelling of an Industrial System

Arnab Naskar

SCSQR, Indian Institute of Technology, Kharagpur, India
arnab.naskar95[at]gmail.com

Abstract: ***Across the industries, there is a growing need of increased Operational Reliability, Availability and Maintainability of the equipment, which comprises of diagnosis and prognosis of a particular problem. As systems are evolving daily to a new level of sophistication, maintenance of those systems needs a critical approach. Hence, more industries are trying to adapt predictive maintenance policies in their critical area of operations. With the advantage of predictive analytics and machine learning techniques, predictive maintenance is gaining its momentum in different industries. In this paper, a framework of predictive reliability modelling has been discussed, which is a part of predictive maintenance. With the help of various machine-learning models along with Deep Neural Network, one predictive reliability model has been made. This paper also projects a model to calculate Remaining useful life (RUL) by incorporating Recurrent Neural Network (RNN).***

Keywords: Industry 4.0, Predictive Analytics, Machine Learning Techniques, Diagnosis, Prognosis, Predictive Maintenance, Predictive Modelling, Deep Learning, Reliability, Maintainability, Availability, Recurrent Neural Network

1. Introduction

Manufactures across different industries carry out different preventive maintenance procedures to keep the failure cost of any system to a minimum. This processes require a routinely check and repair of a particular system. With the advancement of technology, as the systems is becoming complex, maintenance of those systems need a new approach [1]. With the conventional maintenance policies, it is becoming difficult to predict the time to failure of many such systems. In many cases, the calculations are also coming as erroneous, which in turn possess detrimental effects. By leveraging the power of predictive analytics and big data, predictive maintenance policy is bridging this gap. As there is always a growing need to increase operational Reliability and Availability of systems, predictive reliability modelling is turning out to be a game-changer in industries. Predictive modelling provides a solution to assess the present and future health (diagnosis and prognosis) of a system, from where a maintenance action can be planned. After the collection of relevant data from the systems, Exploratory Data Analysis (EDA) has been done. Then those data is feed intostatistical models, which gives the desired result and a validation.

This paper gives a framework how to leverage predictive analytics to build a model for fault isolation and to prevent failure events of any system. Due to the confidentiality reason, all the failure modes and the real name of the parameters have been masked.

2. Methodology

The proposed methodology in performing the analysis is as shown in Figure 4. The important steps, are summarized below.

Step 1. Failure Mode Analysis
Systems fail in different ways. For an example, an automobile failure can be a reason of engine malfunction, transmission system or brake system failure. Further analysis can also tell that failure of one or more components leads to those systems failure. Hence, for any system more than one failure modes can be present [2].

Individual failure mode analysis not only helps us to quantify the impact of that failure mode on system reliability, but also to evaluate how to prevent it from happening. In this paper, a case study has been presented where three failure modes are present in a system failure.

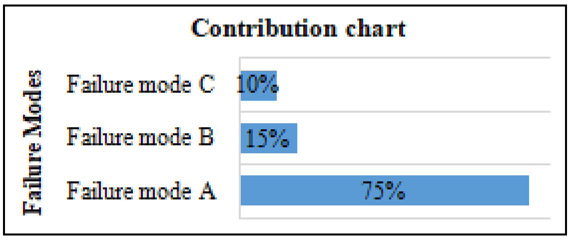

Figure 1: Contribution chart of failure modes

From Figure 1. it can be inferred that, failure mode A consists of almost 75% of system failure. Therefore, analysis of failure mode A is of utmost important.

Step 2. Fault Tree Analysis
Fault tree analysis is a technique to represent a model that graphically and logically represents of various combinations of possible events, which can lead to occurring of a top undesired event within a system [3].

In this case study, a similar fault tree analysis has been done to identify the possible events that can lead to failure mode A. From the analysis, 57 parameters are identified which can lead to failure mode A, which eventually leads to system failure.

Step 3. Failure Data Collection
In order to find out the exact reason of failure mode A, process data for 96 samples has been given. Out of these 96 samples, 16 samples are considered as failure. To prevent

the system from failure mode A, a predictive model needs to be made.

Step 4. Data Class Sorting

In this case study, a binary classification type dataset is given. The failures in the dataset are labeled as Class 1 and the other samples are labeled as Class 0. By observing it can be shown, the dataset distribution of the failure mode A dataset is imbalanced. The percentage of failure data is 19.15% and percentage of good sample data is 80.85%. A quick visualization of the dataset has been shown in Figure 2.

