

Neural Network for Earthquake Prediction Based on Automatic Clustering in Indonesia

Mohammad Nur Shodiq[#], Dedy Hidayat Kusuma[#], Mirza Ghulam Rifqi[#],

Ali Ridho Barakbah^{*}, Tri Harsono^{*}

[#] Department of Informatics, State Polytechnic of Banyuwangi, Jember Street KM13 Kabat, Banyuwangi 68461, East Java, Indonesia

^{*} Electronic Engineering Polytechnic Institute of Surabaya, Kampus ITS Sukolilo, Surabaya 60111, East Java, Indonesia

E-mail: noer.shodiq@poliwangi.ac.id, dedy@poliwangi.ac.id, mirza@poliwangi.ac.id, ridho@pens.ac.id, trison@pens.ac.id

Abstract— A model of artificial neural networks (ANNs) is presented in this paper to predict aftershock during the next five days after an earthquake occurrence in selected cluster of Indonesia with magnitude equal or larger than given threshold. The data were obtained from Indonesian Agency for Meteorological, Climatological and Geophysics (BMKG) and United States Geological Survey's (USGS). Six cluster was an optimal number of cluster base-on cluster analysis implementing Valley Tracing and Hill Climbing algorithm, while Hierarchical K-means was applied for datasets clustering. A quality evaluation was then conducted to measure the proposed model performance for two different thresholds. The experimental result shows that the model gave better performance for predicting an aftershock occurrence that equal or larger than 6 Richter's scale magnitudes.

Keywords— Artificial neural networks, earthquake prediction, cluster analysis

I. INTRODUCTION

Indonesia has a high risk to geological disasters as it lies between three active plates of the world: the Eurasian, Indo-Australian and Pacific plates. Therefore, a system that is able to predict the earthquake will greatly help to minimize the risk of losses that arise. A series of earthquakes are not randomly formed, but follow a spatial pattern with a trigger that results in an earthquake event. In other words, natural disasters (e.g., earthquakes) rarely appear on their own, but instead they tend to form a complex network of interacting faults as in [1-3]. However, an earthquake prediction should state where, when, how big, and how probable the predicted event is [4]. Several studies applied artificial neural networks (ANN) for earthquake prediction as in [5-7].

Seismic clustering is the first stage to analyze earthquake risk which based on a variety of seismology criteria. By using a clustering scheme, it is possible to retrieve spatio-temporal pattern that created by events. Although the automatic identification of the optimal number of clusters on seismic data is a very difficult problem in the process of data clustering [8], but the optimal number of clusters can be determined by measuring variance within and variance between each cluster, this is known as cluster analysis [9-10]. The data used are seismic data of all regions of Indonesia obtained from BMKG and USGS in year 1910 to 2017. This

data is filtered using magnitude of completeness of 5.1 Richter magnitude scale [11].

This study proposed an ANN model to predict an earthquake during the next five days after an earthquake occurrence with magnitude equal or larger than given threshold for a selected cluster in all region of Indonesia. Aftershock prediction can be used by the authorities to deploy precautionary policies.

II. RESEARCH METHOD

This section describes the system design of earthquake prediction. Figure 1 illustrates the overall process.

A. Data Acquisition

The earthquake data covers the entire territory of Indonesia, which is the boundary 5.98 north latitude - 11 south latitude and east longitude 94.12 - 140.98 east longitude. The earthquake dataset in this research is taken from catalogue of Agency Meteorology, Climatology and Geophysics (BMKG) and United States Geological Survey (USGS) from 1910 to 2017. The dataset consists of 82,580 seismic events with magnitude 1 - 10 ML, and depth of 0-650 km.

