

Classification Model Disease Risk Areas Endemicity Dengue Fever Outbreak based Prediction of Patients, Death, IR and CFR using Forecasting Techniques

Agus Qomaruddin Munir
Faculty of Science and Technology
Respati University Yogyakarta
Yogyakarta - Indonesia

Edi Winarko
Dept. of Computer Science and Electronics
Faculty of MIPA Gajah Mada University
Yogyakarta - Indonesia

ABSTRACT

Forecasting model time series data becomes a process of model use to create prediction (forecast) to an event in the future based on past event which is known as data. The data gives benefit to the interrelatedness among factors that can be the predictor variables. Deep investigation to time series data will give information to find out the data pattern. One of the examples of data series is dengue fever phenomena. Dengue fever is a contagious disease which can become an endemic. The spread of dengue fever needs a particular supervision from related party under the supervision of health department. Dengue fever control is used to avoid the disease to stop spreading. It is done to give important information, so the society can know endemic area from a particular social environment. The research is aimed to do improvisation toward dengue fever surveillance system by conducting a forecast of dengue fever occurrences in a particular area. Forecast technique was conducted with three (3) algorithm approaches. The test of three algorithm forecasting was done to predict dengue fever occurrences in 12 months and to help the activity to control spreading cases of dengue fever to stop spreading in the area of Yogyakarta. Forecast experiment approach from three methods in which linear regression, multilayer perceptron, and sequential minimal optimization regression (SMOreg) showed that linear regression had better accuracy compared to sequential minimal optimization regression (SMOreg) and multilayer perceptron.

General Terms

Classification, Data Mining

Keywords

Dengue Fever, Forecast, Linear Regression, Multilayer Perceptron, SMOreg, Time Series.

1. INTRODUCTION

Forecasting time series model is a technique which is used to predict the variables arranged by occurrences or events based on time series (periodic series). Time series data analysis in health issue is significantly needed. Time series data analysis of contagious disease surveillance importantly helps to develop hypothesis to explain and anticipate the dynamics of the phenomena investigated, and it is useful in creating quality control system and resource reallocation [1].

Several forecast techniques with data mining model approaches are Gaussian process, linear regression, multilayer perceptron, and sequential minimal optimization regression (SMOreg). Every predictor uses learning algorithm to find out the most appropriate model for the correlation between set attribute and data class input [2].

Model created by learning algorithm should be suitable to data input which can be predicted correctly toward the record of prediction result that has not seen previously. The objective of learning algorithm is to develop model with good generalization model, so there can be suitable practical model based on time series data [3]. Dengue fever occurrences can be noted as a recording of time series data of a particular contagious disease. Dengue fever is an arboviral disease with wide geographic range and high occurrences. Dengue fever is a dangerous pandemic caused by dengue virus from flavivirus spread by *Aedes aegypti* mosquito [4], [5]. Every year, it is estimated that there are 50 million infection; 500,000 more occurrences with 20,000 lethal occurrences [6].

WHO reports that the return of contagious disease is an alert that the development that has been achieved to global health and global prosperity is failed. The number of people in the world is around 2.5 billion with the 2/5 parts of them live in tropical and sub-tropical area [7]. Those people have high risk to make contact with dengue fever. Dengue fever endemics are in more than 100 countries in Africa, America, East Mediterranean, South East Asia, and West Pacific. South East Asia and West Pacific become the most severe area to get the bad impact of that endemic disease [8].

In Indonesia, dengue fever becomes significant problem of community health during the last 41 years. Since 1968, there has been some increase cases number of the spread of dengue fever province and district endemic area. From 2 provinces and 2 municipalities in 1968 become 32 (97%) and 382 (77%) district/municipalities in 2009. Maluku province from 2002 until 2009 did not report any dengue fever occurrences. Besides, there is always increasing cases of dengue fever; in 1968 there were only 58 cases. It increased to be 158,912 cases in 2009 [9].

