Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Determining the Method of Predictive Maintenance for Aircraft Engine Using Machine Learning

Adryan Fitra Azyus[1]✉ and Sastra Kusuma Wijaya[2]
[12]Universitas Indonesia, Indonesia
Corresponding Author: Adryan Fitra Azyus, **E-mail**: *adryan.fitra01@ui.ac.id*

| ABSTRACT
Predictive maintenance (PdM) is indicated state of the machine to perform a schedule of maintenance based on historical data, integrity factors, statistical inference methods, and engineering approaches that are currently often applied to aircraft maintenance. The Predictive maintenance on aircraft to avoid the worse event (failure) and get information about the status of aircraft machines by applied on Machine Learning (ML) to get high accuracy and precision. The research aims to look for the method and technique of ML, which is the best applied on PdM for aircraft in accuracy indicators. The techniques of ML have been divided by classification and regression, which are compared on three ML methods: Random Forest (RF), Support Vector Machine (SVM), and simple LSTM. The result of the study for classification technique are LSTM 98,7%, SVM 95,6%, and RF 900,3%. On other hand, Regression technique for ML result on MAE and RMSE are LSTM 13,55 and 22,13, SVM 15,77 and 20,51, RF 15,06 and 19,98. Classify technique is better and faster than regression when calculating the PdM on an aircraft engine. The LSTM method of ML is the best applied to it because of the accuracy higher and time process faster than other methods in this study. Finally, the LSTM method is highly recommended while using with classify technique on ML to determine the PdM on an aircraft engine.

| KEYWORDS

Aircraft engine, Machine Learning Methods, Remaining Useful Life, Predictive Maintenance, Classification, Regression.

| ARTICLE DOI: 10.32996/jcsts.2022.4.1.1

1. Introduction

Machine learning (ML) is a method or algorithm capable of learning based on training that is given information about something. It can be used in the future when the algorithm is applied. There are three types of ML: Supervised learning make calculation or prediction based on known or labelled data, unsupervised learning focus on clustering data while process, and reinforcement learning focus on interaction to environment. Deep learning (DL) is the development algorithm from ML to fix its limitation. Both ML and DL algorithms are suitable for prediction, classification, and making cluster processes. Random Forest (RF) and SVM are popular methods in ML, then CNN, RNN, and LSTM are the most frequent implemented methods in DL. As a prediction tool, both ML and DL are divided into two techniques: Regression and Classify technique. The difference between the two techniques is the method's output; while regression technique is leaning on prediction value of a variable including time-series data, labelling of variable (discrete) is the focus for classify technique.

Evaluation is important to indicate the performance of the model that has been trained, so to validate the result of using ML and DL method, *Root Mean Squared Error* (RMSE) is a frequent evaluation method for regression technique, this evaluation indicating how concentrated the output is around the line of best fit on prediction data. As the ML and DL are capable of calculating prediction, this benefit can be applied to the various sector, such as prediction on machinery maintenance or services. The evaluation method for the classify technique are *Accuracy* defined parameter which indicates output values of prediction compared with the actual value, *Precision* describes the level of accuracy between the output of prediction compared with all data, and *Recall* is defined as the level of correct result compared by the number of results should have been returned, which mean the higher recall value, the higher data collected, and there is no data will miss from accuracy and precision process (Wuest et al., 2016).

The capacity of machinery working cannot last forever; sometimes, it will be broken down because of out-date operations. Machinery systems that include sensors are just monitoring the machine's state but cannot make a report of the machine is in good or bad condition. To avoid the worse event (failure) and get information about the status of a machine, a maintenance strategy must apply to the scheduled machinery system. There are three best practices of maintenance strategy; *Corrective, Preventive,* and *Predictive Maintenance*. Predictive maintenance (PdM) is based on preventive maintenance but continuous monitoring of the machine's state, the maintenance performed when it is needed or in an optimal way. PdM indicated the state of the machine to perform schedule to maintenance based on historical data, integrity factors, statistical inference methods, and engineering approaches (Susto et al., 2012)

Prognostic and health management (PHM) on the aviation industry will be expanded because of its effect on economic and human safety. Advance maintenance shall be applied to this industry to inform the aircraft engine condition. PdM is an advanced maintenance technique that can be applied to the aviation industry because of its high precision prediction, reducing cost operation and increasing safety by calculating the remaining useful time (RUL) of aircraft engines (Si et al., 2011). Furthermore, the Combination of PdM as a technique to calculate the RUL and ML as a tool to make a prediction in high accuracy is mx together that accurately forecasting the state of aircraft machine condition and on the best time to get the maintenance or service.

