Examining Equipment Condition Monitoring for Predictive Maintenance, A case of typical Process Industry

Rafif Nova Riantama, Abiyyu Dimas Prasanto, Nani Kurniati*, Dewanti Anggrahini
Department of Industrial and System Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
riantama76@gmail.com, abiyyudimas5@gmail.com,
*nanikur@ie.its.ac.id, dewanti@ie.its.ac.id

Abstract

Maintenance of the production machine plays an important role to keep the production process properly run. Improper maintenance may cause problems in production. Corrective Maintenance strategy, a response to equipment failure, is needed to rectify the failures. However, this strategy may cause higher downtime and higher costs due to loss of production. Predictive Maintenance (refer to Condition-Based Maintenance) utilizing equipment monitoring data can be used for optimizing the maintenance strategy by predicting the future machine condition. This research attempts to examine the condition-based equipment data using the data analytics approach to developing a Predictive Maintenance program. Several methods are applied. K-means for clustering the failure characteristic, Support Vector Regression (SVR) model used for predicting equipment failure. The result and discussion represent that SVR and K-means model suitable for developing equipment failure prediction, useful support for managing the maintenance activities.

Keywords
Data analytics, Predictive maintenance, Condition Monitoring, Support Vector Regression, Cluster.

1. Introduction

Maintenance of the production machines is essential to ensure the production process properly run. The downtime due to failure or maintenance will result in stopping all the production processes. Improper maintenance may cause serious consequences for product quality, equipment availability, the environment, and company competitiveness (Ledger et al., 1999). Based on the observation of a typical manufacturing company in Indonesia, there is a low equipment availability due to high period downtime and large number of breakdowns. During 2015 – 2019, the machine's availability about 70% - 80%, which is far below the company's target of 90%. Further consequences, the production number also below the target.

Proper maintenance activities contribute to overall business performance through their impact on quality, efficiency, and effectiveness of the company's operations (Alsyouf, 2009). Well performed maintenance involves few corrective maintenance actions while performing as little preventive maintenance as possible (Restagari et al. 2013). Corrective maintenance may cause unplanned downtime, while scheduled or preventive maintenance may cause high maintenance costs. Condition-based maintenance (CBM) appears to improve the maintenance system to become more efficient. CBM uses the predictive principle to develop a maintenance plan (Susanto and Kurniati, 2020). This maintenance strategy makes fewer downtimes than scheduled maintenance by removing unnecessary maintenance, as well as reducing corrective maintenance by anticipating the possibility of equipment failure (Sampaio et al., 2019). Condition-based maintenance is included in advance maintenance using sensors to predict equipment failures by monitoring the equipment condition (Elsayed, 2013). As stated in Jardine (2006), CBM could create an effective maintenance system, so it can optimize the equipment's life and performance. Moreover, CBM can optimize the production process and increasing productivity (Fathy, 2017 and Prajapati et al., 2012).The decision-making steps on CBM can be divided into two steps, diagnostic and prognostic. Diagnostic deals with the identification when the failure occurs, while prognostic is predicting when the failures occur. Since large data needed to be proceeded for taking that decision, data analytic approaches should be applied to conducting the decision-making process.

Several studies that using data analytics on condition-based maintenance showed the machine learning model is applied to solve industrial cases by predictive maintenance (Krenek, et al., 2016). In research by Moura, et al. (2011), the Support Vector Regression (SVR) displayed great performance to predict machine failure on time series data. Research by Suyuti, et al. (2013) performing an Artificial Neural Network (ANN) model to classify the machine failure types based on the failure characteristic. Some machine learning models are applied e.g. Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest, Random Tree Model, to predict equipment failure (Sampaio, et al., 2019). This study reported the ANN performing best accuracy than the other models. All those previous research shows that data analytics can be implemented for the CBM strategy.

This research attempts to develop a condition-based maintenance system using data analytic approach for a typical process industry i.e. cement industry. In this study, two machines are examined e.g. raw mill and kiln, as a study case. Different data analytics methods will be applied to each machine. An SVR model will be implemented on a Rawmill machine for prognostic steps of machine failure prediction. To the Kiln machine, K-means clustering will be implemented to determine the number of clusters and analyze the characteristic of every cluster. The output of this research is expected to solve the maintenance problem of similar characteristics industry.