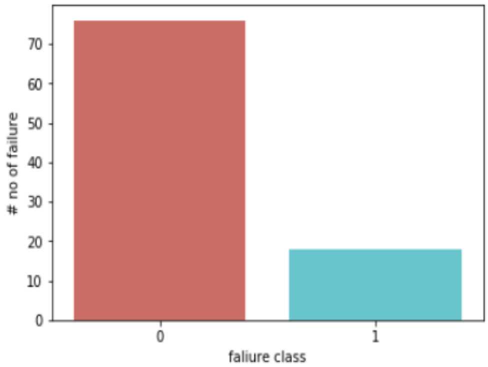

Figure 2: Imbalance Dataset

Step 5. Synthetic Minority Over-sampling Technique (SMOTE)

By leveraging the advantage of various machine learning models and Deep Learning, Predictive Reliability Modelling has been made. The performance of machine learning algorithms is typically evaluated using predictive accuracy. However, this is not appropriate, when the dataset is imbalanced. As a solution, SMOTE algorithm can be applied. Synthetic Minority Over-sampling Technique (SMOTE) is an over-sampling method in which for each minority class sample x, some samples are randomly selected from their k-nearest neighbors, and a new sample is constructed. In this way, new minority class can be created [4]. Therefore, applying SMOTE, the imbalance dataset can be balanced. Results is shown in Figure 3. after applying SMOTE.

```
length of oversampled data is  112
Number of no subscription in oversampled data 56
Number of subscription 56
Proportion of failure data in oversampled data is  0.5
Proportion of good sample data in oversampled data is  0.5
```

Figure 3: Balanced Dataset after SMOTE

Step 6. Data Preprocessing

In the data preprocessing steps, imputing missing data and feature scaling are one of the important steps to follow. In this analysis, by the help of "Missingno" library, missing data visualization has been made. Feature scaling has also been done. For many machine-learning models, standard scaling of the features is an important step which need to be followed. In Deep Learning also, it helps up in speeding up the calculation of the algorithm. The aim of this is to achieve Gaussian with zero mean and unit variance in order to avoid a particular feature gets dominance over the other features.

Step 7. Multicollinearity check

For the tree-based classifiers, multicollinearity does not possess any problem. As a tree-based classifier, Random Forest and Decision Tree have been used. For these two cases, there is no need to check for multicollinearity.

Variation Inflation Factor (VIF) is used to measure the multicollinearity for any dataset [5]. In the given problem, multicollinearity issue has not been seen.

Step 8. Applied Machine Learning Models

In this analysis, three machine leaning models (Logistic Regression, Random Forest Classifier, & Decision Tree Classifier) have been implemented in order to get the best-suited model for the dataset. Depending upon the accuracy parameter of each model, highest accuracy model has been chosen.

Usually the dataset has been divided into three parts. First part is training set, on which the model is trained. Second part is validation set, which is used to tune the hyperparameters of each model. The last part is test set, in which the trained model is tested. In this analysis, the dataset has been divided into two parts by merging the validation and test set. By splitting the dataset, 75% dataset has been put under "Training set" and the rest 25% as "Validation/Test set".

Step 9. Testing Evaluation and Performance Measure

The test data are put into machine-learning models for prediction, and the results are obtained. For performance estimation, precision, recall and F1 value performance measure are taken. Confusion matrix is used and shown in Table 1 [6].

Table 1: Confusion Matrix

True Label	*Predicted Label*	
	Failure	*Good*
Failure	True Positive (TP)	False Negative (FN)
Good	False Positive (FP)	True Negative (TN)

To calculate the *Precision*, *Recall* and *F1 score*, certain formulas are applied.

$$P(Precision) = \frac{TP}{TP + FP}$$

$$R(Recall) = \frac{TP}{TP + FN}$$

$$F1\ Score = \frac{2 * P * R}{P + R}$$

From F1 score, accuracy of the model can be derived.

Step 10. Best Accuracy Model

As previously discussed, three machine learning models have been applied to this dataset. Depending upon the accuracy of the models, the best fitted model has been chosen.

Step 11. Feature Selection

Feature selection helps to make sense of the features of a dataset and shows the importance of any feature [7]. Currently, there are three methods, which are useful in extracting features, which are:

1) Filter Method
2) Wrapper Method
3) Embedded Method

Under the Wrapper method, Recursive Feature Elimination (RFE) eliminates worst performing features from a particular model one after another until the best subset of features are known.