This article aims to determine the best method and technique of Machine Learning applied to Predictive Maintenance to calculate the RUL on aircraft machines. The rest of this article is arranged as follows. Section 2 explain the literature that according to the study about Predictive Maintenance and the applied-on aviation industry. The next section, Section 3, discuss the methodology proposed in this study. Results and Discussion are presented in section 4, and finally, in section 5 is presented the conclusion.

2. Literature Review

Schedule-based airline maintenance does not consider the prediction of faults in advance, resulting in undue maintenance situations where components are still kept in service even though they might have exceeded their defined limit of failure in the aviation industry. So, the unprecedented growth of the industrial sector has led to an exponential increase in the amount of industrial IoT (Internet of Things) data, and that is sprouting increased interest in the data-driven. Accurate data makes sensible decision-making regarding aviation safety and rational maintenance. A predictive maintenance model was proposed to identify the most crucial attributes and the critical relationship among the attributes for fault detection of individual equipment(s) in the turbofan aircraft, which uses the data-driven prognostic method for RUL estimation with multiple operating conditions. Predictive maintenance is a prominent strategy that can achieve increased reliability and safety of CPS (cyber-physical systems) while attaining reduced maintenance cost by estimating the current health status and the remaining useful life (RUL) (Behera et al., 2019)

Predictive maintenance is the method of scheduling maintenance based on the prediction about the failure time of any equipment. The prediction can be made by analyzing the data measurements from the equipment using Machine learning, a technology by which the outcomes can be predicted based on a model prepared by training it on past input data and its output behaviour. The model developed can be used to predict machine failure before it actually happens. Comparing 10 different Machine learning methods to calculate RUL (remaining useful lifecycle) for predictive maintenance of turbofan engine; as a result, RF is lowest RMSE than other 9 methods. The machine learning models were constructed based on the datasets from turbo fan engine data from the Prognostics Data Repository of NASA. Using a training set, a model was constructed and was verified with a test data set. The results obtained were compared with the actual results to calculate the accuracy and the algorithm (Mathew et al., 2020).

Machine learning (ML) application for predictive maintenance of a water-cooling system in a hydropower plant located in Vientiane province, Lao PDR. Data used for the learning algorithm is from log sheets 31 months, compiled by a temperature in/out heat exchanger unit and maintenance history. The data is separated into two sets: training and testing sets. This paper uses the Classification Learner Application to train the model. The application supports 22 classifier types, which can be organized into six major classification algorithms, including Decision Trees, Discriminant Analysis, Support Vector Machines (SVM), Logistic Regression, k-Nearest Neighbors (KNN), and Ensemble Classification. The best classification algorithms are the SVM and Decision Trees algorithms. The 5-fold-cross-validation option is used to evaluate all classifiers' mean error rates during training and testing the model. The prediction model has good efficiency while also predicting to good accuracy. The prediction result shows acceptance criteria. This model is more convenient for the operator; they can further visualize and monitor this system in the hydropower plant (Xayyasith et al., 2019).

3. Methodology

This study aims to compare machine learning techniques, regression and classification, and determine which technique is suitable for predictive maintenance for an aircraft engine. The difference between the technique is on regression the result about the prediction of a quantity and the classification predicting a label. On the other hand, the similar objectives of both techniques are predictive modelling, which is in this study is predictive maintenance on an aircraft engine. Predictive modelling is the problem of developing a model using historical data to make a prediction on new data where we do not have the answer. Predictive modelling can be described as the mathematical problem of approximating a mapping function (f) from input variables (X) to output variables (y). This is called the problem of function approximation. The job of the modelling is to find the best mapping function we can give the time and resources available. The machine learning algorithm is used as an important role in comparing both techniques: Random Forest, Support Vector Machine, and LSTM.