2. Literature Review

2.1. Condition Based Maintenance

According to Fu et al. (2004), the Condition Based Maintenance is a maintenance strategy focuses on the prediction of the degradation process of the product, which is based on the assumption that most abnormalities do not occur instantaneously. There are three steps in condition-based maintenance according to Jardine et al. (2006).

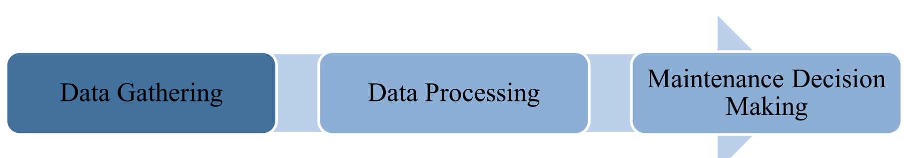

Figure 1 Condition Based Maintenance Steps

A. Data Gathering

There are two types of data to be gathered, event data and condition measurement data. Event data can be in the form of event or action (damage, inspection, repair, replace, etc.), while condition measurement data is data form of sensor monitoring on equipment (temperature, vibration, etc.)

B. Data Processing

Data processing including data cleaning and data analysis. This step used to prepare data for maintenance decision making

C. Maintenance Decision Making

The decision-making process can be categorized into two categories, diagnostic and prognostic. Diagnostic deals with the identification when the failure occurs, while prognostic is predicting the failures occur.

2.2. Data Analytics

Data analytics is a process of collecting, organizing, analyzing data to gain useful knowledge. It is a set of technique require integration of datasets to reveal hidden value (Youssra & Sara, 2018). The type of data analytics is classified into four categories as follows.

1. Descriptive Analytics

   It is a preliminary stage using historical data to knowing patterns of information. Provide future trends and probabilities.
2. Diagnostics Analytics

   It uses to find the problem root cause. By determining the reason why the behavior of data happens.
3. Predictive Analytics

   It is using historical data to predict future data behavior. By using data mining and artificial intelligence technique.
4. Perspective Analytics

   Dedicated to determining the appropriate action from the information that has been gained.

2.3. K-Means Clustering

K-means clustering is a technique for grouping data objects into k clusters using dissimilar measures (Santosa & Umam, 2018). A measure of dissimilarity using the Euclidean distance. The greater the distance, the data produced from the centroid of data, the higher the dissimilarity of data. The k-means clustering technique is often used for large data. The data used is the result of transformation data which is from pre-processing data. The Euclidean distance formula can be mathematically stated as follows.

$$E= \sum_{j=i}^{k} \sum_{i=1}^{n_j} ||x_i^j - c_j||^2 \tag{1}$$

where:

Xji : i-object to j-cluster

Cj : centeroid of j-cluster

k : number of cluster

Moreover, the formula of the sum of squared error can be written using the following equation.

$$SSE = \sum_{i=1}^{k} \sum_{x \varepsilon D_i} \|x - m_i\|^2| \tag{2}$$

Where:
Di : i-cluster
X : i-data from i-cluster
Mi : centeroid of i-cluster

The steps in solving problems using the k-means clustering technique are as follows.
1. Determine the number of clusters (k) based on information already known.
2. The use of Euclidean distance to determine the dissimilar of data with cluster centroid.
3. Recalculate the new centroid with the cluster members formed.
4. Repeat steps (2) and (3) with the new centroid until has not changed anymore. So that the clustering process can be declared convergent and complete.

2.4. Support Vector Regression (SVR)
Support Vector Regression is the development of the SVM for regression cases. In the case of regression, the output data is in the form of a real or continuous number (Santosa & Umam, 2018). With *$\mathcal{E}$-intensitive loss function*, SVM can be generalized for doing regression (Scholkopf & Smola, 2002). The purpose of SVR is to find a function f(x) that has the largest deviation ($\mathcal{E}$) from the actual target (y_i) for all data training.