One of the provinces which becomes the epidemic area of dengue fever is Yogyakarta Special District. The data from Health Department Yogyakarta Province (2013) shows that in 2013 Case Fatality Rate (CFR) of dengue fever in Yogyakarta is 0.5 (national <1) with IR in 2012 were 28.8/1,000 people. There was CFR decrease 0.21 in 2012 with 971 numbers of dengue fever cases (Health Department of Yogyakarta, 2013). Here it is the table of dengue fever occurrences in Yogyakarta province.

Table 1. DF Occurrence in Yogyakarta Province in 2013

No	Distric	Case
1	Kota Yogyakarta	905

2	Bantul	1203
3	Kulon Progo	144
4	Gunung Kidul	310
5	Sleman	736

Until in the end of April 2014, the occurrence numbers of dengue fever in Yogyakarta Province were 3298 cases. According to Health Department data, Yogyakarta is an endemic area because the occurrence case happened in the last 3 years respectively in all districts in the region [10]. Although there have been some steps to overcome the spread of the disease, there should be analysis method to do forecasting from time series data, so people can make any anticipation before [11].

This research discusses about forecasting model and time series data analysis of dengue fever. Technically, time series data attribute gives an indicator that can be a clue in making forecasting related to dengue fever case in a particular period. Forecasting result can be a guidance as a determiner of the risk of dengue fever endemic. The objective of the study is to make forecasting and to analyze toward the number of dengue fever cases in Yogyakarta area in a particular time during 12 months in 2007 by using time series data set from 1997 to 2005. Forecasting was done by forecasting data mining forecast approaches which are comparison three algorithms: linear regression, multilayer preception, and sequential minimal optimization for regression (SMOreg).

This study is divided into 7 stages. The first part is an introduction to the background of the issues related to the research topic. The second part contains the rules of dengue endemic area. The third section discusses general overview of the forecasting model in data mining. The fourth section discusses the research methodology comprising the steps of research, related data sources, and methods used. The fifth section contains discussion and examination results and forecasting models. The sixth part of the conclusion. The seventh part reference.

2. ENDEMIC CLASSIFICATION AREA OF DENGUE FEVER

Based on the guidance of Disease Surveillance Epidemiology (PEP), Contagious Disease Eradication and Environmental Health General Directorate, Indonesian Health Department, the classification of pandemic area of dengue fever is as follow:

- Risky Area I (endemic) is the area which always reports any dengue fever occurrences in three (3) last months.
- Risky Area II (sporadic) is the area which there is any dengue fever occurrences in the last three months.
- Risky Area III (potential) is the area which there is no reported case of dengue fever, but the area has dense population, risky transportation, and found mosquito larvae > 5%.
- Free Area is the area which never reports any dengue fever case.

Based on the explanation, there can be a rule upon the statement above as follows:

Table 2. Rule of Dengue Fever Endemic Area

RULES
K1: if always a case of DF 3 consecutive months then endemic areas
K2: if cases of dengue 3 consecutive months then sporadic areas
K3 : if there is no case and densely populated and vulnerable transportation and larvae > 5% then the potential area
K4: if not unheard of dengue cases then the free area

3. FORECASTING MODEL TIME SERIES DATA

Data forecasting is a technique in doing a prediction toward a particular event which correlates to the fact found in the field based on time series analysis (historical data) in the previous time series. Time series analysis is an analysis which is conducted based on past value from a particular past variable with a particular objective to determine the pattern in a historical data time lines and to reflect the time series as the forecasting condition of the future.

The research uses three forecasting method approaches such as linear regression, multilayer perceptron, and sequential minimal optimization regression (SMOreg). From those three approaches, there was calculation toward forecasting result obtained, and there were calculation 2 kinds of evaluation index which were MAE (mean absolute error) and RMSE (root mean squared error). Here it is some theories related the three methods used.

3.1 Linear Regression

Linear regression is a particular method used to calculate forecasting time series data in a particular period. Regression analysis can be separated into two parts; those are linear regression and nonlinear regression. Linear regression is divided into two; those are linear regression with one variable and linear regression with multivariable.