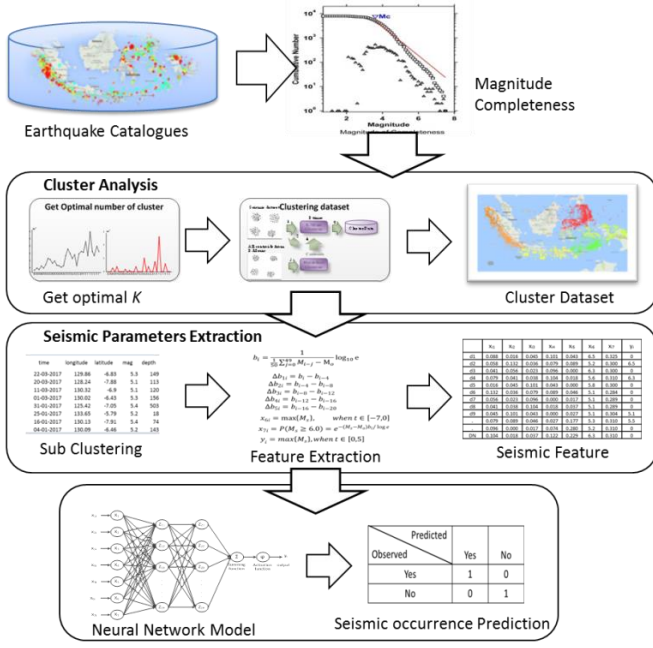


Figure 1 Research approach for earthquake prediction

The magnitude of completeness (MC) of the catalogue for earthquake dataset is determined first before implements clustering process. It uses the Gutenberg-Richter frequency size distribution [12]. Magnitude of completeness also called as threshold or cut-off magnitude. Seismic events with magnitude under the MC will be eliminated [11].

Incompleteness of seismic data will give the result in seismic risk parameters resulting into overestimated or underestimated. In this research, we use magnitude 5.1 ML for magnitude of completeness (MC) according to [11]. So, total number of seismic data after implementing magnitude of completeness is 9,233 events.

B. Cluster Analysis

The main objective of cluster analysis is to generate groups of data from a large dataset in to small group with the same characteristics. Good clustering results will produce an accurate representation of the behaviour of a system [8]. This concept is to extract the most similar (or dissimilar) separated sets of objects according to a given similarity (dissimilarity) measure [12]. Thus, this algorithm is the basis for extracting the dataset to obtain useful information for finding pattern recognition in the data [8].

Clustering process is divided into 2 schemes. The first process is to find the number of cluster. The number of clusters are the amount of cluster to group large data into small groups. There are many algorithms to determine the number of cluster. The easiest method to find the number of cluster is to generate randomly. But, this method does not guarantee to find the global optimal. Choosing a large number of clusters does not necessarily imply better for clustering. On the contrary, results could be less accurate representation. Therefore, the optimal number of clusters has to be determined. Since the number of clusters selected to cluster dataset is one of the most critical decisions in clustering techniques, several approaches have been developed in order to determine this number [11]. In this paper, Valley Tracing and Hill Climbing algorithm are

implemented to select the optimal number of clusters [9]. It analyzes the moving variance of clusters for each stage of cluster construction and observes the pattern to find the global optimum as well as to avoid the local optima.

The algorithm to find the optimal number of clusters [9] is described as follows:

1. Set as each data of A, where A is attribute of n-dimensional vector.
2. Set K as the predefined number of clusters $K_i \{k_i | i = 2, \dots, n-1\}$
3. Apply clustering algorithm with number of clusters K_i
4. Record the calculation of variance as $V_i \{v_i | i = 1, \dots, n-2\}$
5. Increment $j=j+1$
6. Repeat from step 3 while $j < n-2$.
7. Calculate the moving variance with valley tracing -hill climbing $\partial \{ \partial_i | i = 1, \dots, n-1 \}$
8. Put threshold λ on moving variance ∂ for automatic clustering
9. Rank the moving variance $\max(\partial) \geq \lambda$, and i as optimal number of cluster

The second process is to cluster the dataset. into a particular cluster based on the most similar. Before running this process, it requires initial cluster. Therefore, the number of clusters obtained from the first process, becomes the initial clusters for clustering dataset. In this paper, data clustering uses Hierarchical K-means clustering. It combines two clustering algorithms: Hierarchical algorithm and K-means algorithm [13].

The Hierarchical K-means clustering algorithm is as follows [13]:

1. Set as each data of A, where A is attribute of n-dimensional vector.
2. Set K as the predefined number of clusters.
3. Determine p as numbers of computation
4. Set $i=1$ as initial counter
5. Apply K-means algorithm.
6. Record the centroids of clustering results as $C_i \{c_{ij} | j = 1, \dots, K\}$
7. Increment $i=i+1$
8. Repeat from step 5 while $i < p$.
9. Assume $C_i \{c_i | i = 1, \dots, p\}$ as new dataset, with K as predefined clusters
10. Apply hierarchical algorithm
11. Record the centroids of clustering result as $D_i \{d_i | i = 1, \dots, K\}$
12. Then, $D_i \{d_i | i = 1, \dots, K\}$ as initial cluster centers for K-means clustering.