To develop the SVR model, this research using R-studio software with the help of the 'SVM' function of library 'e1071'. There are three parameters (gamma, cost, epsilon) that must be set the value. And three kernel options, radial basis, sigmoid and linear. The best parameter values combinations selection is conducted using the grid search method. Below is the detail of three kernel options uses on the SVR model (Santosa & Umam, 2018):

$$\begin{aligned} &\textit{Kernel Radial basis} && = e^{\left(\frac{\left\|x-x_i\right\|^2}{2\sigma^2}\right)} \\ &\textit{Kernel Tangent Hyperbolic} && = \tanh(\beta x^T x_i + \beta_i) \\ &\textit{Kernel Linear} && = x^T x \end{aligned} \qquad (3)$$

3. Research Methods
The research step for data analytics is divided into six steps, i.e. data integration, data cleaning, data selection, data transformation, data mining, and gain knowledge. Data integration is a process to combine data from many sources to obtain important data. Data that has been obtained from data integration then to be processed with data cleaning. The purpose of data cleaning is to eliminate missing value from the data. Then, data selection is a process to determine useful data types from data cleaning. The fourth step is data transformation. The data that has been obtained from data selection convert into another form of data but not change the value. It will make data easier for the next processing and provide accurate results. One of the data transformation methods is scaling. Scaling is a method to convert the data into a certain range. The next step is data mining, which in this study divided into K-Means Clustering and Support Vector Regression. K-Means Clustering is a method for grouping the data into a certain cluster based on its characteristics. Grouping data into a certain cluster using dissimilarity approach usually with Euclidean distance. While SVR is a method for solving regression cases, result in data in the form of a real or

continuous number to predict or forecast the future condition of equipment or machine. And the final step is gaining knowledge from the results of data analytics that have been done.

4. Result and Discussion

A typical cement industry as the object of this study has already applied the condition-based maintenance principle by applying sensors for real time equipment monitoring. Sensor monitoring data is only used for performing corrective maintenance decisions, which means that maintenance is carried out when the machine fails. By the availability of the sensor monitoring data, the predictive instead of the corrective task could be optimized the maintenance program.
This research begins with collecting data from several data sources. Sensor monitoring data are obtained from the company's database from 2015 – 2019. Table 1. shows a list of sensors for both machine, Rawmill machine, and Kiln machine.

Table 1 List of Sensor for Raw Mill and Kiln

Rawmiill Machine	Kiln Machine
RPM sensor (motor belt conveyor) (%)	Motor EPFAN (RPM)
RPM sensor (motor bucket elevator) (%)	Motor IDFAN 1 (RPM)
RPM sensor (rotary feeder) (%)	Motor IDFAN 2 (RPM)
Pressure Sensor (millibar)	Kiln Feed (Ton/Hours)
Vibration Sensor (mm/s)	Temperature Calciner (°C)
	Main drive Kiln (RPM)
	Coal Feeder (Ton/Hours)

All sensor monitoring data will be used as input for developing the data analytics model. Since this research using two machines as a case study with different data analytics models, the detail of each work will be explained further in the following subchapter.

4.1. Support Vector Regression Model on Raw Mill Machine

The failure prediction model will be developed for four main equipment on Rawmill Machine, i.e. motor belt conveyor, motor bucket elevator, rotary feeder, and sealing air fan equipment. The Regression approach on data analytics will be applied using SVR Model. As a general regression model, the SVR model requires predictor variables and the response variable. Sensor monitoring data will be used as predictor variables, while the response variable will be developed through Mean Residual Life (MRL) calculation. Both variables will be used as input data for developing the SVR model on each equipment.

To prevent bias and unclear results in the prediction process due to different ranges of each variable, data transformation is needed to ease the model in recognizing and detecting data trends. Data scaling is being used to transform the data using min-max normalization. Before this process, each variable has a different range of values, e.g. the value of the RPM motor belt conveyor is in the range of 0 – 99, the value of the vibration sensor is in the range of 0 -10. By this data transformation all variables, both predictor and response, have the value in the range of 0 -1. The result of data transformation will be used as input data on modeling steps.