3.1.1 Linear Regression with Single Variable

This model is the simplest model to investigate the correlation between independent variable x and independent variable y . the object relation is the connector between two variables.

3.1.2 Linear Regression with Multivariabel

Basic hypothesis for regression model multivariable lays beyond five hypotheses from one regression model variable, and the demand of all independent variables is linear with the others.

The value can be jumbled in n sample for estimation parameter [12]. The equation can be seen from the calculation below:

$$y_i = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k + u_i \quad i=1,2,\dots,n$$

.....(1)

3.2 Multilayer Perceptron

Multilayer perceptron (MLP) is one of popular network techniques. MLP can be used to investigate classification problem, construction model, forecasting time series data, and discrete control [13]. This neuron network model is developed by adding hidden layer for simple perceptron. Every layer employs some neurons, and every neuron on the layer is connected to the layer close to one another with different load [14].

For forecasting subject matter, a back propagation algorithm (BP) is used to conduct training < LP [15]. This algorithm gives numerical solution for the problem by minimizing nonlinear function extending parameter capacity of the function.

3.3 Sequential Minimal Optimization Regression (SMOreg)

SMOreg is an algorithm based on support vector machine for the solution of regression problem [11]. SMOreg is used to conduct the calculation of forecasting time series data and calculation of linear regression model. Regression analysis is a statistical mathematical model used to take care the correlation among variables which can give solution to prediction problem and forecasting problem. SVR (Support vector regression) is an algorithm SVM (support vector machine) [12].

SVM (support vector machine) is an algorithm which is closely related to regression function problem to predict, and it is able to used in nonlinear regression [17].

4. RESEARCH METHOD

4.1 Pre-processing Step

This research was started by collecting dengue fever occurrence cases data in Yogyakarta province through Health Department. Collected data was in the form of time series data about dengue fever occurrence cases in 5 Districts in Yogyakarta Province from 1997 until 2006. Five attributes used in the forecasting were the number of dengue fever patients in every district, the number of died dengue fever patients, incident rate (IR), and case fatality rate (CFR). Based on data for 1997-2006 there were 120 datasets such occurrences month for each districts.

The dataset consists of a number of 120 dengue incidence data from January 1997-December 2005 a number of 108 datasets, while testing the data taken from January 2006-December 2006, a total of 12 datasets. Illustration of the use of training data and testing data is in the period January 1997 to December 2006 is shown in Figure 1.

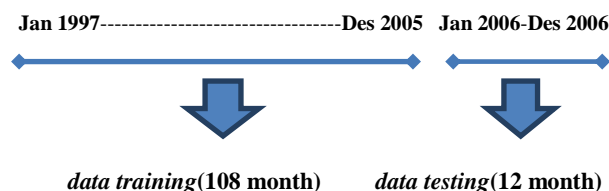


Figure 1. Time Series Illustration of Training and Testing Data

Forecasting technique was done by using 3 methods; those were linear regression, multilayer perceptron, and sequential minimal optimization regression (SMOreg). Each method is tested with the same parameter which was time unit to forecast with 12 numbers, time stamp in the form of date, periodicity automatically and confidence intervals in the rate

of 95%. Forecasting attribute was the number of dengue fever in January 2006 until December 2006.

The next step is test the training data using the three algorithms, namely linear regression, multilayer perceptron and SMOreg. To calculate comparison method by looking at the value of accuracy, MAE (mean absolute error) and RMSE (root mean squared error). The comparison then choose the method with best results.

Selected methods were then tested using the data of testing 5 districts, and then compared with the data returned by the rules of grouping endemicity of the Ministry of Health for the data Yogyakarta Province. Research steps can be seen in Figure 2.