C. Proposed Neural Network

The architecture of the neural network model for predicting the earthquake occurrence is shown in Figure 2.. The input neuron layer has seven nodes representing the seven seismicity indicators. They are used to recognize the pattern of earthquake occurrences. It can also be employed for predicting the nature of the future events and as well as mitigation of earthquake risk [11], [14]. The seven seismicity indicators are shown in Table 1.

To find seismic parameters, firstly, the earthquake data should be cluster into small data groups based on the optimal number of cluster. Then, distribution of data is utilized to investigate the seismic parameters for each clusters, for example, b value.

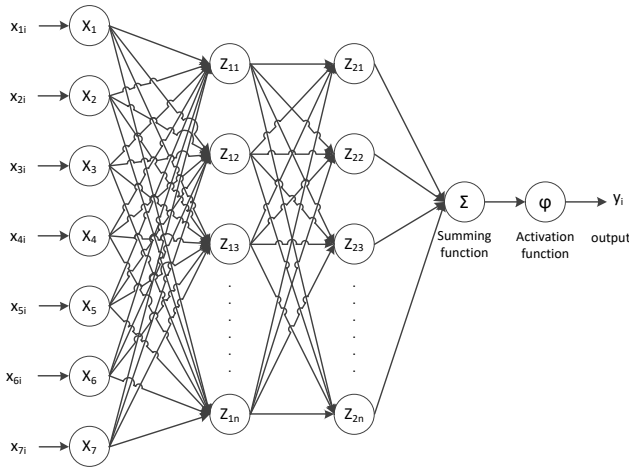


Figure 2 Architecture of the proposed neural network

Table 1 The seven seismicity indicators

No.	Notation	Description
1	x_{1i}	$b_i - b_{i-4}$
2	x_{2i}	$b_{i-4} - b_{i-8}$
3	x_{3i}	$b_{i-8} - b_{i-12}$
4	x_{4i}	$b_{i-12} - b_{i-16}$
5	x_{5i}	$b_{i-16} - b_{i-20}$
6	x_{6i}	Maximum magnitude during the last week
7	x_{7i}	Dynamic GR's law

Five input parameters (b -value) were obtained from the Gutenberg-Richter law's. The b parameter characterizes the relative size of distribution of events and it depends on the stress regime and seism tectonic behaviour of the region. In this paper, it is calculated using the last 50 recorded quakes[15]. The b parameter value is determined as:

$$b_i = \frac{1}{\frac{1}{50} \sum_{j=0}^{49} M_{i-j} - M_o} \log_{10} e \quad (1)$$

Where M_i is the magnitude for the i -th earthquake in the clustering dataset. M_o is magnitude completeness or cut off magnitude.

Then, increments of b are calculated:

$$\Delta b_{1i} = b_i - b_{i-4} \quad (2)$$

$$\Delta b_{2i} = b_{i-4} - b_{i-8} \quad (3)$$

$$\Delta b_{3i} = b_{i-8} - b_{i-12} \quad (4)$$

$$\Delta b_{4i} = b_{i-12} - b_{i-16} \quad (5)$$

$$\Delta b_{5i} = b_{i-16} - b_{i-20} \quad (6)$$

From these equations, it can be concluded that 70 earthquakes are required to calculate the x_i values, therefore, to obtain the five of the ANN inputs.

The sixth input variable x_{6i} is the maximum magnitude M_s from the recorded quakes during the last week in the analyzed area [15]. It is defined as:

$$x_{6i} = \max\{M_s\}, \text{ when } t \in [-7, 0] \quad (7)$$

where the time t is measured in days.

The seventh input variable x_{7i} identifies the probability of recording an earthquake with magnitude larger or equal to M_s . This input is to include Gutenberg–Richter's law in a dynamic way which is calculated from the probability density function (PDF) [15]. For example, this study uses magnitude M_s with 6.0 M_L so the parameter x_{7i} is defined as:

$$x_{7i} = P(M_s \geq 6.0) = e^{-(M_s - M_o)b_i / \log e} \quad (8)$$

The output parameter has one variable y_i . y_i parameter is the maximum magnitude M_s observed in the cell under analysis during the next five days. It has been set to 0 for such situations where no earthquake with magnitude equal or greater to M_s . In this scenario, M_s is set to 6.0 M_L .

$$y_i = \max\{M_s\}, \text{ where } t \in [0, 5] \quad (9)$$

where the time t is measured in days.