Modelling step used for developing the SVR model for every equipment on Rawmill machine. The objective of the modeling step is obtaining the best parameter for the SVR model on every equipment. Table 2 shows the trial parameters used in the process for creating SVR model.

The grid search method is being used in this research to determine the best parameter. Technically, this method will combine every parameter value in Table 2 to select the best parameter combination. The best combination of the parameter is a combination that produces the smallest RMSE value and does not indicates an overfitting condition. Overfitting is the condition when the data training resulting in a very good performance but giving a low performance on data testing. Good performance indicates a smaller value of RMSE. When the data testing shows the same or smaller RMSE value than RMSE of the data training, it represents the model does not indicate overfitting condition. Table 3 showing the best parameter combination for every equipment.

Table 2 Trial Parameters Used for SVR Model

Gamma (γ)	*Cost of Constraints* (C)	*Insensitive-loss function* (*epsilon*, ϵ)	Jenis *kernel*
0,001	0,1	0,4	*Radial basis*
0,01	1	0,6	*Sigmoid*
0,1	10	0,8	*Linear*
1	100	1	
10	1000	1,2	

Table 3 The Best Parameter Combination for Every Equipment

Equipment	**Parameter**		**RMSE data test**	**RMSE data train**	**Data test – data train**
Motor Belt Conveyor	Gamma	0.01	3.5865388	3.641100322	**-0.05456**
	Cost	1000			
	Epsilon	0.8			
	Kernel	Radial Basis			
Motor Bucket Elevator	Gamma	0.1	3.6111	3.6639	**-0.05456**
	Cost	0.1			
	Epsilon	0.4			
	Kernel	Radial Basis			
Rotary Feeder	Gamma	0.01	5.8882	6.1901	**-0.3073**
	Cost	1000			
	Epsilon	0.6			
	Kernel	Radial Basis			
Sealing Air Fan	Gamma	0.1	42.9535	43.2337	**-0.2801**
	Cost	0.1			
	Epsilon	1.2			
	Kernel	Sigmoid			

According to the testing result in Table 3, those parameter combinations result in the smallest RMSE data testing than the other combination of related equipment, which means the chosen parameter does not cause overfitting on the model.

4.2. K-Means Clustering Model on Kiln Machine

K-means clustering begins with data preprocessing which is aiming to get more accurate analysis results with less data computing time. Data preprocessing consists of data cleaning, data selection, and data transformation. Data cleaning is used to removed missing value from data, i.e. no value data, error data, and incomplete data. Then, in data selection using sensor data from 2015 to 2018. Data transformation makes data values more standardized by changing data into a specific range without changing the information value. By using the scaling technique to convert data into range 0 to 1. Scaling technique using coding execution from Matlab software. So in this way it is expected to speed up the calculation time.

K-means clustering begins with determining the number of clusters. The optimum number of clusters obtained through two methods, elbow method and silhouette index. Calculation elbow method using the sum of squared error which is the sum of squared differences between each observation with the centroid of the cluster. The graph that forms the elbow angle indicates the sum of squared value at that point become the optimum number of cluster. The sum of squared error calculated based on equation (2).

The calculations using Matlab software produce a trend of decreasing the value of the sum of square error (SSE). This is because the greater number of clusters will cause a smaller sum of square error (SSE) value. The greater number of iterations, the more stable the sum of square error (SSE) values generated. But this is also influenced by the algorithm of Matlab software. Where the sum of square error (SSE) values can change even if it is not significant. Figure 2 provides the SSE calculation. A significant decreasing value of SSE shows for cluster 4 to 5.