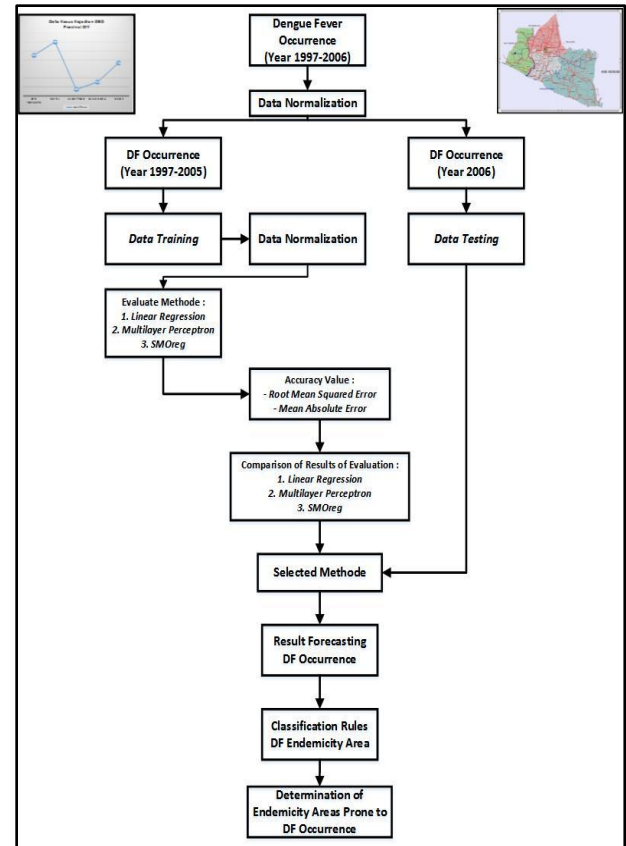


Figure 2. Framework Forecasting of Dengue Fever Area Endemic Model

4.2 Data Normalization Step

Data normalization was done to keep the data range is not too far between the others, so the data will not be that far. Normalization is needed to achieve scale rate of attribute value from the data, so the data can be placed in a particular range. The technique used in normalization was conducted with softmax which becomes the development of linear transformation, so the range output was 0-1. The equation of that is as follow:

$$r_{ij} = x_{ij} / \sqrt{\sum x_{ij}^2} \dots\dots\dots(2)$$

In which :

r_{ij} = new data

x_{ij} = record data

5. RESULT AND DISCUSSION

In this section will discuss the step of data forecasting, classification of endemicity areas, testing models, and comparison of accuracy.

5.1 Forecasting Data Step

Forecasting data step in this research is by creating data forecasting model. Data forecasting conducted in this research used Weka application 3.7.11 by adding plug in in the form of forecast. The number of data set in this research was 120 data with 108 months used as training data and 12 months as testing data. Four predictor attributes used in forecasting data was the number of dengue fever cases, the number of died patients, IR, and CFR. The presentation of the data can be seen in the table 3.

Table 3. Occurrences Cases of Dengue Fever from 1997 to 2005 in Yogyakarta Province

Year	Dengue Fever			
	Patient	Death	IR	CFR
1997	1190	42	0,34	3,52
1998	3631	78	1,102	2,14
1999	508	8	0,15	1,57
2000	917	10	0,28	1,09
2001	1019	10	30,95	0,98
2002	860	12	26,08	1,4
2003	1531	35	46,42	2,29
2004	2206	31	66,29	1,41
2005	971	12	29,18	1,24
2006	?	?	?	?

5.2 Classification of Endemicity Area

Classification of endemicity area made with reference to the rules laid down by the Ministry of Health, which consists of four types, namely the region endemic, sporadic region, the potential region and region free.

Classification formed four classes of K1, K2, K3 and K4. Determination of endemicity area classification is done by counting the number of instances of each month in each province DIY terditi of 5 districts of the city of Yogyakarta, Bantul, Kulon Progo, Gunungkidul and Sleman.

After forecasting incidence of dengue cases in each district and then seeing the observed period. Next determine endemicity area by observing the incidence of DHF each region which have a sequence of instances between January to December in particular with reference to the time series data on the previous year.

Table 4 shows the results of grouping according to the rules of dengue endemic area of endemicity of the Ministry of Health with real data in the period January 2006 to December 2006. Based on the classification of each district of Yogyakarta, Bantul, Gunung Progo and Sleman at the level of K1 that is endemic areas because of the 3-month period there is always a case of dengue incidence.

