Fault Classification of Pump Using *Support Vector Machine* (SVM) Method

Avie Aura Dzilfadhilah
Engineering Management Department
Universitas Internasional Semen Indonesia
Gresik, Indonesia
avie.dzilfadhilah17@student.uisi.ac.id

Anindita Adikaputri Vinaya
Engineering Management Department
Universitas Internasional Semen Indonesia
Gresik, Indonesia
anindita.vinaya@uisi.ac.id

Nicky Yesica
Engineering Management Department
Universitas Internasional Semen Indonesia
Gresik, Indonesia
nicky.yesica15@student.uisi.ac.id

***Abstract*— A machine is a mechanical or electrical equipment using the working principle of converting energy to assist human activities or produce certain products. The condition of a machine should be maintained and monitored in a good condition. Therefore, the condition of the machine needs to be detected before serious damage occurs. This study aims to detect the type of pump damage with a machine learning approach. The object of this study was water pump with Panasonic GP-129 type. The goal in this study is to classify three fault of pump conditions, those are misalignment, unbalance, and bearing fault. Based on the results obtained, the classification of pump fault using SVM methods had average accuracy of 98.35% on the Linear SVM and Cubic SVM models, and average accuracy of 100% on the Quadratic SVM model.**

Keywords*— *Fault, Pump, SVM

I. Introduction

In industrial activities, rotating machines such as pumps, compressors, and etc. are very important assets. The condition of a machine must be maintained and monitored in good condition because any severe damage will reduce machine productivity and increase maintenance costs. It is necessary to maintain the condition of the machine. In general, the damage to the rotating machine that often occurs is misalignment, unbalance, and bearing fault. Misalignment can occur because the coupling shafts are not on the same axis line. The unbalance occurs because the rotor mass is not evenly distributed, causing centrifugal force. Bearing fault occurs due to defects in the bearings which are characterized by the appearance of amplitude peaks at high frequencies [1].

Rotating engine can be detected damage based on acoustic signal. In 2018, Vinaya et al identified machine errors based on time and frequency domain parameters using fuzzy logic. Four engine conditions, namely normal, misalignment, unbalance, and bearing fault were detected with an accuracy of up to 90% [2]. Every method has respective advantages and disadvantages. The advantages using fuzzy logic is intuitive, adaptive in terms of optimization, and efficient computating especially in terms of nonlinear dynamic system. Fuzzy logic has the advantages of being intuitive, adaptive in terms of optimization, and computationally efficient, especially in terms of nonlinear dynamic systems. However, this approach requires a human understanding of inputs, outputs and the rules that link them all [3].

Machine learning is an area of artificial intelligence that has emerged as part of the ongoing quest to build intelligent machines capable of learning. Machine Learning is a system to process and analyze big data easily. Supervised Learning is a type of machine learning to classify machine errors. Support Vector Machine (SVM) is a classifier method that is commonly used and has good performance. SVM will plot the classes based on the hyperplane [4].

Awalin (2019) identified disturbances in the distribution system using phase voltages and current signals using SVM. Based on his research, the SVM method is able to be an alternative to detect types of errors in distribution systems with an error identification accuracy using SVM reaching 100% [5]. Gao Guahua (2006) diagnoses dental disorders with vibration signals using SVM. [6]. Amadi (2015) diagnosed bearing defects in eight classes using the SVM method and the diagnosis accuracy with the SVM method reached 99% [7]. In this study, SVM will be used to classify the type of pump damage. The object of this research uses 3 "Panasonic GP-129" type water pumps which have been damaged, namely misalignment, unbalance, and bearing fault.

.

II. Literature

A. Type of Machine Fault

Misalignment is a condition where the centerline of two ahaft are not collinear (not in the same axis line). In the misalignment of additional dynamic loads will accelerate engine damage. This type of vibration often comes from reactive forces in the coupling between the two rotating shafts. Generally, misalignment presents fundamental harmony is at 1x, 2x, 3x rpm due to the strain induced in the shaft.

Unbalance as the un-equal distribution of weight around the center of rotation. It will cause a heavy spot at one point. This will create a deflection on the formation of centrifugal force in the radial direction. Unbalance present fundamental harmony is at 1x rpm. In frequency domain, unbalance will show a shaft with an amplitude in direct proportion to the centrifugal force [8].