The training vector associated to the i -th earthquake can be expressed as:

$$T_i = \{x_{1i}, x_{2i}, x_{3i}, x_{4i}, x_{5i}, x_{6i}, x_{7i}, y_i\} \quad (10)$$

The input vector (x) has seven elements, and the output vector (y) has a single element.

The number of hidden layers and the number of nodes in each hidden layer are determined by trial and error for best results. First, one hidden layer and several nodes are arranged. The number of nodes is increased gradually to see the pattern of the movement of the result. When the number of nodes is increased but no improvement in its value, the experiment is stopped and the best result is taken.

The activation function in this paper uses sigmoid activation function. It is defined as:

$$\phi(x_i) = \frac{1}{(1 + e^{-x_i})} \quad (11)$$

where,

$$x_i = \phi(\sum_i w_{ij} x_j) \quad (12)$$

where, w_{ij} are the connection weights between unit i and unit j .

The learning method in this paper is back propagation. It consists of two main phases: propagation and weight update.

D. Performance evaluation

In this prediction, an earthquake is said to occur during the next five days when its magnitude reaches the specified threshold (M_s). Several parameters used to evaluate the performance of system [2] are:

1. *True positives (TP)*. The number of times that the ANN predicted an earthquake and an earthquake did occur during the next five days.

2. *True negatives (TN)*. The number of times that the ANN did not predict an earthquake and no earthquake occurred during the next five days.
3. *False positives (FP)* or false alarm. The number of times that the ANN predicted an earthquake but no earthquake occurred during five days.
4. *False negatives (FN)*. The number of times that the ANN did not predict an earthquake but an earthquake did occur during the next five days.
5. *Negative predictive value (NPV)* denoted as P_0 . The proportion of negative result, calculated as follow:

$$P_0 = \frac{TN}{TN + FN} \quad (13)$$

6. *Positive predictive value (PPV)*, denoted as P_1 . The proportion of positive result, calculated as follow:

$$P_1 = \frac{TP}{TP + FP} \quad (14)$$

7. *Sensitivity (S_n)*. The proportion of positives that are correctly identified, calculated as follow:

$$S_n = \frac{TP}{TP + FN} \quad (15)$$

8. *Specificity (S_p)*. The proportion of negatives that are correctly identified, calculated as follow:

$$S_p = \frac{TN}{TN + FP} \quad (16)$$

III. RESULTS AND DISCUSSION

In this section, it is explained the results of research and at the same time is given the comprehensive discussion.

A. Dataset Clustering

Grouping dataset in to small group is well known as clustering. There are three steps in this stage. First step is to get optimal number of cluster, while the second step is data clustering process, and the last step is visualizing on map.

Earthquake dataset in this paper is clustered using epicentre parameter, that consists of longitude and latitude. The optimal number of cluster is obtained by applying Valley Tracing and Hill Climbing algorithm. Six clusters are the optimal number resulted from these processes. It becomes an initial number of clusters for dataset clustering. Afterwards, the second step is data clustering process, it uses Hierarchical K-means clustering algorithm.

To visualize the result of clustering process, it uses PHP/Java Bridge tools and Google map on Indonesia map [16]. Every cluster has different colours in order to differ one cluster to another. Visualization of spatial data distribution based on clustering can be seen in Figure 3. A number of dataset of each cluster is shown by Table 2, while the Indonesian territory belonging to each cluster can be seen in Table 3.

Table 2 The amount of data member of clusters

Cluster	Dataset member of each cluster	Average magnitude	Maximum magnitude
Cluster0	3006	5.47	8.3
Cluster1	669	5.54	9
Cluster2	1178	5.47	8.5
Cluster3	1112	5.46	7.8
Cluster4	1597	5.51	8.2
Cluster5	1671	5.46	8.5

Table 3 Dataset Member of each cluster based on region

No.	Cluster	Region	No.	Cluster	Region
1	cluster0	Borneo, Halmahera, Minahasa, Sulawesi	4	cluster3	Bali region, Flores region, Java Sea, Lombok region, south of Sumbawa, Sumbawa region, Timor Region
2	cluster1	Southwest of Sumatra, Mentawai region, Nias region, northern Sumatra	5	cluster4	Banda Sea, Biak region, Ceram Sea, Halmahera, Papua Region, Sulawesi near Papua
3	cluster2	Borneo, Java Sea, Mentawai region, southern Sumatra, southwest of Sumatra, Sunda Strait	6	cluster5	Banda Sea, East Timor region, Flores region, Timor Sea, Sulawesi, Tanimbar islands, Kai islands