Table 4. Results of Grouping Endemicity Regency in Province based on Real Data 2006

Waktu	Regency Yogyakarta Province				
	Yogyakarta	Bantul	KP	GK	Sleman
01-Jan-06	K1	K1	K1	K1	K1
01-Feb-06					
01-Mar-06					
01-Apr-06	K1	K1	K1	K1	K1
01-May-06					
01-Jun-06					
01-Jul-06	K1	K1	K1	K1	K1
01-Aug-06					
01-Sep-06					
01-Oct-06	K1	K1	K1	K1	K1
01-Nov-06					
01-Dec-06					

*KP = Kulonprogo, GK = Gunungkidul

Training data was tested using three algorithms (linear regression, multilayerperceptron and SMOREG) with a time stamp in the form of date, periodicity is automatically and confidence intervals at a rate of 95% to obtain the data forecasting results of each algorithm. Then the next step is compare the accuracy three algorithms.

5.3 Testing Model

This section discusses the testing of models created based on the selected algorithm in subsection 5.2. A total of 108 training data for each county in which the municipality of Yogyakarta Special Region of Yogyakarta, Bantul, Kulon Progo, Gunungkidul and Sleman then testing the data 12 to be tested using the selected algorithm.

Testing model is used to predict incidence of DF cases with four attributes, namely the incidence of dengue cases, the rate of death, incident rate and case fatality rate. Every district in the province of Yogyakarta tested with linear regression forecasting method for the period starting in January 1997 to December 2005. Then in 2006 is used to test accuracy of data pattern in the previous year. For example, to test one districts of the city of Yogyakarta incidence of dengue cases are as follows:

Table 5. Data from Real incidence of dengue cases in Yogyakarta Municipality January 2006-December 2006

Year	Attribute			
	Case DF	Death	IR	CFR
1-Jan-06	172	1	34.33	0.58
1-Feb-06	229	1	45.71	0.44
1-Mar-06	247	1	49.31	0.40
1-Apr-06	63	2	12.58	3.17
1-May-06	31	0	6.19	0.00
1-Jun-06	39	0	7.79	0.00
1-Jul-06	27	0	5.39	0.00
1-Aug-06	31	0	6.19	0.00
1-Sep-06	14	0	2.79	0.00
1-Oct-06	14	0	2.79	0.00
1-Nov-06	11	0	2.20	0.00
1-Dec-06	10	0	2.00	0.00

From data incidence of dengue cases in real district of Yogyakarta city in the period January 2006-December 2006 is compared with data of Yogyakarta city district forecasting results in Table 6.

Table 6. Data from Yogyakarta Municipality Forecast January 2006-December 2006

Year	Attribute			
	Case DF	Death	IR	CFR
1-Jan-06	157.75	0.58	31.49	0.37
1-Feb-06	78.06	0.13	15.58	0.16
1-Mar-06	30.90	1.12	6.17	3.61
1-Apr-06	82.33	0.67	16.43	0.81
1-May-06	98.32	1.10	19.63	1.12
1-Jun-06	66.24	1.45	13.22	2.20
1-Jul-06	34.17	1.16	6.82	3.39
1-Aug-06	14.68	2.75	2.93	18.71
1-Sep-06	28.11	0.96	5.61	3.42
1-Oct-06	39.72	0.33	7.93	0.83
1-Nov-06	25.99	2.82	5.19	10.86
1-Dec-06	22.45	5.62	4.48	25.03

Forecasting results above were tested on four other districts namely Bantul, Kulon Progo, Gunungkidul and Sleman the same period, on January 2006 until December 2006 by which it can be seen in Table 7.