Bearing Fault is the result of fatigue bearing material. Under normal operating conditions, fatigue failure begins with small cracks located within the surface of the raceway track and rolling element. The vibration signal pattern consists of successive oscillations that repeat with each path of the moving component through the disturbance. The frequency of impact repetition depends on the fault position [9].

B. Support Vector Machine (SVM)

Machine learning helps discover patterns in the data and use these patterns to make predictions about new data. To achieve this, however, it is important to clean, explore and prepare the data to improve the overall quality of dataset. The simplest way to assess into two subsets, the training set and the test set. First, the machine learning algorithm is fitted using the training set. The model is then evaluated using the test set, this step consists in assessing the performance of the learned function in unseen data, which the model did not use for training [4].

SVM is a harmonious series of concepts in the field of pattern recognition introduced by Vapnik in 1992 at the Annual Workshop on Computational learning theory. SVM is one of machine learning method that works on the principle of Structural Risk Minimization (SRM) which aims to find the best hyperplane that separates two classes in the input space [10]. SVM is an usefu algorithm with supervised machine learning that can be used for solving classification problem. Any training samples that fall on the marginal hyperplanes. A hyperplane is a line that separates and classified into two classes, clickbait and non-clickbait. Margin is the distance between the marginal hyperplanes.

The SVM takes the training dataset and finds the optimal hyperplane, and then separates all the featured data objects into two classes, Clickbait and Non-clickbait. The features $x_1 \ldots x_n$ are the Tern Frequency Inverse Document Frequency (TFIDF) of title, TFIDF of body and cosine similiarity and the class label, y_1 is either clickbait or non-clickbait. The output class y_i is classified into two classes, clickbait ($y_i = +1$) and non-clickbait ($y_i = -1$).

The Hyperplane *H* and Marginal hyperplane *H1* and *H2* equations are:

$$H : w^T x_i + b = 0$$

$$H1 : w^T x_i + b = -1$$

$$H2 : w^T x_i + b = 1$$

Where, w^T represents transpose of weight vector and b representas bias. The data poins that were correctly classified should satisfy the inequality:

$$y_i(w^T x_i + b) \geq 1 \quad \text{for} \quad x_i, i = 1, 2, ... [11]$$

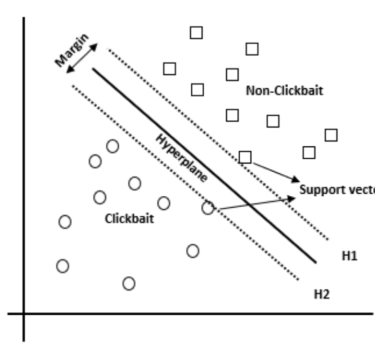

Fig. 1. Support Vector Machine [11]

C. Predictor Variables

Predictor variable is variable used to predict the response. Response variable is the variable is trying to predict. In this study, the response variable was the machine fault.and the predictor variables are Root Mean Square, Kurtosis, Skewness, Crest factor, Shape factor, and Instantaneous Frequency.

Root Mean Square is an indicator of vibration signal. RMS is good at identifying the overall noise level. RMS is good to identify noise level, however, does not provide the information of fault component location [12]. N is the number of data, x_i is the data value. RMS formulated, as in:

$$\mathrm{RMS} = \frac{1}{N}\sqrt{\sum_{i=1}^{N} x_i^2} \tag{1}$$

Kurtosis is distortion of normal curve. Kurtosis is measured by comparing the tapering shape of the curve and normal curve. μ is the average value, σ is standard deviation. Kurtosis can be symbolized by $\propto_4$, as in:

$$\propto_4 = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^4}{\sigma^4} \tag{2}$$

Skewness is the degree of asymmetry from distribution normal curve [13]. The formula of skewness ($\propto_3$), as in:

$$\propto_3 = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^3}{\sigma^3} \tag{3}$$

Crest Factor is ratio between the peak value of wave to the RMS. Crest Factor is also one of the important characteristics that can be used for engine condition. The formula, as in:

$$\mathrm{CF} = \frac{x_{max}}{\mathrm{RMS}} \tag{4}$$

Shape Factor is ratio RMS value to average value. Shape factor is one of symtoms variable Symtoms variables is use to identify machine condition because can show the indicate information of measure signal [12]. The formula, as in:

$$\mathrm{SF} = \frac{\mathrm{RMS}}{\mu} \tag{5}$$

Instantaneous Frequency is converted time domain by Fast Fourier Transform (FTT) algorithm. FTT is a very efficient method for calculating coefficients from discrete Fourier to a finite sequence from complex data. S(f) is a signal in frequency domain, s(t) is a signal in time domain, is $e^{-j2\pi ft}$ constant of a signal, f is the frequency, and t is the time domain [14]. The formula, as in:

$$S(f) = \int_{-\infty}^{\infty} s(t) e^{-j2\pi ft} dt \tag{6}$$

D. Accuracy in Classification

Accuracy in classification is the percentage of accuracy of data records that are classified correctly after testing the classification results. The higher the accuracy results, the more effective the classification model. The formula, as in:

$$\mathrm{Accuracy} = \frac{TP + TN}{P} \tag{7}$$

TP (True Positive) is the number of positive records that are correctly labeled by the classification algorithm. TN (True Negative) is the number of negative records that are labeled incorrectly by the classification algorithm. P is the total of all evaluated records [15].

III. EXPERIMENTAL SETUP

The object of research used 3 modified water pumps type Panasonic GP-129. Each pump has one fault, including, a pump modified into pump with misalignment, a pump modified into pump with unbalance, and a pump modified into pump with bearing fault. The recording was carried out in the Laboratory of acoustic engineering and building physics in the Department of Physics Engineering, ITS Surabaya.

The record used a microphone at 5 cm from pump for 5 second and frequency sampling was 48000 Hz. Microphone used type DBX_RTA Microphone. Microphone had connected with interface, which converted analog data to digital data. The interface used type USB DAC Multi Channel (focusrite scarlett 18i8) with 4 inputs and 4 outputs. The software to save the record was Adobe Audition CC 2015, so output as audio track with .wav. format. Each pump took 20 times recording. Total data used in this research was 60 data recording.

Fig. 2. Panasonic GP-129 Pump

Fig. 3. DBX_RTA Microphone

Fig. 4. USB DAC Multi Channel (focusrite scarlett 18i8) Interface

A. Variable Extraction

The pump recording results need to be extracted into predictor variables to facilitate the characterization of the type of pump failure. The variables used in this study were RMS, kurtosis, skewness, crest-factor, shape-factor, and instant frequency. The extraction of the 1-10th wavfiles from each pump to preditors variables will be used as training data, while the 11-20th wavfiles will be used as test data. The total training dataset is 30 and the total test dataset is 30. The following is the extraction result from the misalignment pump.

TABLE I. VARIABLE EXTRACTION OF MISALIGNMENT PUMP

No.	RMS	Kurtosis	Skewness	Crest-Factor	Shape-Factor	Instantaneous Frequency (Hz)
1	0,0042	2,4217	0,1544	3,7171	617,43	49,62
2	0,0041	2,4598	0,1564	3,427	520,85	49,62
3	0,0041	2,4552	0,1529	3,7374	1000,63	49,62
4	0,0041	2,4629	0,1419	3,3887	398,37	49,62
5	0,0041	2,4167	0,1404	3,4762	352,87	49,62
6	0,0042	2,3905	0,1325	3,5921	657,46	49,62
7	0,0044	2,4091	0,0754	3,2217	1345,77	49,62
8	0,0043	2,4932	0,0992	3,2195	636,67	49,62
9	0,0043	2,4731	0,1061	3,1365	380,03	49,62
10	0,0044	2,4928	0,0926	3,2712	1576,83	49,62
11	0,0043	2,5087	0,0934	3,1301	493,78	49,62
12	0,0043	2,494	0,1036	3,2952	1041,54	49,62
13	0,0043	2,5372	0,0902	3,2734	643,48	49,62
14	0,0044	2,5185	0,0937	3,3139	545,08	49,62
15	0,0043	2,5112	0,0907	2,9414	659,46	49,62
16	0,0043	2,5161	0,0833	3,1763	774,96	49,62
17	0,0043	2,5001	0,0841	3,1096	473,39	49,62
18	0,0043	2,4956	0,0859	2,8793	909,94	49,62
19	0,0043	2,4966	0,0943	3,0568	533,76	49,62
20	0,0043	2,4941	0,0882	3,2994	491,13	49,62

B. Variable Evaluation

We evaluated the relationship of the predictor variables used to the response variables. The correlation test in this study was used to determine how the tendency of the relationship between predictor variables and response variables was. Predictor variables are the result of data extraction carried out in the previous stage, including RMS, kurtosis, skewness, crest-factor, shape-factor, and instantaneous frequency. The response variable is the actual engine failure condition. Based on Table II, all variables have a tendency to be related to the response variable so that all variables will be used as predictor variables.