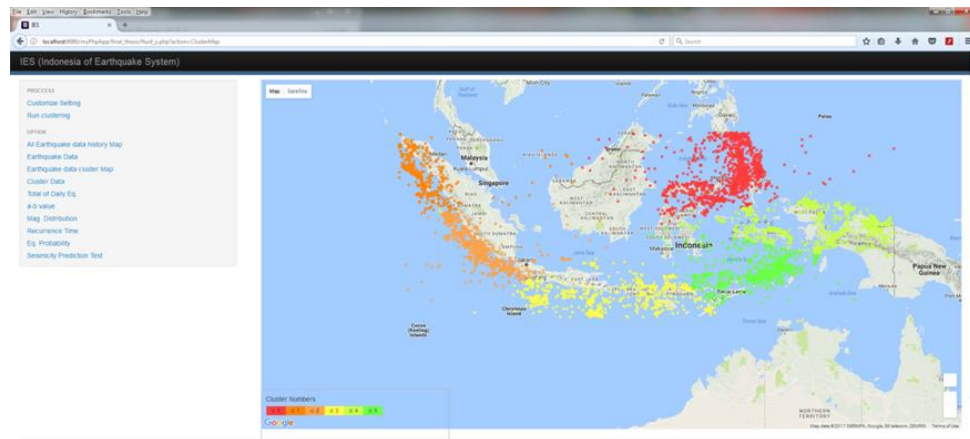


Figure 3 Magnitude distribution based on cluster from 1910 to 2017 with Magnitude completeness MC 5.1 ML, and total number of earthquake 9,233 events

B. Seismic Parameters Extraction

The earthquake catalogue used in this paper has been obtained from the BMKG and USGS Services. Before analyzing earthquake parameters, the completeness of the data each area should be analyzed first. Year of completeness of seismic catalogue can be graphically calculated. On the graph, the time is inversely plot to the x-axis and the accumulated number of quakes on the y-axis.

The slope of the curve, which represents the annual rate in the data, has a straight line on the curve showing that at that time span is constant. From Figure 4, the earthquake data is obtained in the estimated timeframe from January 1975 to May 2008 for data training. The rest of data between June 2008 to December 2016 is used as sample testing datasets.