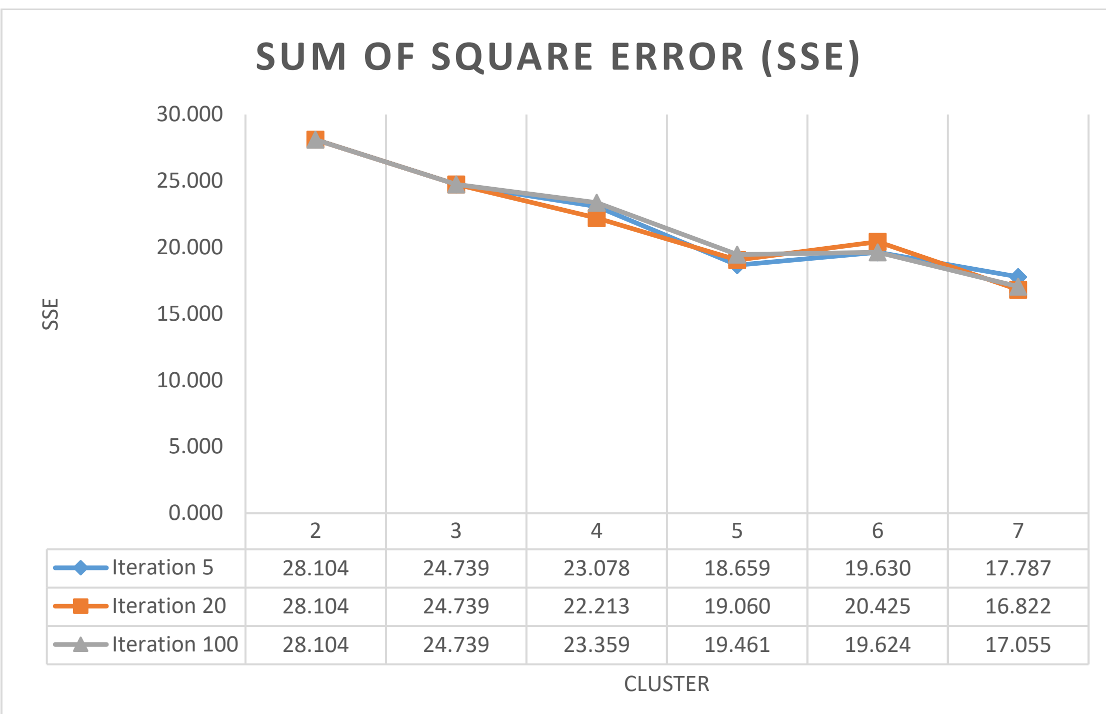

Figure 2. The SSE Calculation

This is also supported by the elbow method by looking at the formation of the elbow angle at the point of the number of clusters 5. Then the number of iterations explains that 100 iteration has a stable sum of square error (SSE) value. So it indicates that the number of cluster 5 with 100 iterations is resulting in the optimum number of clusters.

The second method is using the silhouette index. This method is used to know the quality and robustness of a cluster. This method is a combination of cohesion and separation methods. The silhouette index has a variation in value between -1 to 1. Where the value of silhouette index close to 1 implies that it is close to the optimum cluster. Whereas the silhouette index approaching -1 implies that the number of clusters is less than optimum.