TABLE II. RELATIONSHIP OF THE PREDICTOR VARIABLES TO RESPONSE VARIABLE

Predictor Variables	**r value**	**Correlation to response variable**
RMS	0,776	Strong
Kurtosis	0,726	Strong
Skewness	-0,362	Weak
Crest-Factor	0,688	Strong
Shape-Factor	0,540	Moderate
Instantaneous Frequency	0,866	Very strong

C. SVM Classification

The training process will use 5-fold cross-validation. Cross-Validation is one of the most widely used data resampling methods to estimate the true model prediction error [16]. The training uses the Linear SVM, Quadratic SVM, and Cubic SVM methods. The accuracy of the training data is presented using a confusion matrix to facilitate the detection of prediction errors in machine learning.

IV. RESULTS

The results of the training process using SVM are displayed in a matrix. In the confusion matrix, the y axis was True Class, which is the actual condition of the pump fault. The x axis was Predicted Class, which is the result of machine failure conditions predicted by SVM Linear. The value of 10 in the blue box was 10 predicted data according to the actual condition of the pump damage. A white box that has no value was no a prediction error. Amount of true predicted was 10 data bearing faults, 10 data misalignments, and 10 data unbalances. Total of true predicted was 30 data and the total data was 30. Accuracy of training data using Linear SVM was 100%.

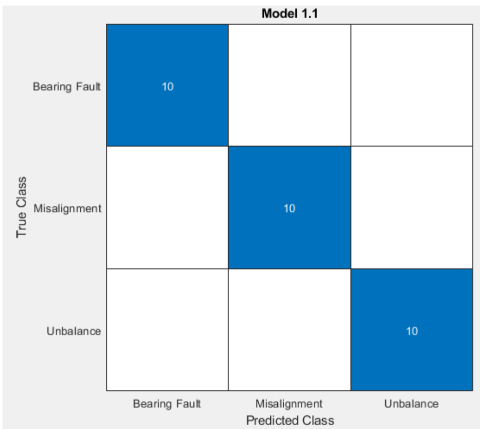

Fig. 5. Linear SVM Confusion Matrix

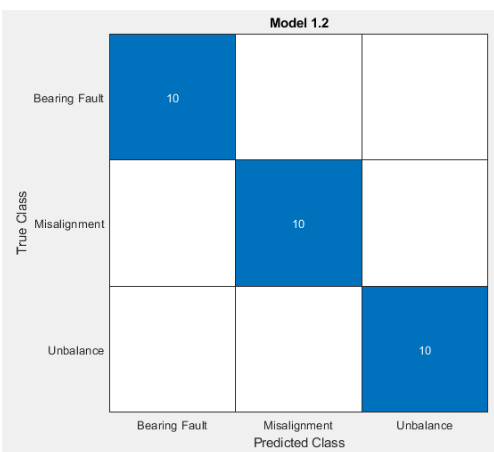

Fig. 6. Quadratic SVM Confusion Matrix

Based on Fig.6, the matrix showed no prediction error. Amount of true predicted was 10 data bearing faults, 10 data misalignments, and 10 data unbalances. Total of true predicted was 30 data and the total data was 30. Accuracy of training data using Quadratic SVM reach 100%. The Cubic SVM also got the same result

The prediction of pump fault will be done based on the model that has been obtained from training data learning. Data extraction is carried out with 10 other wavfile from each pump as testing data.