Under the Embedded method, tree based models comes, in which it calculates the best performing features for a specific dataset. This feature importance is calculated based on Gini Index, and Entropy.

In this analysis, these two methods have been used to extract most important features, which are responsible for failure.

Step 12. Remaining Useful Life (RUL) Estimation

For the prognostic measure, remaining useful life (RUL) calculation is an important step to consider. By taking into account of RUL, engineers can schedule maintenance plans to prevent unplanned failures.

For calculating RUL, there are three models available.
- Physics based Modelling
- Data Driven Modelling
- Hybrid Modelling

In our dataset, by using Data Driven Modelling approach, RUL has been estimated. For this part, Recurrent Neural Network (RNN) and Long Short Term Memory (LSTM) have been applied [8] [9].

A simulation study has been done up to 975 hours from the initiation of the machine. All the recorded probability of failure has been represented graphically, from which a forecasted model has been made.

By stacking four layers of LSTM, with a dropout of 0.2, a model has been made for forecasting the probability of failure for the next 100 hours. A threshold probability value is set, which will act as a condition indicator that will detect failure.

The proposed steps has been made in Figure 4 as a form of flow chart.

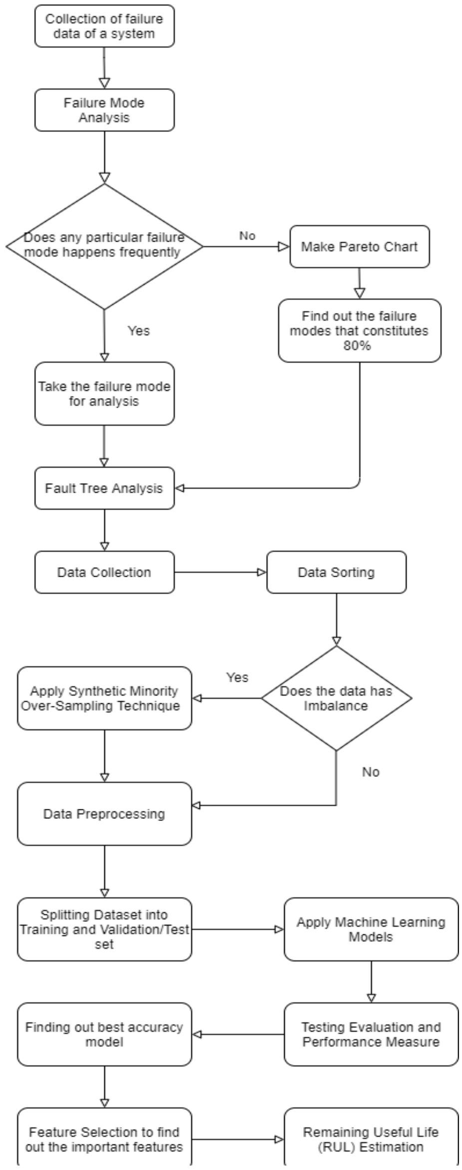

Figure 4: Predictive Model Flowchart

3. Case Study

In this present work, one industrial system has been analyzed. After failure mode analysis, it is found that failure mode A constitutes almost 75% cases of system failure. Hence, a framework of predictive reliability model has been made to prevent the system from failure mode A. As a diagnostic part, important parameters of this failure mode has been found. Remaining Useful Life (RUL) is also calculated as a part of prognostic part of the predictive reliability model.

From the fault tree analysis, 57 parameters have been identified as the cause of failure. However, in practical field, due to software lagging, unavailability of sensor mounting space or economic constraint, collection of real time sensor data of those 57 parameters may not be possible. Hence, to find the root causes and to reduce the number of parameters, machine learning techniques have been deployed.

4. Results

Aim of this case study is to project a framework to make a predictive reliability model by leveraging the predictive analytics techniques. Here several machine learning models has been applied to find out the best-accuracy model for the dataset.

Table 2. Benchmarking for fault classification

Algorithm Name	Performance Measure			
	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.82	0.85	0.82	0.82
Random Forest Classifier	0.97	0.97	0.97	0.97
Decision Tree Classifier	0.94	0.94	0.94	0.94

In this dataset, Random Forest Classifier is having the highest accuracy. By applying feature selection on random forest classifier, 10 significant parameters have been chosen [10]. The results have been shown in **Table 3.**

From the chosen 10 parameters, probability of failure can be calculated for each test set data. Similar steps can be applied By splitting the dataset into training/test set, RNN-LSTM has been applied. Train and Test loss has been taken into account. Figure 5 shows the model loss with the no of epochs.