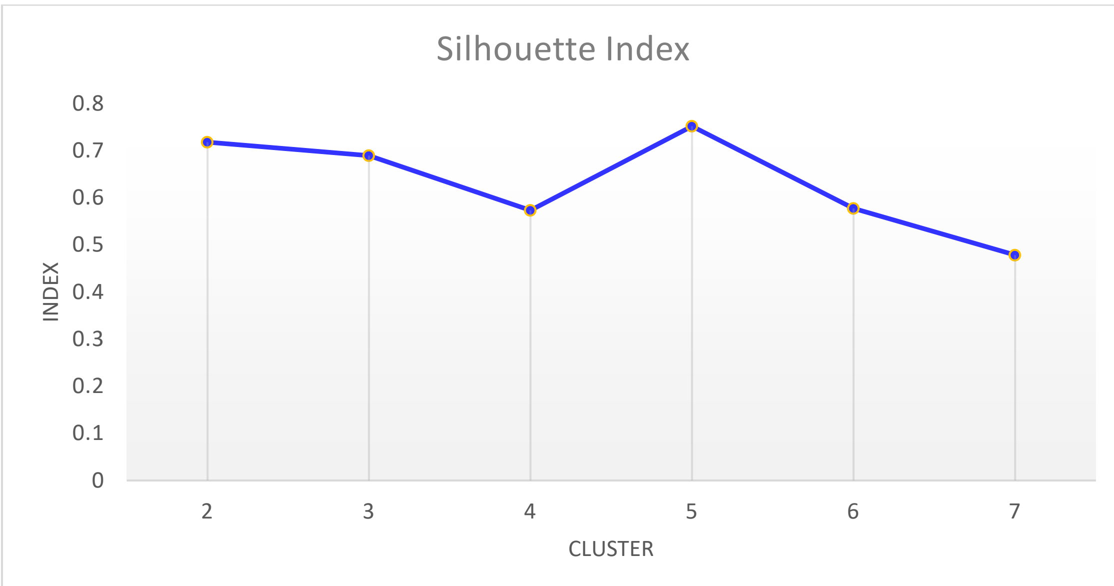

Figure 3. Silhouette Index Calculation

The theoretical analysis of the silhouette index can be proven through the calculation results. Based on the graph it can be seen that the silhouette index graph has fluctuations with a decreasing trend. However, the number of cluster 5 has increased with the highest silhouette index, 0.752. So the silhouette index results confirm that the number of cluster 5 is the optimum cluster.

Table 4 Failure Time Cluster Result

Cluster	Failure Time			
	Avg	Max	Min	Stdev
1	101.122	520	1	123.120
2	145.782	1021	1	170.730
3	309.857	1491	5	494.087
4	111.538	443	3	131.313
5	203.926	1320	4	279.954

Table 5 Repair Time Cluster Result

Cluster	Repair Time			
	Avg	Max	Min	Stdev
1	19.024	166	1	35.818
2	14.527	145	1	1
3	355.429	461	208	104.922
4	1	1	1	0
5	6.037	16	1	4.168

Based on failure time and repair time values, the results of data that have been grouped in 5 clusters with their characteristics. Cluster 3 is a cluster with heavy damage. It can be known from the highest average failure time and repair time, 309.8571 hours, and 355.442 hours. And cluster 4 is a cluster with light damage. It can be known from the lowest average failure time and repair time, 111.53 hours, and 1 hour. While cluster 1, cluster 2, and cluster 5 are clusters with moderate damage. The moderate damage is being classified d into 3 levels based on failure time and repair time. Cluster 5 is a light level, cluster 2 is a moderate level, and cluster 1 is a heavy level.

5. Conclusion

According to the discussion and analysis result that already explained, the data analytics method successfully can be implemented for developing a Condition-Based Maintenance strategy. Support Vector Regression model can be used as a model to predict equipment failure time using sensor monitoring data. K-means clustering model can be used as a model to gain accurate information using data clusters based on their characteristic. By using this method in the typical company, the sensor monitoring data can be used optimally to predict the future condition of machines or equipment. Therefore, the company can decide the right maintenance policy by referring to the prediction result. Besides the advantages that offered, opportunities are wide open to improving this research, as examining other potential machine learning techniques.

6. References

Alsyouf, I. (2009), “Maintenance practices in Swedish industries: Survey results”, International Journal of Production Economics, 121(2), 133-149

Elsayed, E. A., 2012. *Reliability Engineering.* 2nd ed. New Jersey: John Wiley & Sons.

Fathy, A. M. M., 2017. *Condition Based Maintenance,* Cairo, Egypt: Faculty of Engineering and Technology, Mechanica Engineering Department.

Fu, C., Ye, L., Liu, Y., Yu, R., Lung, B., Cheng, Y., Zeng, Y. et al., 2004. Predictive maintenance in intelligent control-maintenance-management system for hydroelectroic generating unit. *IEEE Transactions on Energy Conversion,* pp. 79 - 86.

Jardine, A., Lin, D. & Banjevic, D., 2006. A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance. *Mechanical System and Signal Prosessing,* pp. 1483 - 1510.

Krenek, J. et al., 2016. Application of Artificial Neural Networks in Condition Based Predictive Maintenance. *Studies in Computational Intelligence,* pp. 1 -11.

