Maintenance Prediction by Machine learning: Study Review of some Supervised Learning Algorithms

Mouna Tarik and Khalid Jebari
Laboratory of Mathematics and Applications, University Abdelmalek Essaadi (UAE)
Faculty of Sciences and Techniques.
Tangier, Morocco
mouna.tarik@etu.uae.ac.ma, k.jebari@uae.ac.ma

Abstract

Predictive maintenance is a prominent strategy to minimize downtime, associated costs and failure risks. In this paper, a study review of some supervised learning algorithms is presented. Multi-layer perceptron (MLP), Support vector machine (SVM) and decision tree (DT) are compared in terms of prediction accuracy. The data considered for simulation is often used in literature; it is applied to aircraft engine sensors measurements to predict in-service engine failure. The performance of the algorithms, cited above, has been compared in terms of classification accuracy, precision, recall and F-score.
In this study, the support vector machine provides better results than other techniques for maintenance prediction.

Keywords
Predictive maintenance, Artificial neural network, Support vector machine, Decision tree, Accuracy.

1. Introduction:

Maintenance costs are a major part of the total operating costs of all manufacturing plants (R.Keith 2002). Every year, it is estimated that U.S. industry trade spends two hundred billion dollars on maintenance of plant equipment and facilities.

The ineffective maintenance results in a loss of more than sixty billion dollars (Mobley et al 2002). In fact, it is fully required for industries to detect faults early and accurately.

In industrial plants, two maintenance management systems are frequently used, which are run to failure and preventive maintenance. Run to failure means that the maintenance interventions are performed only after the occurrence of failures. This is clearly the simplest approach to deal with maintenance (Gian et al 2014). The major loss associated with it is high overtime, labor costs, high machine downtime and lower production availability (R.Keith 2002). Preventive maintenance takes place where maintenance actions are performed according to a planned schedule based on process iterations. Therefore, failures are usually prevented, but unnecessary corrective actions are often performed, leading to inefficient use of resources and increased operating costs (R. Keith 2002).

The predictive maintenance allows the monitoring of equipment to avoid future failures and detect abnormalities, identify the root cause of issues and schedule maintenance when it is needed. It is performed through historical data and decision making tools

In this paper, machine learning (ML) for maintenance prediction of aircraft engines are used.

Referring to literature, ML is the most appropriate to deal with prediction issues. The aim of the study is to analyze and compare some of the most supervised learning algorithms used for classification issues, so the most accurate one can be identified for prediction purpose.

This paper is organized as follows: Section 2 provides a brief literature review to introduce the ML algorithms used. Section 3, introduces an overview of maintenance requests and methodology. Section 4 presents the experimental results. Finally, section 5 concludes the obtained result.

2. Literature review:

2.1 Machine learning:

Machine learning is a sub-discipline of artificial intelligence used in a variety of application domains for classification, pattern recognition, clustering, etc. The objective is to recognize systematic classification of patterns

and enables computers in building models from a data based on features extraction (Sekeroglu1 et al 2019) in order to facilitate decision making. In literature, as shown in Table.1, mainly four different types are used to perform prediction: supervised learning, semi-supervised learning, unsupervised learning and reinforcement learning**.**
Depending on the type of output, two classes of supervised problems are possible: classification (if the output concludes categorical values) or regression (if the output concludes continuous values). We will focus on the classification field in the context of selecting the best supervised learning algorithm.