3.1 Classification Predictive Modeling

Classification predictive modelling is the task of approximating a mapping function (f) from input variables (X) to discrete output variables (y). The output variables are often called labels or categories. The mapping function predicts the class or category for a given observation; it is common for classification models to predict a continuous value as the probability of a given an example belonging to each output class. The probabilities can be interpreted as the likelihood or confidence of a given an example belonging to each class. A predicted probability can be converted into a class value by selecting the class label that has the highest probability. There are many ways to estimate the skill of a predictive classification model, but perhaps the most common is to calculate the classification accuracy. The classification accuracy is the percentage of correctly classified examples out of all predictions made.

3.2 Regression Predictive Modeling

Regression predictive modelling is the task of approximating a mapping function (f) from input variables (X) to a continuous output variable (y). A continuous output variable is a real value, such as an integer or floating-point value. These are often quantities, such as amounts and sizes. Because a predictive regression model predicts a quantity, the model's skill must be reported as an error in those predictions. There are many ways to estimate the skill of a predictive regression model, but perhaps the most common is to calculate the root mean squared error, abbreviated by the acronym RMSE. A benefit of RMSE is that the units of the error score are in the same units as the predicted value. An algorithm that is capable of learning a predictive regression model is called a regression algorithm. Some algorithms have the word "regression" in their names, such as linear regression and logistic regression, which can make things confusing because linear regression is a regression algorithm, whereas logistic regression is a classification algorithm.

3.3 Classification vs Regression

Classification predictive modelling problems are different from regression predictive modelling problems, Classification is the task of predicting a discrete class label, and Regression is the task of predicting a continuous quantity. There is some overlap between the algorithms for classification and regression; for example, A classification algorithm may predict a continuous value, but the continuous value is in the form of a probability for a class label, but a regression algorithm may predict a discrete value, but the discrete value in the form of an integer quantity. Some algorithms can be used for both classification and regression with small modifications, such as decision trees and artificial neural networks. Some algorithms cannot or cannot easily be used for both problem types, such as linear regression for predictive regression modelling and logistic regression for classification predictive modelling. Importantly, the way that we evaluate classification and regression predictions vary and do not overlap.

3.4 Dataset

The dataset is taken from NASA (National Aeronautics and Space Administration) data repository (A. Saxena, 2008). The dataset selected includes the run-to-failure sensor measurements from degrading turbofan engines. Each engine has 21 sensors collecting different measurements related to the engine state at runtime. These sensor devices fed the information periodically, then the system in this study for both machine learning techniques learned the behaviour of these data and provided estimation regarding maintenance conditions in the future or predictive maintenance.

3.5 Machine Learning Algorithm

There are 3 Machine learning methods used in this study that can run on both techniques, Regression and Classification, which are Random Forest, Support Vector Machine, and LSTM.

Support Vector Machine (SVM): In SVM, the numeric input variables in the data, which are in the different columns, form an n-dimensional space. A hyperplane is a line that splits the input variable space. In SVM, a hyperplane is selected to best separate the points in the input variable space by their class. The SVM algorithm is implemented in practice using a kernel.

Random Forest is an ensemble learning method; it operates by constructing a set of decision trees at training time and outputting the mean prediction of the individual trees. It generates n different subsets of training data sets using the bootstrap sampling approach from an original dataset. Subsequently, n different decision trees are constructed using training datasets of the subsets.

LSTM or Long-Short Term is one of the methods used in this study. LSTM is a development from the RNN method, which has its own advantage, such as there is architecture to remember and forget the output that is will be as input. On the other hand, the other ability of LSTM can maintain errors that occur when doing backpropagation, so it does not allow errors to increase.

4. Results and Discussion

The study result from both technique classification and regression algorithm, which are Random Forest (RF), Support Vector Machine (SVM), Long-Short Term method (LSTM). This study calculated the classification and regression data by implementing Python using CPU AMD Ryzen 5 4500 with 8Gb RAM. The processing time for execution is about 20 minutes, and the output is in the form of graphical comparison and evaluation for each method and technique. The secondary data from NASA have been used for calculated PdM. Predictive modelling result using classification shows on table 1, and among 3 methods, LSTM method obtain 98,7% accuracy, 92,3% precision, and 96% of recall which highest to other method with SVM for evaluation value 95,6%, 95,6%, and 88%, and RF 90,3%, 94,4%, and 68%. As a result of using classification, the technique decided LSTM is the best for predictive modelling. Moreover, table 2 shows the result of the regression technique with evaluation value using MAE, R2, and RMSE. The RF method obtain the evaluation value of MAE 0.762, R2 0.762 dan RMSE 19,98, which comparing with LSTM evaluation value sequentially 13.65, 0.74, and 20.91, then the SVM for the evaluation value 15.77, 0.755, and 20.512. The RF method is the best when implemented regression technique for Predictive Maintenance.