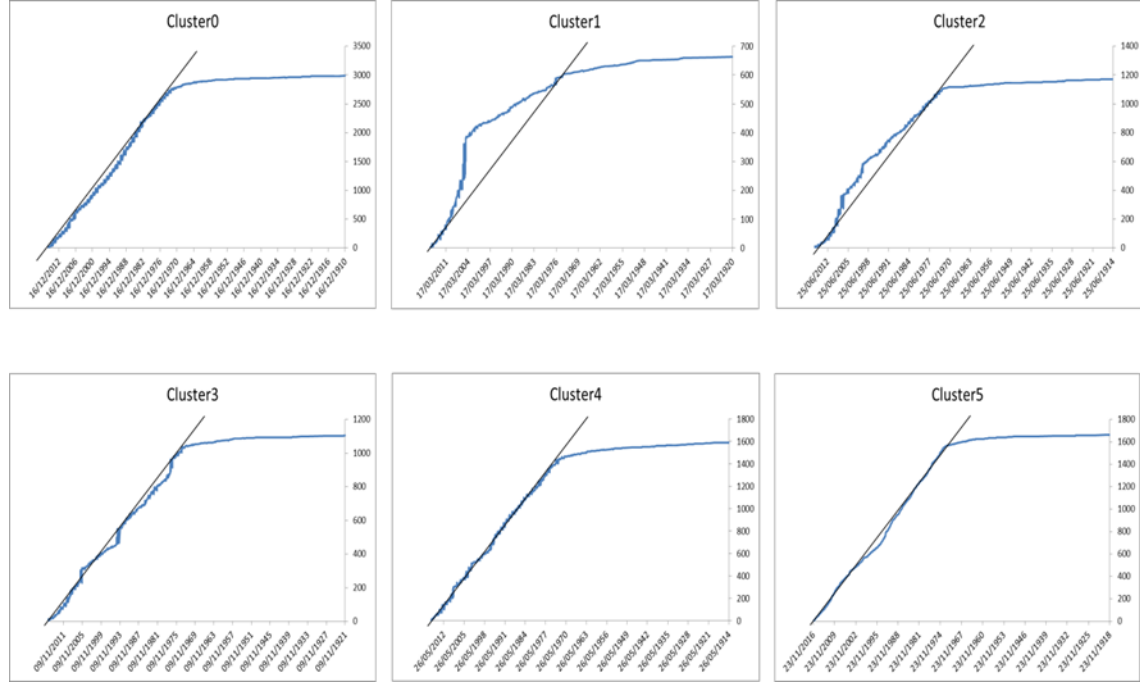


Figure 4 Year of completeness for all studied areas

C. Neural Network Training and Testing

Network training uses earthquake catalogues between January 1975 to May 2008. It consists of 2081 training data sets for cluster0, 369 training datasets for cluster1, and 759, 739, 1009, 1113 training datasets for cluster2, cluster3, cluster4, cluster5. Sample of training datasets for cluster0 are presented in Table 4.

The model was tested using earthquake catalogue between June 2008 and December 2016. It also consists of 489 datasets for cluster0, 152 data sets for cluster1, and datasets

for cluster2, cluster3, cluster4, cluster5 are 237, 199, 307, 318. Sample of testing datasets for cluster0 showing seven-element input and one output parameter are presented in Table 5.

The best result according to trial and error is the ANN model with learning rate sets to 0.1, two hidden layers with thirty-two nodes each other and 10000 iterations.

Table 4 Sample of training datasets in cluster0

No.	b_i	b_{i-4}	b_{i-8}	b_{i-12}	b_{i-16}	b_{i-20}	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}	x_{i6}	x_{i7}	y_i
1	1.658	1.529	1.540	1.467	1.374	1.551	0.128	0.011	0.073	0.093	0.177	5.3	0.217	0
2	1.670	1.519	1.562	1.448	1.551	1.597	0.152	0.044	0.115	0.103	0.046	5.3	0.215	5.3
3	1.645	1.477	1.519	1.438	1.540	1.477	0.168	0.041	0.080	0.102	0.063	5.1	0.220	5.3

Table 5 Sample of testing datasets in cluster0

No.	b_i	b_{i-4}	b_{i-8}	b_{i-12}	b_{i-16}	b_{i-20}	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}	x_{i6}	x_{i7}	y_i
1	1.780	1.670	1.498	1.383	1.187	1.131	0.110	0.173	0.114	0.197	0.056	5.2	0.194	0
2	1.765	1.780	1.457	1.285	1.200	1.102	0.014	0.323	0.172	0.085	0.097	5.2	0.197	5.2
3	1.710	1.795	1.410	1.300	1.200	1.054	0.085	0.385	0.110	0.101	0.146	5.2	0.207	0

Table 6 and Table 7 show the experimental result of testing scenarios of proposed ANN model for quality parameters. In

the first scenario used the input node $x7i = P(Ms \geq 5.5)$ and the second one used the input node $x7i = P(Ms \geq 6)$.

Table 6 ANN performance of proposed model with $x7i = P(Ms \geq 5.5)$

Parameters	Value					
	Cluster0	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
TP	15	15	8	11	13	1
TN	298	84	155	133	201	236
FP	13	20	10	22	7	1
FN	163	33	64	33	86	80
P_0	65%	72%	71%	80%	70%	75%
P_1	54%	43%	44%	33%	65%	50%
S_n	8%	31%	11%	25%	13%	1%
S_p	96%	81%	94%	86%	97%	100%
Average	56%	57%	55%	56%	61%	56%

Table 7 ANN performance of proposed model with $x7i = P(Ms \geq 6)$

Parameters	Value					
	Cluster0	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
TP	2	5	1	5	1	0
TN	424	135	203	182	269	291
FP	2	3	0	4	2	0
FN	61	9	33	8	35	27
P_0	87%	94%	86%	96%	88%	92%
P_1	50%	63%	100%	56%	33%	0
S_n	3%	36%	3%	38%	3%	0%
S_p	100%	98%	100%	98%	99%	100%
Average	60%	72%	72%	72%	56%	0

IV. CONCLUSION

An ANN model has been built to predict an earthquake during the next five days after an earthquake occurrence in a selected cluster. The best result according to trial and error is the ANN model with learning rate sets to 0.1, two hidden layers with thirty-two nodes each other and 10000 iterations. The proposed model gave better performance for magnitude threshold(Ms) ≥ 6 ML.