Leger, J.-B., Iung, B. and Morel, G. (1999b), Integrated design of prognosis, diagnosis and monitoring processes for proactive maintenance of manufacturing systems, *IEEE Systems, Man, and Cybernetics Conference SMC'99 Proceedings*, vol.4, 492-498

Moura, M. d. C., Zio, E., Lins, I. D. & Droguett, E., 2011. Failure and Reliability Prediction by Support Vector Machines Regression of Time Series Data. *Reliability Engineering and System Safety,* Volume 1534, p. 1527.

Prajapati, A., Bechtel, J. & Ganesan, S., 2012. Condition Based Maintenance. *Journal of Quality in Maintenance Engineering,* Volume IV, pp. 384 - 400.

Restagari, A., Salonen, A., Bengtsson, M., Wiktorsson, M., 2013. Condition Based Maintenance in Manufacturing Industries: Introducing Current Industrial Practice and Challenges. *22nd International Conference on Production Research, ICPR, Parana, Brazil*.

Sahu, H., Shrma, S. & Gondhalakar, S., 2010. A Brief Overview on Data Mining Survey. *International Journal of Computer Technology and Electronics Engineering (IJCTEE),* I(3), pp. 114 - 121.

Sampaio, G. S., Filho, A. R. d. A. V., da Silva, L. S. & da Silva, L. A., 2019. Prediction of Motor Failure Time Using An Artificial Neural Network. *Sensors,* pp. 1 - 17.

Santosa, B. & Umam, A., 2018. *Data Mining dan Big Data Analytics.* 2 ed. Yogyakarta: Penebar Media Pustaka.

Susanto, B.S.P. & Kurniati, N., 2020. Multi Sensor-Based Failure Diagnosis using the Mahalanobis Taguchi System. *IOP Conf. Series: Materials Science and Engineering* 847 012036.

Suyuti, A., Tahir, Z., Alhari, F. & Rafrin, M., 2013. *Analisis Manajemen Perawatan Mesin Industri dengan Menggunakan Jaringan Saraf Tiruan Feed Forward-Back Propagation (Studi Kasus PT. Semen Bosowa Maros).* Makassar, Seminar Nasional Teknik Informatika (SNATIKA).

Youssra, R. & Sara, R., 2018. Big Data and Big Data Analytics: Concepts, Types and Technologies. *International Journal of Research and Engineering,* 5(9), pp. 524-528.

Rafif Nova Riantama is the final year undergraduate student at the Department of Industrial and Systems Engineering (ISE). Institut Teknologi Sepuluh Nopember (ITS) Indonesia. He is actively involved in student organizations. He was the coordinator of the student assistant at the Manufacturing System Laboratory. He also enthusiastically craft the curiosity by challenging himself to actively participate in several student competitions. He already got several prizes for those competitions, both national or international.

Abiyyu Dimas Prasanto is also an active student assistant at the Manufacturing System Laboratory, Department of Industrial and Systems Engineering (ISE). Institut Teknologi Sepuluh Nopember (ITS) Indonesia. He is involved in helping teachers who deliver the courses under the Manufacturing System Labs responsibility. The deliverables he has proposed are delivering the software and calculation tutorial in class, as well as helping for developing the module for practicum.

Nani Kurniati is an active associate professor at the Department of Industrial and Systems Engineering (ISE). Institut Teknologi Sepuluh Nopember (ITS) Indonesia. She is interested in and has experience researching the reliability and maintenance field, as well as in the quality engineering and management field. She was a Deputy of Quality Assurance at University level, work at Office of Quality Assurance, ITS. Now, she is the Secretary of Department for Academic and Student Affairs.

Dewanti Anggrahini is an active lecture at the Department of Industrial and Systems Engineering (ISE). Institut Teknologi Sepuluh Nopember (ITS) Indonesia. She is mainly interested in manufacturing system analysis. She shows great experience in developing Lean concepts and implementation. Teaching and assisting students in Laboratory and Student Organization are her very good performance reported.