Table 1 Classification Evaluation Result

Metode	**Accuracy**	**precession**	**recall**
LSTM	0.987742	0.923077	0.96
Random Forest	0.903226	0.944444	0.68
SVM	0.956989	0.956522	0.88

Table 2 Regression Evaluation Results

Metode	**Mean Absolute Error**	**R2**	**Root Mean Square Error**
LSTM	13.653602	0.740260	20.910238
Random Forest	0.762622	0.762622	19.989856
SVM	15.770433	0.750044	20.512645

The result of each technique shows that LSTM is best implemented for classification and RF best for regression. The classification technique results in figure 1 that shows the comparison for all methods indicated LSTM value approaching the real value of dataset, then figure 2 also shows the comparison of all methods using regression technique with the real value, and the regression technique is not inaccurate as a classification because of regression using continue data, and it's hard to obtain the accurate value of R2, and in this study, the highest R2 of the evaluation value is 0,762, and in the classification technique, the highest accuracy is 98% which mean 98% of predicted data is matched with a real dataset. The predictive model using classification made the researcher easier to predict because the data is in continuous form. It has been changing on labelled form, so the predictive process is more accurate and quick than the regression technique. All the above factors determine a recommendation for using the classification technique to model predictive maintenance for aircraft engines.

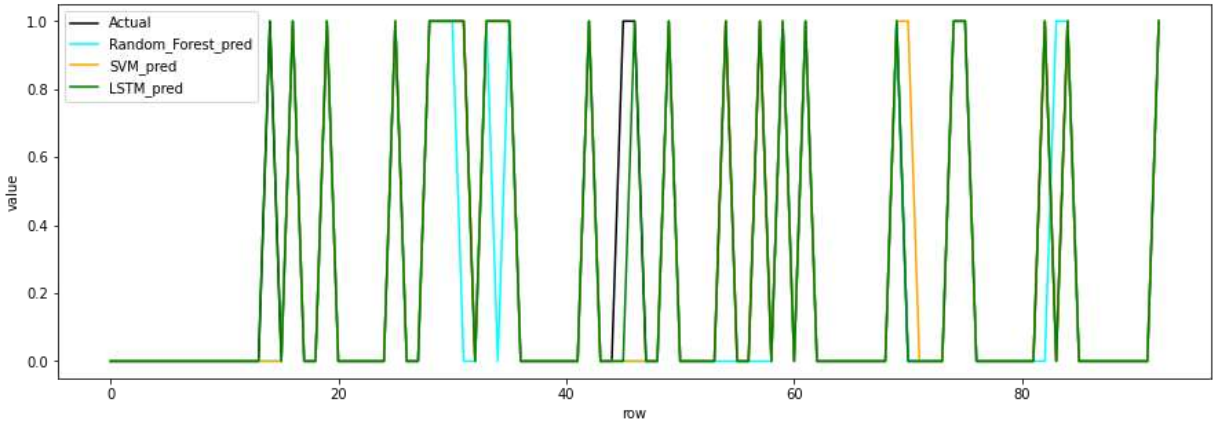

Figure 1 Comparison between Predictive Modelling Method on Classification Technique