Table1: Types of learning

Learning type	Description
Supervised learning	Training data includes the desired output
Unsupervised learning	Training data does not include the desired output
Semi-supervised learning	Training data includes few desired outputs
Reinforcement learning	Rewards from the sequence of actions

2.2 A review of maintenance prediction studies:

Various investigative studies have been conducted on downtime, fault prognosis and maintenance prediction. Table 2 presents a recapitulation of the studies in the maintenance fields;

Table2: Recapitulation of classification techniques applied to predictive maintenance

Authors	Year	Method	Application	References
Hack-Eun Kim et al.	2012	SVM	Prognosis of bearing faults	(Hack-Eun et al 2012)
P. Konar& Chattopadhyay	2011	SVM	Bearing fault detection in induction motor	(Konar et al 2011)
Bo-Suk Yang et al.	2008	Random forest	Diagnosis of induction motors	(Bo-Suk et al 2008)
V.Muralidharan et al.	2013	Bayes Algorithm	Diagnosis of mono-block centrifugal pump	(Muralidharan et al 2012)
M. Unala et al.	2014	GA-ANN	Defects identification for rolling bearings	(Unala et al. 2014)
Hongfei Li et al.	2014	DT , SVM	Improving rail network velocity	(Hongfei et al. 2014)
Piero Baraldi et al.	2016	KNN	fault diagnostics of automotive bearings operating under variable conditions	(Baraldi et al. 2016)
Wathiq Abed et al.	2014	Dynamic Neural Network	Diagnosis of bearing fault of brushless DC motor	(Abed et al. 2014)
Akhand Rai et al.	2017	FCM	Degradation assessment of bearing	(Akhand et al 2017)
Zaharah Allah Bukhsha	2019	DT	Railway switches	(Bukhsha et al 2019)
Dazhong Wu et al.	2017	Random forest	Tool Wear prediction	(Dazhong et al. 2017)
Shuangshuang Jin et al.	2017	GA-SVM	Remaining Life Prediction of the Fan Bearing	(Jin et al 2017)
Achmad Widodo et al.	2007	SVM	machine condition monitoring and fault diagnosis	(Widodo et al 2007)

In summary, according to aforementioned fields of literature, it is found that most of machine learning algorithms are able to detect the faults information and deal with diagnosis issues depending on the data used and the features extraction.

2.3 Machine Learning algorithms:

Many Machine Learning techniques are designed to analyze large amounts of data and are capable to handle high dimension very well (Thorsten et al 2016). Thus, ML techniques have been considered to develop maintenance prediction models.

In this research, Artificial Neural Networks (ANN), Support Vector Machine (SVM), and Decision Trees (DT) are considered for the comparison in order to predict aircraft engine failure. For that, the present study identifies the model that best fits the relationship between the attribute sets and class labels of the input data.

a) Artificial neural network

Neural networks, as abstract mathematical models, are based on the functioning of the brain. ANNs are characterized by the network architecture; it's composed of many computing elements called nodes and organized in a feed forward manner. Each node receives an input signal from other nodes or external inputs and then after processing the signals through an activation function, it generates a transformed signal to other nodes (Guoqiang et al 1999). The multi-layer perceptron (MLP) is the popular form of ANNs, which consists of three different layers:

The input layer: the external information is received in this layer. It transmits signal to the following layer;

The hidden layers: the number of neurons used in this layer depends on the problem that the MLP tries to learn;

The output layers: in this layer, the network provides the desired outcome.

The MLP with one hidden layer and two output layers is shown in Figure.1.

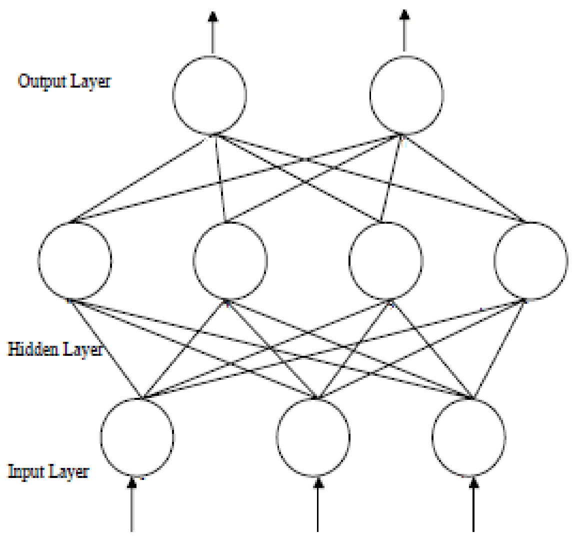

Figure.1: MLP with one hidden layer

The MLP works through the optimized weight values; the weights are estimated before that the training phase begins. The learning process allows reaching the optimized weight values.