V. ACKNOWLEDGEMENT

The authors thank to Directorate of Research and Community Service (DRPM) from Indonesian Ministry of Research, Technology and Higher Education for funding this research through college cooperation research (PKPT) scheme.

REFERENCES

- [1] Georgoulas, G., A. Konstantaras, E. Katsifarakis, C.D. Stylios, E. Maravelakis, and G.J.Vachtsevanos. Seismic-mass Density-based Algorithm for Spatio-temporal Clustering. *Expert Systems with Applications*. 2013; 40(10): 4183-4189
- [2] Alvarez, R. J., J. C. Echeverria, A. Ortiz-Cruz, and E. Hernandez. Temporal and Spatial Variations of Seismicity Scaling Behavior in Southern México. *Journal of Geodynamics*. 2012; 54:1-12.
- [3] Ales, S., and Jiri Vanek. Earthquake Clustering in the Tectonic Pattern and Volcanism of the Andaman Sea Region. *Tectonophysics*. 2013; 608: 728-736.
- [4] C.R. Allen, *Responsibilities in Earthquake Prediction*. Bulletin of the Seismological Society of America. 1982; 66(6): 2069–2074.
- [5] E.I. Alves. Earthquake Forecasting using Neural Networks: Results Future Work. *Nonlinear Dynamics*. 2006; 44(1–4): 341–349.
- [6] H. Adelli and K. Panakkat. A Probabilistic Neural Network for Earthquake Magnitude Prediction. *Neural Network*. 2009; 22(7): 1018-1024.
- [7] J. Reyes, A. Morales-Esteban, and F. Martínez-Álvarez. Neural Networks to Predict Earthquakes in Chile. *Applied Soft Computing*. 2013; 13(2): 1314–1328.
- [8] A. Morales-Esteban, F. Martínez-Álvarez, S. Scitovski, R. Scitovski. A Fast Partitioning Algorithm using Adaptive Mahalanobis Clustering with Application to Seismic Zoning. *Computers & Geosciences*. 2014; 73: 132-141.
- [9] A.R. Barakbah and K. Arai. *Determining Constraints of Moving Variance to Find Global Optimum and Make Automatic Clustering*. Industrial Electronics Seminar (IES). Surabaya. 2004: 409-413.
- [10] A.R. Barakbah and K. Arai. Reserved Pattern of Moving Variance for Accelerating Automatic Clustering. *EEPIS Journal*. 2004; 2(9): 15-21.
- [11] M.N. Shodiq, A. R. Barakbah, T. Harsono. *Spatial Analysis of Earthquake Distribution with Automatic Clustering for Prediction of Earthquake Seismicity in Indonesia*. The Fourth Indonesian-Japanese Conference on Knowledge Creation and Intelligent Computing (KCIC). Surabaya/Bali. 2015: 47-55.
- [12] D.A. Yuen, J.K. Benamin, F. B. Evan, W. Dzwinel, A.G. Zachary, and R.S. Cesar. Clustering and Visualization of

- Earthquake Data in a Grid Environment. *Visual Geoscience*. 2005; 10(1): 1-12
- [13] K. Arai and A.R. Barakbah. Hierarchical K-means: an Algorithm for Centroids Initialization for K-means. *Reports of the Faculty of Science and Engineering, Saga University, Japan*. 2007: 36(1).
- [14] A. Zamani and R.M. Sorbi. Application of Neural Network and ANFIS Model for Earthquake Occurrence in Iran. *Earth Science Informatics*. 2013; 6(2): 71-85.
- [15] A. Morales-Esteban, F. Martínez-Álvarez, J. Reyes. Earthquake prediction in seismogenic areas of the Iberian Peninsula based on computational intelligence. *Tectonophysics*. 2013; 593: 121–134
- [16] M.N. Shodiq, A.R. Barakbah, T. Harsono. Cluster Oriented Spatio Temporal Multidimensional Data Visualization of Earthquake in Indonesia. *International Journal of Engineering Technology (EMITTER)*. 2015; 3(1): 53-67.