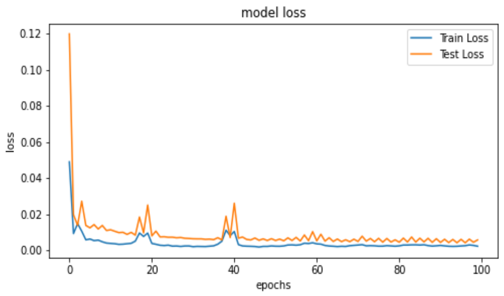

Figure 5: Model Loss

Finally, RUL has been plotted with the given set of data. It has been observed that at 1048th hour, the probability of failure will cross the threshold of 0.5, which will mark as the failure criteria of that particular machine.

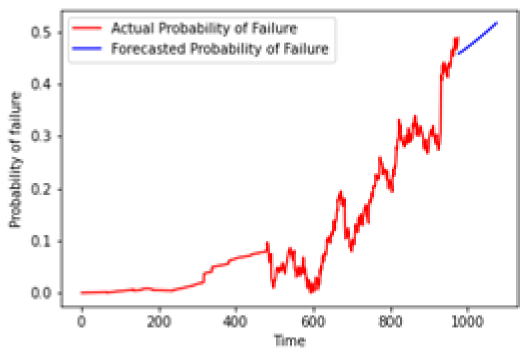

Figure 6: RUL Plot

Hence RUL is calculated as (1078 – 975) = 73 hour.

5. Conclusion

Initially from the dataset of 57 parameters, 10 significant parameters are identified as a part of diagnosis, which are responsible for Failure mode A on the system. In the prognosis part, by using RNN-LSTM method, an RUL plot has been made.

A simulation of 975 hours has been made for the RUL calculation. It has been seen that at 1078 hour the probability of failure is reaching 0.5. Hence, considering that hour as the threshold of failure, residual life has been calculated. Therefore, the residual life of the showerhead is 73 hours (the forecast is from current time instant of 975 hours).

References

[1] Andreja Rojko, "Industry 4.0 Concept: Background and Overview" iJIM- Vol. 11 No. 5, 2017

[2] Necip Doganaksoy, Gerald J.Hahn and William Q. Meeker,"Reliability Analysis By Failure Mode", Article- January 2003, ResearchGate.

[3] Dimitri B. Kececioglu, Reliability & Life Testing Handbook

[4] Tince Etlin Tallo, Aina Musdholifah, "The Implementation of Genetic Algorithm in SMOTE (Synthetic Minority Oversampling Tecnique) for Handling Imbalanced Dataset Problem",2018 4th International Conference on Science and Technology (ICST), Yogyakarta, Indonesia.

[5] Michael H. Graham, "Confronting Multicollinearity in Ecological Multiple Regression", Ecology, 84(11), 2003, pp. 2809-2815 © 2003 by the Ecological Society of America.

[6] L. Chen and H.L.Tang, "Improved Computation of Beliefs based on Confusion Matrix for Combining Multiple Classifiers", Electronics Letters, 19th February 2004, Vol. 40, No. 4.

[7] S. Visalakshi, V. Radha, "A Literature Review of Feature Selection Techniques and Applications", 978-1-4799-3975-6/14/$31.00 ©2014 IEEE.

[8] Jianjing Zhang, Kosta Ristovaski, Ahmed Farahat and Chetan Gupta, "Long Short Term Memory for machinelife prediction", In Journal of Manufacturing Systems 48, pp 78-86, 2018.

[9] Felix O. Heimes "Recurrent Neural Networks for Remaining useful Life Estimation", In IEEE International Conference on Prognostics and Health Management, 2008.

[10] Musa Peker, Ayse Arslan, Baha Sen, Faith V., Abdulkardir But, "A Novel Hybrid Method for Determining the Depth of Anesthesia Level: Combining ReliefF Feature Selection and Random Forest Algorithm (ReliefF + RF), 978-1-4673-7751-5/15/$31.00 ©2015 IEEE.

Author Profile

Arnab Naskar is a postgraduate student of Subir Chowdhury School of Quality and Reliability (SCSQR), IIT Kharagpur. He has recently received his M.Tech degree in Quality and Reliability Engineering. He holds a bachelor's degree in Mechanical Engineering from Jadavpur University, India. His research interests include reliability and predictive analytics domain.