Once the learning process is completed, the trained NN with optimal weights will be able to produce the output with the needed accuracy.

Let n is the total number of inputs of a neuron, $x_1 . . ,x_n$, the inputs, and the corresponding weights are w1,. . .,wn. The output of the adder is:

$$u= \sum_{i=1}^{n} w_i x_i \quad (1)$$

And

$$Y=f\,(u+b) \quad (2)$$

Where Y is the output of the neuron, b is the bias and f ( ) is the activation function.

In practice, a variety of activation functions are used, the most common ones are:

1: Sigmoid: computes $\frac{1}{1+e^{-x}}$

2: tanh computes tang(x)

3: Relu computes max (0, x)

b) Support Vector machine:

SVM was introduced by Vapnik, with the aim of minimizing an upper bound of the generalization error by maximizing the margin between the separating hyperplane and the data (Yongqiao et al 2005). Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Through these support vectors, we can maximize the margin of the classifier. The simplest formulation of SVM is the linear one.

We consider the training sample :{xi, yi }, i = 1, . . . , N with input data xi and output data yi with class labels {-1, +1}.

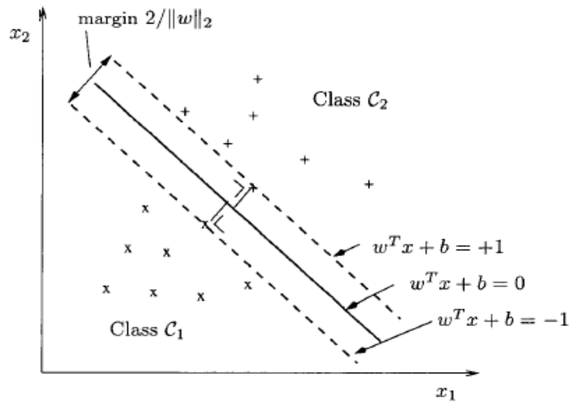

Figure.2: Definition of a separating hyperplane illustrated in a two-dimensional input space: Linear classification:

In practice, for classification fields; train an SVM means solve the following minimization problem:

$$Y[w^T xi+b]>1, \quad i=1,\ldots,N \tag{3}$$

Where wεRn is a weight vector and the bias b is a scalar.

The general methodology is to start formulating the problem in the primal weight space as a constrained optimization problem, next formulate the Lagrangian, then take the conditions for optimality and finally solve the problem in the dual space of Lagrange multipliers [10].

$$Y=\text{sign}\ (\textstyle\sum_{i=1}^{n} y_i\alpha_i(x.\mathrm{x_i}) + b \tag{4}$$

Where Y is the outcome, b and αi are parameters that determine the hyperplane.

For the non-linearly separable case, a high-dimensional version of Eq. (4) is given as follows:

$$Y=\text{sign}\ (\textstyle\sum_{i=1}^{n} y_i\alpha_i K(x,\mathrm{x_i}) + b) \tag{5}$$

Where K(x,xi) is defined as the kernel function. Namely polynomial and radial basis functions are the basic kernel functions for SVM.