Table 7. Number of Cases of dengue incidence DIY Province in 2006 with Time Series Data Forecasting Model

Year	Regency Yogyakarta Province				
	Yogya	Bantul	KP	GK	Sleman
1-Jan-06	157.75	190.10	21.26	25.84	196.92
1-Feb-06	78.06	145.67	38.44	20.35	203.82
1-Mar-06	30.90	131.76	52.25	16.59	211.09
1-Apr-06	82.33	129.67	116.04	6.24	182.11
1-May-06	98.32	97.23	129.31	11.72	151.94
1-Jun-06	66.24	42.98	130.04	22.35	103.32
1-Jul-06	34.17	28.26	130.49	15.36	80.78
1-Aug-06	14.68	27.81	129.67	7.23	60.25
1-Sep-06	28.11	31.17	129.76	15.43	59.97
1-Oct-06	39.72	34.44	134.13	2.68	73.42
1-Nov-06	25.99	36.80	142.03	3.25	59.52
1-Dec-06	22.45	61.70	150.57	3.08	65.43

*KP = Kulonprogo, GK = Gunungkidul

After getting the data forecasting the incidence of cases of dengue endemicity classification results are obtained on the territory of each district are shown in Table 8.

Thus it can be concluded that there is incidence of dengue cases in each month between January 2006 and December 2006 so that in each district in the province of Yogyakarta has a tendency endemic region who are in K1. (K1 = Always incidence of dengue cases for 3 consecutive months). In addition, in October, November, December, January, February and March are the months that have cases of dengue incidence is high.

Table 8. Classification endemicity level of regency in Yogyakarta province in year 2006

Waktu	Regency Yogyakarta Province				
	Yogyakarta	Bantul	KP	GK	Sleman
01-Jan-06					
01-Feb-06					
01-Mar-06	K1	K1	K1	K1	K1
01-Apr-06					
01-May-06					
01-Jun-06	K1	K1	K1	K1	K1
01-Jul-06					
01-Aug-06					
01-Sep-06	K1	K1	K1	K1	K1
01-Oct-06					
01-Nov-06					
01-Dec-06	K1	K1	K1	K1	K1

*KP = Kulonprogo, GK = Gunungkidul

5.4 Comparison of Accuracy

Test results then generate forecasting value (forecast) which is then compared to the value of real events (real) in the next year by looking at the percentage of the highest value. In Table 9 are presented concerning accuracy of the comparison between three methods used.

Table 9. Comparison of Accuracy Linear Regression Testing, Multilayer Perceptron and SMOREg

Algorithm	Linear Regression	Multilayer Perceptron	SMOREg
Accuracy Forecasting (%)	80,11	46,39	72,67
Mean Absolute Error	0.0212	0.0137	0.0271
Root Mean Squared Error	0.0453	0.0227	0.0523

In Table 9 shows that the MAE and RMSE value of the smallest to the largest is the method Multilayer Perceptron then Linear Regression and SMOreg. The smaller the value MAE and RMSE, the better the performance of the algorithm.

Results forecasting time series data that comes closest is the Linear Regression method with the data trend forecasting generate similarity rate of 80.11% followed Sequential Minimal Optimization methods for Regression (SMOreg) with a percentage of 72.67% and the third is the method of Multilayer perceptron with a percentage of 46.39%. Linear Regression method has smaller MAE and RMSE value than SMOreg and Multilayer Perceptron. Therefore, it can be concluded that Linear Regression < SMOreg < Multilayer Perceptron.

6. CONCLUSION

Based on the first study conducted, by comparing the three algorithms, linear regression algorithm has a higher accuracy rate for the forecasting data model than Multilayer Perceptron algorithm and the Sequential Minimal Optimization regression (SMOreg) which is equal to 80.11%.

DBD forecasting data is based on four attributes of items, namely the number of cases of Patients, mortality, incident rate and case fatality rate. Data forecasting results are grouped into regions of dengue endemicity rules referenced from the Ministry of Health items, namely (endemic, sporadic, and potential-free areas) to Determine the which ones have the potential areas endemic for dengue outbreak.

Furthermore, this region of endemicity classification model can be used as an alternative for grouping the endemicity of an area based on four attributes mentioned above.