TABLE III. PREDICTION RESULT WITH SVM

No	True Class	Linear SVM	Quadratic SVM	Cubic SVM
1	Misalignment	Misalignment	Misalignment	Misalignment
2	Misalignment	Misalignment	Misalignment	Misalignment
3	Misalignment	Misalignment	Misalignment	Misalignment
4	Misalignment	Misalignment	Misalignment	Misalignment
5	Misalignment	Misalignment	Misalignment	Misalignment
6	Misalignment	Misalignment	Misalignment	Misalignment
7	Misalignment	Misalignment	Misalignment	Misalignment
8	Misalignment	Misalignment	Misalignment	Misalignment
9	Misalignment	Misalignment	Misalignment	Misalignment
10	Misalignment	Misalignment	Misalignment	Misalignment
11	Unbalance	Unbalance	Unbalance	Unbalance
12	Unbalance	Unbalance	Unbalance	Unbalance
13	Unbalance	Unbalance	Unbalance	Unbalance
14	Unbalance	Unbalance	Unbalance	Unbalance
15	Unbalance	Misalignment	Unbalance	Misalignment
16	Unbalance	Unbalance	Unbalance	Unbalance
17	Unbalance	Unbalance	Unbalance	Unbalance
18	Unbalance	Unbalance	Unbalance	Unbalance
19	Unbalance	Unbalance	Unbalance	Unbalance
20	Unbalance	Unbalance	Unbalance	Unbalance
21	Bearing Fault	Bearing Fault	Bearing Fault	Bearing Fault
22	Bearing Fault	Bearing Fault	Bearing Fault	Bearing Fault
23	Bearing Fault	Bearing Fault	Bearing Fault	Bearing Fault
24	Bearing Fault	Bearing Fault	Bearing Fault	Bearing Fault
25	Bearing Fault	Bearing Fault	Bearing Fault	Bearing Fault
26	Bearing Fault	Bearing Fault	Bearing Fault	Bearing Fault
27	Bearing Fault	Bearing Fault	Bearing Fault	Bearing Fault
28	Bearing Fault	Bearing Fault	Bearing Fault	Bearing Fault
29	Bearing Fault	Bearing Fault	Bearing Fault	Bearing Fault
30	Bearing Fault	Bearing Fault	Bearing Fault	Bearing Fault

The white cell in Table I was the predicted result suitable with actual condition (no error). However, the red cell was had error in predicted result. Total of testing data was 30. The predicted result using Linear SVM and Cubic SVM each produced an error.

TABLE IV. COMPARISON OF ACCURACY ON THE SVM MODEL

Method	Accuracy		
	Training Data	Testing Data	Average
Linear SVM	100%	96.7%	98.35%
Quadratic SVM	**100%**	**100%**	**100%**
Cubic SVM	100%	96.7%	98.35%

Based on the the accuracy results shown in the Table IV, prediction of pump fault with training data and testing data using Linear SVM, Quadratic SVM, and Cubic SVM were appropriated with actual machine conditions. Quadratic SVM model has the highest accuracy among others. Linear SVM and Cubic SVM had the same average accuracy of 98.35%.

Prediction of machine damage on training data using SVM Linear, SVM Quadratic, and SVM Cubic methods is in accordance with the actual machine conditions. In the test results with new data, SVM Linear and SVM Cubic have an accuracy rate of 96.7% each, with an error while SVM Quadratic produces the highest level of accuracy, reaching 100%. The error lies in the unbalance damage. It can be caused the predictor variables in the unbalance test data number 15 (Tabel V) have skewness and shape-factors that are close to the misalignment class (Tabel I).

TABLE V. VARIABLE EXTRACTION OF UNBALANCE PUMP

No.	RMS	Kurtosis	Skewness	Crest-Factor	Shape-Factor	Instantaneous Frequency (Hz)
1	0,0032	2,5517	-0,2840	2,8875	618,15	49,62
2	0,0033	2,5034	-0,2942	3,0057	1397,54	49,62
3	0,0034	2,4763	-0,2909	2,8801	267,84	49,62
4	0,0034	2,2275	-0,1132	2,8865	1083,60	49,62
5	0,0036	2,443	-0,3128	3,0201	525,18	49,62
6	0,0035	2,3395	-0,2206	2,9845	549,76	49,62
7	0,0041	2,2542	-0,2162	2,9496	483,92	49,62
8	0,0040	2,2579	-0,1897	3,0062	368,46	49,62
9	0,0040	2,2021	-0,1663	2,9334	587,34	49,62
10	0,0039	2,2951	-0,1752	2,9028	656,34	49,62
11	0,0039	2,2786	-0,1276	3,1080	351,22	49,62
12	0,0039	2,2788	-0,1324	2,9042	416,96	49,62
13	0,0034	2,408	-0,0377	3,2624	374,03	49,62
14	0,0035	2,3257	-0,0186	3,3864	372,67	49,62
15	0,0034	2,3273	0,0135	3,5831	351,96	49,62
16	0,0035	2,3085	0,007	3,2791	353,19	49,62
17	0,0035	2,323	-0,0391	2,9881	613,13	49,62
18	0,0035	2,2997	0,0181	3,1919	527,20	49,62
19	0,0035	2,3121	0,00043	3,1395	368,41	49,62
20	0,0034	2,2471	0,037	3,0135	597,34	49,62

Both types of damage also have the same instantaneous frequency, which is 49.62 Hz. The difference in the prediction results on the test data for the SVM Linear, SVM Quadratic, and SVM Cubic methods is also due to differences in the shape of the hyperplane (class separator) based on the algorithm of each method.