The extension of linear SVMs to the non-separable case was made by Cortes & Vapnik in 1995, it is done basically by taking additional slack variables in the problem formulation. In order to tolerate misclassifications, the set of inequalities is changed to:

Minimize
$$\frac{1}{2}w^2+\mathrm{C}\textstyle\sum_{i=1}^{m}\xi^2 \tag{6}$$

Subject to:

$$y_i(\mathrm{w^T x_i}+\mathrm{b})\geq 1-\xi \quad , i=1,\ldots\ldots, M$$
$$\xi_i\geq 0, \quad i=1,\ldots\ldots\ldots, M \tag{7}$$

Where ξi is measuring the distance between the margin and the examples $\mathrm{x_i}$ that exist on the wrong side of the margin. In this study, the aim of the SVM in maintenance prediction is to maximize the margin between support vectors in order to get the best separation between failure and non-failure of engines.

c) Decision tree:

The concept of Decision Tree (DT) has been implemented using different algorithms e.g. ID3 (Quinlan, 1986), C4.5 (Quinlan, 1993), and CART (Classification and Regression Trees) (Breiman et al., 1984) (Bukhsha et al 2019). The goal is to divide the training set into homogeneous subsets. The tree has three nodes:

• Root node is the node that has no incoming edges and zero or more outgoing edges;

• Internal nodes have exactly one incoming edge and two or more outgoing edges;

• Leaf or terminal: each node has one incoming edge and no outgoing edges.

Decision Tree models are created using 2 steps: Induction and Pruning.

1. Induction is where the tree is built, set of all the hierarchical decision boundaries based on the data. The key to construct decision tree is how to choose a better logical attribute. Because of the nature of training decision trees, they can be prone to over- fitting.

2. Pruning is the process of removing the unnecessary structure from a decision tree, effectively reducing the complexity to avoid noise with the added possibility of making it even easier to interpret.

Decision tree algorithms use information gain to split a node. Gini index (Wenliang 2012) or entropy is the criterion for calculating information gain. Let S be a data set S that contains examples from n classes, the Entropy(S) and the Gini(S) are defined as follows:

$$\text{Entropy(S)} = -\sum_{i=1}^{n} P(cj)\ \ logP(cj) \tag{8}$$

$$\text{Gini(S)} = 1-\sum_{j=1}^{n} P(cj)^2 \tag{9}$$

Where P(cj) is the probability of class cj in a node .
Information gain is the entropy of parent node minus sum of weighted entropies of child nodes. It's computed as follows (If attribute A is used to partition the data set S):

$$\text{Gain(S, A)} = \text{Entropy(S)} - \sum_{v \epsilon A} \frac{|Sv|}{|S|} * Entropy(Sv) \tag{10}$$

$$\text{Gain(S, A)} = \text{Gini(S)} - \sum_{v \epsilon A} \frac{|Sv|}{|S|} * Gini(Sv) \tag{11}$$

Where v represents any possible values of attribute A; Sv is the subset of S for which attribute A has value v; |Sv| is the number of elements in Sv; |S| is the number of elements in S (Wenliang 2012).

3. Maintenance request and methodology:

3.1 Maintenance request:

Airlines are interested to predict engine failures in order to reduce the flight delays. Observing engine's health and condition through sensors historical data is assumed to facilitate predicting failures of in-service equipment in the near future. The goal is to predict if the engine will fail in the specific period or not. It means that the machine learning model will classify the label into two different classes (1 (healthy) or 0 (faulty)).

3.2 Data Description:

The data set is composed of the measurements of 21 sensors, simulations of run-to-failure events of an aircraft engine, together with operational settings. The sensors measurements inform about the progress of the degradation pattern of the engine. The training data file contains 20000 cycle records for 100 engines, while 100 records are contained in the testing data records. The training data consists of multiple multivariate time series with "cycle" as the time unit, together with 21 sensor readings for each cycle. Each time series can be assumed as being generated from a different engine of the same type. In this simulated data, the engine is assumed to be operating normally at the start of each time series. It starts to degrade at some point during the series of the operating cycles. The degradation progresses and grows in magnitude. When a predefined threshold is reached, the engine is considered unsafe for further operation. All features columns are numeric with no missing values. The table below shows the data with necessary details;

Table 3: Description of dataset

Datatype	Multivariate
Attribute types	Integer, Real
Number of instances	100
Number of attributes	26
Number of classes	2

3.3 Development Environment:

The code was performed using python3, Pandas and Numpy for data analysis, sickit-learn libraries for ML algorithms, Matplotlib for visualization.
The codes were executed on the core I5-73000 CPU processor with the operating system Linux Debian.