This model is able to classify the degree of endemicity areas based on time series data patterns obtained from incident cases in the previous year. In addition, this model can also be used as a supporting epidemiological analysis of infectious disease caused by the dengue disease with data mining models approach, so that the process of determining the endemic regions can be maximized.

Apart from the above conclusion determination dengue endemic area can be used as a reference to the local Health Department to be able to anticipate in each region and in the months of October, November, December, January, February and March in order to increase awareness of the existence an outbreak of dengue fever. To improve the forecasting accuracy data occurrence of dengue cases area are need additional predictor variables related with the environment and other predictor variables as well as the density of *Aedes aegypti*.

7. REFERENCES

- [1] Kuhn L, Davidson LL, Durkin MS. (1994). Use of poisson regression and time series analysis for detecting changes over time in rates of child injury following a prevention program. *Am J Epidemiol*. 140(10):943–955.
- [2] Diana, N., Tarmizi, A., Jamaluddin, F., Bakar, A. A., Othman, Z. A., Zainudin, S., & Hamdan, A. R. (2013). Malaysia Dengue Outbreak Detection Using Data Mining Models, 4(August), 96–107.
- [3] J. Han and M. Kamber. (2006). Data mining: concepts and techniques. 2nd Edition. Morgan Kauffman.
- [4] World Health Organization (WHO). (2002). Dengue and Dengue Haemorrhagic Fever. Fact Sheet No. 117. Geneva: World Health Organization.
- [5] Kouri G. Dengue, a growing problem of health in the Americas. (2006). *Rev Panam Salud Publica*. pp:143–145.
- [6] World Health Organization (WHO). (2007). Available: <http://www.who.int/csr/disease/dengue/impact/en/index.html>. Accessed 2014 Nov 09.
- [7] Chen, C.-C., & Chang, H.-C. (2013). Predicting dengue outbreaks using approximate entropy algorithm and pattern recognition. *The Journal of Infection*, 67(1), 65–71. doi:10.1016/j.jinf.2013.03.012
- [8] World Health Organization (WHO). (2013) Comprehensive Guidelines for Prevention and Control of Dengue Haemorrhagic Fever. Regional Office for South East Asia.
- [9] Achmadi U.F., (2010). Demam Berdarah Dengue di Indonesia Tahun 1968-2009, *Buletin Jendela Epidemiologi Vol-2*, pp. 1-13.
- [10] <http://www.jogjapro.go.id>
- [11] Wang, L., Tan, L., Yu, C., & Wu, Z. (2012). Study and Application of Non-Linear Time Series Prediction in Ground Source Heat Pump System, (19005114009), 3522–3525.
- [12] Yang, M., Chen, Y., & Li, T. (2009). Application Research of Road Passenger Transport Volume Forecasting Using Regression Methods. 2009 *Second International Symposium on Knowledge Acquisition and Modeling*, (1), 277–280. doi:10.1109/KAM.2009.73
- [13] D. Hush, C. Abdallah, B. Hore. (1990). Model following using multilayer perceptrons Decision and Control”, *Proceedings of the 29th IEEE Conference*, pp. 1730–1731.
- [14] Member, S. V. I., Aparicio, J., & Senabre, C. (2010). Comparative Analysis of Self Organizing Maps vs . Multilayer Perceptron Neural Networks for Short - Term Load Forecasting, (1).
- [15] M. Kazeminejad, M. Dehghan, M.B. Motamadinejad, H. Rastegar. (2006). A New Short-Term Load Forecasting Using Multilayer Perceptron Information and Automation”. *ICIA 2006. International Conference*, pp. 284–288.
- [16] Kenneth Levenberg. (1944). A Method for the Solution of Certain Non- Linear Problems in Least Squares. *The Quarterly of Applied Mathematics*, 2, pp. 164–168.
- [17] Jinfang Yang, Yongjie Zhai, Daping Xu, et al. (2007). SMO Algorithm applied in time series model building and forecast[C]. *Hong Kong International Conference on Machine Learning and Cybernetics*, pp. 2395–2400.