V. CONCLUSION

Classification of pump fault using SVM methods had been carried out in this study. Some of the SVM methods used are Linear SVM, Quadratic SVM, and Cubic SVM . The Quadratic SVM method did not produced prediction error. Based on the results obtained, the classification of pump engine damage using SVM methods has an average accuracy rate of 98.35% on the Linear SVM and Cubic SVM models, and 100% on the Quadratic SVM model. Linear and Cubic SVM method produced 1 prediction error. The average prediction error is in the condition of *unbalance* was detected as *misalignment.*

VI. REFERENCES

[1] F. Huda, Nazaruddin and M. Dovani, "Analysis Of Rotordinamic Sound Due to Unbalace, Misalignment, and Looseness," *Annual National Machine Engineering Seminar XIV (SNTTM XIV),* 2015.

[2] A. A. Vinaya, Q. A. M. O. Arifianti, N. Yessica, D. Arifianto and S. Aisjah, "Fault Diagnosis of Water Pump Based on Acoustic Emission Signal Using Fast Fourier Transform Technique and Fuzzy Logic Inference," *Internasional Conference on Engineering, Science, and Industrial Applications (ICESI),* 2019.

[3] D. Syahputra, Tulus and Sawaluddin, "The Accuracy Of Fuzzy Sugeno Method With Antropometry On Determination Natural Patient Status," *IconICT,* no. 930, 2017.

[4] A. Machelli and S. Vieira, Machine Learning Methods and Applications to Brain Disorders, London, United Kingdom: Academic Press, 2020.

[5] L. J. Awalin, K. Naidu and H. Suryono, "Fault Types Classification Using Support Vector Machine (SVM)," *AIP Conference Proceedings,* vol. 2129, no. 020132, 30 July 2019

[6] G. Gao, Y. Zhang, Y. Zhu and G. Duan, "Data Fusion and Multi-fault Classification Based On Support Vector Machines," *9th Joint International Conference on Information Sciences (JCIS-06),* 2006.

[7] D. N. Amadi, "Penerapan Metode Suport Vector Machine (SVM) untuk Diagnosis Kerusakan pada Bantalan Gelinding," *Agri-tek,* vol. 16, no. 1, pp. 62-73, 1 March 2015.

[8] A. Nejadpak and C. X. Yang, "Misalignment and Unbalance Faults Detection and Identification Using KNN Analysis," in *CANCAM 201726th Canadian Congress of Applied Mechanics*, Victoria, BC Canada,2017.

[9] C. Harlişca and L. Szabo, "Bearing Faults Condition Monitoring – A Literature Survey," *Researchgate,* January 2015.

[10] Sukendi, I. Isranuri and Suherman, "Analisa Karakteristik Getaran dan Machine Learing untuk Deteksi Dini Kerusakan Bearing," *Widya Teknika,* vol. 23, no. 2, pp. 41-49, 2015.

[11] S. R. Dam, S. P. Panday and T. B. Thapa, "Detecting Clickbaits on Nepali News using SVM and RF," *Proceedings of 9th IOE Graduate Conference,* vol. 9, pp. 140-146, March 2021.

[12] A. Ramali D, Setiyono B. and Hidayatno A., "Identification Of Machine Fault Based on Vibration Signal Using Fuzzy Logic Method," *Thesis, Electro Engineering, University of Diponegoro,* 2012.

[13] D. Fahmeyzan, S. Soraya and D. Etmy, "Normality Analysis of Monthly Turnover of Microeconomic Actors in Senggigi Village Using Skewness and Kurtosis," *Jurnal Varian,* vol. 2, no. 1, pp. 31- 36, 2018.

[14] D. Koesman S, "Modification of Portable Welding Based on DC Electric Motor for TIG Welding on Steel Plate Parallel Connection Type," *Doctoral Dissertation, University of Diponegoro,* 2017.

[15] M. Yusa, E. Utami and E. T. Luthfi, "Comparative Analysis of Classification Algorithm Performance Evaluation on Diabetes PatientReadmission," *Buana Informatika,* vol. 7, no. 4, pp. 293-302, 2016.

[16] D. Berrar, "Cross-Validation," *Encyclopedia of Bioinformatics and Computational Biology,* 2018.