4. Experimental results:

This section presents the results of all models MLP, SVM and Decision tree for the engine's dataset.
The testing data is used to evaluate the model's performance. The metrics used for evaluation purposes are explained as follows:
First we calculate the accuracy. It shows how the model classifies the instances correctly. It's calculated below:

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{12}$$

Where TP and TN are True Positive and True Negative values, and FP and FN are respectively False Positive and False Negative values.
The Table below shows the results for the considered dataset:

Table 4: Accuracy results

Model	**Accuracy(testing dataset)**
MLP	0,82
Decision tree	0,88
SVM	0,91

Then we calculate the F-score (also called the F1 score or F measure), it is mainly used for binary classification problem and defined as the weighted harmonic mean of the test's precision and recall:

$$\text{F-score} = \frac{2*Precision*Recall}{Precision+Recall} \tag{13}$$

Where the precision determines the exactness of the model. It is a ratio of correctly predicted positive instances (TP) to the total positively predicted instances (TP+FP) (Bukhsha et al 2019):

$$\text{Precision} = \frac{TP}{TP+FP} \tag{14}$$

In contrast, recall provides a measure of model's completeness. It is a ratio of correctly predicted positive instance to the total instance of positive class (TP+FN) in test data (Bukhsha et al 2019). Recall is calculated as:

$$\text{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \tag{15}$$

Table 5 shows the calculation result for the models used:

Table 5: F-score result

Model	F-score
MLP	0,5
Decision tree	0,7
SVM	0,79

Table 6 shows the result of maintenance need prediction:

Table 6: Recapitulation of the results

Model	Precision	Recall	F-score	Accuracy
MLP	0,875	0,56	0,682	0,87
Decision tree	0,933	0,56	0,7	0,88
SVM	0,944	0,68	0,79	0,91

In order to visualize the performance of various models, we will use the AUC-ROC (Area Under the Receiver Operating Characteristics) curve. AUC for ROC and Precision-Recall curves were plotted as shown below:

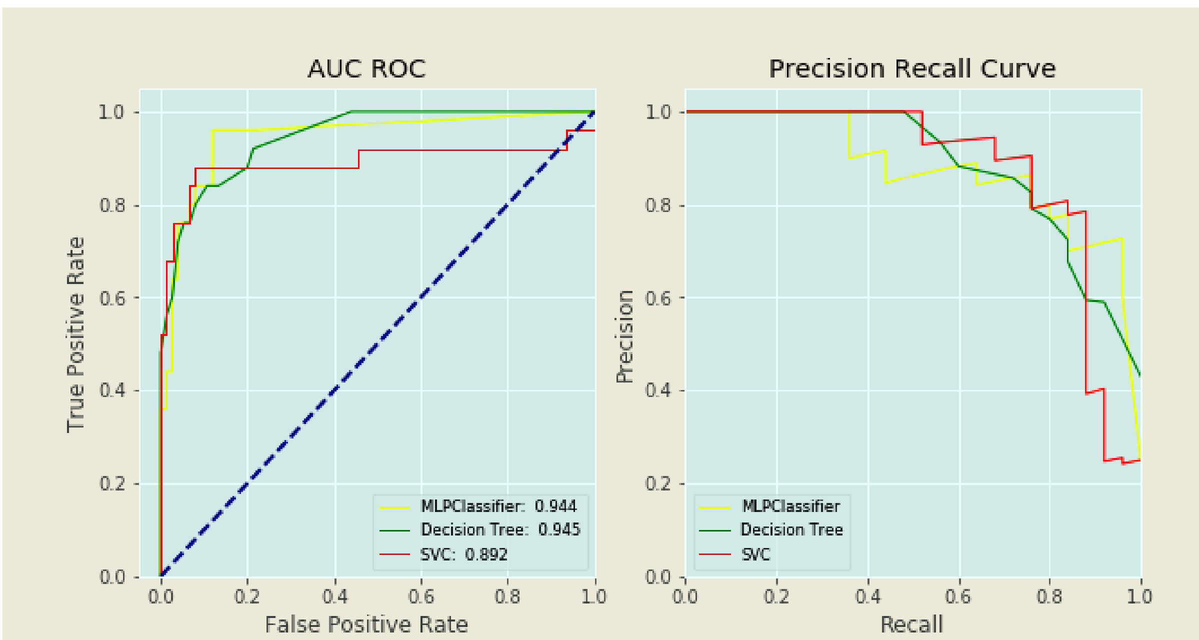

Figure3. AUC ROC and precision recall curves

A false positive rate of the ROC curve plots on an x-axis (The probability that true value is zero for the target is equal to one) against a true positive rate on a y-axis (The probability that the true value is one when the target is equal to one). The curve of Multi-layer perceptron, Decision tree and SVM has been showed.
SVM has the lowest AUC-ROC and scored better than other classifiers in precision-recall curve.

6. Conclusion:

This paper has presented a comparative study of different machine learning (ML) algorithms used for classification tasks in maintenance prediction. In the experiment, a typical use case of aircraft engine failure detection was studied. We have included tree ML predictive techniques; Multilayer perception (MLP), support vector machine (SVM) and decision tree (DT), to perform a binary classification study in order to predict if the engine will fail in a specific period or not. The best ML model was selected based on different metrics. In our case, the support vector machine is most likely to have a good performance for failure prediction; it scored better in Precision, Recall and Accuracy. The multi-layer perceptron and the decision tree have shown a poor performance although they scored better in AUC-ROC. Future work will include the implementation of more machine learning algorithms with more numerical datasets and will focus on unsupervised learning and new hybrid models.

References

R.Keith Mobley. *An introduction to predictive maintenance. Elsevier Science (USA),2002*.

Gian Antonio Susto, Machine Learning for Predictive Maintenance: a Multiple Classi- er Approach. IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS,11(3):812- 820,2014.

Tom M. Mitchell. *Machine learning.* McGraw-Hill Science/Engineering/Math, 1997.

Guoqiang Zhang a, Michael Y. Hu b, B. Eddy Patuwo b, Daniel C. Indro b. *Artificial neural networks in bankruptcy prediction. General framework and crossvalidation analysis*,116(1):16-32,1999

Robert E. Dorsey,Robert O. Edmister,John D. Johnson. *Bankruptcy Prediction Using Articial Neural Systems. Research Foundation of The Institute of Chartered Financial Analyst*,1995.

Boran Sekeroglu1, Shakar Sherwan Hasan2, and Saman Mirza Abdullah. *Comparison of Machine Learning Algorithms for Classification Problems.Science and Information Conference,*491-499,2019.

Z. Allah Bukhsha, A.Saeedb, I. Stipanovica, Andre G. Doreea. *Predictive maintenance using tree-based classification techniques: A case of railway switches. Transportation Research Part C: Emerging Technologies*,vol 101:35-54,2019

Yongqiao Wang, Shouyang Wang, and K. K. Lai. *A New Fuzzy Support Vector Machine to Evaluate Credit Risk*,13(6):820-831,2005.

Kyung-Shik Shin, Taik Soo Lee, Hyun-jung Kim. *An application of support vector machines in bankruptcy prediction model, Expert systems with applications,* 28(1):127- 135,2005.

Johan A. K. Suykens Jony Van Gestel. *Least Squares Support Vector Machines. Neural processing letters*,9(3):293-300,1999.