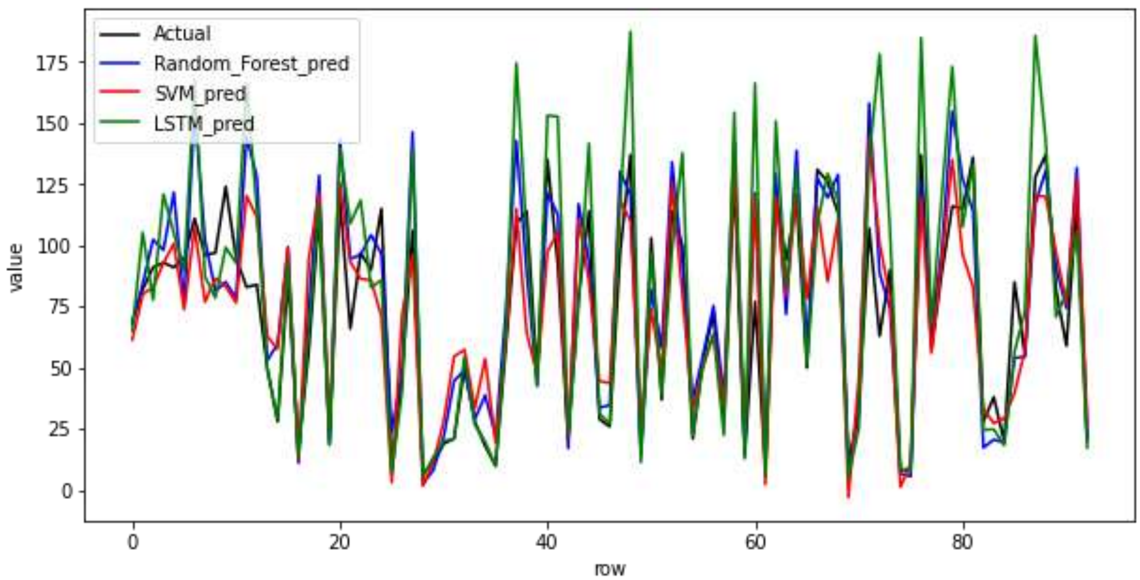

Figure 2 Comparison between Predictive Modelling Method on Regression Technique

5. Conclusion

Predictive Maintenance for aircraft engines has been done by the predictive model of machine learning and also can be implemented using classification and regression techniques. Both techniques, classification and regression, in this study applied on 3 machine learning methods which are Random Forest (RF), Support Vector Machine (SVM), and Long-short Term Method (LSTM). The result of this research for both techniques is the solution for Predictive Maintenance. As a detailed discussion, the classification techniques with the LSTM method is the best practice from other methods, and on regression technique, the Random Forest method is optimal on the result. Based on the aims of this study to compare both techniques, the result of the classification method is better than regression because of its easier and the accuracy of the predictive model, so the classification with LSTM method recommended implemented on Predictive Maintenance on an Aircraft engine. Therefore, for other research that implemented the predictive maintenance is dependent on the objectives of the research itself, if the result consistently continues, the regression technique is better than classification, but if the higher value on accuracy and the easier the calculating process, the classification will be as recommendation technique.

References

[1] Behera, S., Patel, Y. S., Choubey, A., Misra, R., Kanani, C. S., & Sillitti, A. (2019). Ensemble trees learning-based improved predictive maintenance using IIoT for turbofan engines. *Proceedings of the ACM Symposium on Applied Computing*, *Part F147772*, 842–850. https://doi.org/10.1145/3297280.3297363

[2] Mathew, V., Toby, T., Singh, V., Rao, B. M., & Kumar, M. G. (2020). 2020 IEEE 2nd International Conference on Circuits and Systems, ICCS 2020. *2020 IEEE 2nd International Conference on Circuits and Systems, ICCS 2020*, *Iccs*, 306–311.

[3] Si, X. S., Wang, W., Hu, C. H., & Zhou, D. H. (2011). Remaining useful life estimation - A review on the statistical data-driven approaches. *European Journal of Operational Research*, *213*(1), 1–14. https://doi.org/10.1016/j.ejor.2010.11.018

[4] Saxena, A., & Goebel, K. (2008). Turbofan engine degradation simulation data set. *NASA Ames Prognostics Data Repository*, 878-887.

[5] Susto, G. A., Beghi, A., & De Luca, C. (2012). A predictive maintenance system for epitaxy processes based on filtering and prediction techniques. *IEEE Transactions on Semiconductor Manufacturing*, *25*(4), 638–649. https://doi.org/10.1109/TSM.2012.2209131

[6] Wuest, T., Weimer, D., Irgens, C., & Thoben, K. D. (2016). Machine learning in manufacturing: advantages, challenges, and applications. *Production and Manufacturing Research*, *4*(1), 23–45. https://doi.org/10.1080/21693277.2016.1192517

[7] Xayyasith, S., Promwungkwa, A., & Ngamsanroaj, K. (2019). Application of Machine Learning for Predictive Maintenance Cooling System in Nam Ngum-1 Hydropower Plant. *International Conference on ICT and Knowledge Engineering*, *2018-November*, 43–47. https://doi.org/10.1109/ICTKE.2018.8612435