Nagdev Amruthnath , Tarun Gupta. *A Research Study on Unsupervised Machine Learning Algorithms for Early Fault Detection in Predictive Maintenance.5th International Conference on Industrial Engineering and Applications (ICIEA)*

Mobley, R Keith, *An Introduction to predictive maintenance*, 2002, 2nd ed, ISBN 0- 7506-7531-4,p:355-361,2018.

Thorsten Wuest,Daniel Weimer,Christopher Irgens Klaus-Dieter Thoben. *Machine learning in manufacturing: advantages, challenges, and applications.Production and Manufacturing Research,*4(1):23-45,2016.

Hack-Eun Kim Andy C.C.Tan JosephMathew Byeong-Keun Choi. *Bearing fault prognosis based on health state probability estimation.Expert Systems with Applications*,39(5):5200-5213,2012.

P. Konar P. Chattopadhyay. *Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVMs). Applied Soft Computing*, 11(6):4203-4211,2011.

Bo-Suk Yang, Xiao Di and Tian Han. *Random forests classifier for machine fault diagnosis, Journal of mechanical science and technology*,22(9):1716-1725,2008.

V. Muralidharan , V.Sugumaran. *A comparative study of Naive Bayes classifier and Bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis, Applied Soft Computing*,12(8):2023-2029,2012

Muhammet Unala Mustafa Onat Mustafa Demetgul HalukKucuk *Fault diagnosis of rolling bearings using a genetic algorithm optimized neural network. Measurement*,vol 58:187-196,2014.

Hongfei Li , Dhaivat Parikh , Qing He, Buyue Qian, Zhiguo Li , Dongping Fang , Arun Hampapur. Improving rail network velocity: *A machine learning approach to predictive maintenance. Transportation Research Part C: Emerging Technologies,* vol 45:17-27,2014.

Piero Baraldi, Francesco Cannarile, Francesco Di Mai , Enrico Zio. *Hierarchical k-nearest neighbours classification and binary differential evolution for fault diagnostics of automotive bearings operating under variable conditions. Engineering Applications of Artificial Intelligence*, vol 56:1-13,2016.

Wathiq Abed ,Sanjay Sharma , Robert Sutton. *Diagnosis of bearing fault of brushless DC motor based on dynamic neural network and orthogonal fuzzy neighborhood discriminant analysis.UKACC International Conference on Control (CONTROL) IEEE*, Pages:378-383, 2014.

Akhand Rai, S.H. Upadhyay. *Bearing performance degradation assessment based on a combination of empirical mode decomposition and k-medoids clustering.Mechanical Systems and Signal Processing*, vol 93: 16- 29, 2017.

Dazhong Wu, Connor Jennings, Janis Terpenny, Robert X. Gao. *A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests. Journal of Manufacturing Science and Engineering*, 139(7):71-81, 2017.

Shuangshuang Jin , Zhenhuan Wu , Yudong Fu and Guoan Yang. The Remaining Life Prediction of the Fan *Bearing Based on Genetic Algorithm and Multi-parameter Support Vector Machine. 5th International Conference on Mechanical, Automotive and Materials Engineering (CMAME) IEEE ,*pages:42-47,2017

Achmad Widodo, Bo-Suk Yang. *Support vector machine in machine condition monitoring and fault diagnosis. Mechanical systems and signal processing,21(6):*2560-2574,2007.

Wenliang Du, Zhijun Zhan : *Building Decision Tree Classifier on Private Data [27] Lidia Ceriani,Paolo Verme*. *The origins of the Gini index: extracts from VariabilitÃ e MutabilitÃ (1912) by Corrado Gini, The Journal of Economic Inequality, 10(3*): 421-443,2012.

Biography

Mouna Tarik is an industrial engineer with a proven experience of almost eight years in different positions in industrial plants. She holds a master degree in industrial engineering at The Faculty of science and techniques, Fès, Morocco. She is a PHD student at University Abdelamalek Essaidi, Tangier, Morocco. She has published one conference paper. Her research interests include machine learning, prediction, industrial engineering and manufacturing.