Automatic Cluster-oriented Seismicity Prediction Analysis of Earthquake Data Distribution in Indonesia

Ali Ridho Barakbah[#], Tri Harsono[#], Amang Sudarsono[*]

[#] Department of Information and Computer Engineering, Politeknik Elektronika Negeri Surabaya, Surabaya 60111, Indonesia
E-mail: ridho@pens.ac.id, trison@pens.ac.id

[*] Department of Electrical Engineering, Politeknik Elektronika Negeri Surabaya, Surabaya 60111, Indonesia
E-mail: amang@pens.ac.id

Abstract— **Many researchers have analyzed the earthquakes to predict the earthquake period occurrences. However, they commonly faced the difficulty to project the prediction into the region adjusted to the earthquake data distribution and to provide an interpretation of the prediction for the region. This paper presents a new system for cluster-oriented seismicity prediction analysis, and semantic interpretation of the prediction result projected to the region. The system applies our automatic clustering algorithm to detect some clusters automatically depending on the earthquake data distribution and create clusters of the earthquake data for the prediction. The semantic interpretation is presented in the system to provide easier information from the seismicity prediction analysis. The system consists of four main computational functions: (1) Data acquisition and pre-processing, (2) Automatic clustering of earthquake data distribution, (3) Seismicity prediction of earthquake time period occurrence based on cluster with confidence levels of seismic event using the Guttenberg-Richter law, and (4) Region-based seismicity prediction analysis and semantic interpretation of the prediction for each cluster. For experiments, we use earthquake data series provided by the Advanced National Seismic System (ANSS) in the year 1963-2015 with the location of Indonesia. We made a series of experiments for earthquakes in Nias (2005), Yogyakarta (2006), and Padang (2009), with respectively 6.3, 7.6 and 8.7 Richter magnitude level. Our system presented the seismicity prediction analysis from each earthquake cluster and provided an easy interpretation of the prediction probability.**

Keywords— **seismicity prediction analysis; earthquake prediction; automatic clustering; semantic interpretation.**

I. INTRODUCTION

Earthquakes are earth events rocked as a result of the release of energy on earth. The accumulation of earthquake energy resulting from the breaking of rock layers in the earth's crust, thus moving tectonic plates. The resulting energy is transmitted in all directions in the form of an earthquake wave so that the effect can be felt down to the surface of the earth. Indonesia is an area prone to earthquakes as it is traversed by the meeting point of three tectonic plates, namely: Indo-Australian Plate, Eurasian Plate, and Pacific Plate. Fig. 1 shows the tectonic plate lane.

The Indo-Australian plate moves relatively north and is infiltrated by the Eurasian plate, while the Pacific plate moves relative to the west. Indonesia is also included in one of the countries whose territory is located on the Ring of Fire. The Ring of Fire is a frequent area of earthquakes and volcanoes that surround the Pacific Ocean basin [1]. The meeting road of these plates is in the waters, so that when a massive earthquake with shallow depths would have the potential to occur tsunami [2]. This condition makes Indonesia vulnerable to the tsunami.
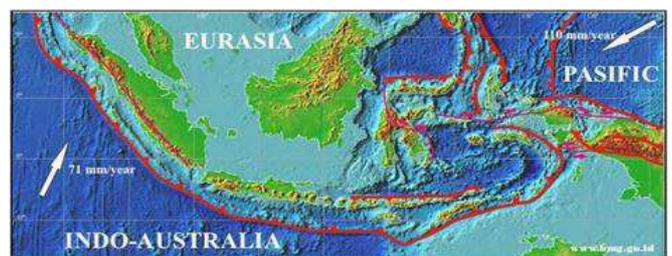

Fig. 1 Tectonic plate lane surrounding Indonesia

During the years 1897 to 2009, there was an earthquake in Indonesia of over 14000 with a magnitude of 5.0 magnitude on the Richter level. With such an earthquake, it has caused thousands of lives, damage and damage to thousands of infrastructure and buildings, and has spent much for rehabilitation and reconstruction [3]. The technical solution becomes very important for earthquake prediction by providing information for the probability of future earthquakes projected to the regions of Indonesia.

Analysis of earthquake distribution is essential, especially in countries that often occur earthquakes. Many researchers studied the field of earthquake. Sadeghian and Jalali-Naini [4] found a new probability density function (PDF) for forecasting the time of earthquake occurrence. Rusnardi et al. [5] constructed area earthquake source model and estimated the frequency magnitude relationship by using the catalogs compiled. Das and Henry [6] examined where aftershocks occur using data from several recent large earthquakes. Irsyam et al. [7] presented the development of spectral hazard maps for Sumatra and Java islands, Indonesia. Fujiwara et al. [8] developed an open web system, includes the hazard map results and data on seismic activity, source models, and underground structure. Faizah et al. [9] have developed the probability of an earthquake in future events using conditional method probability. Moatti et al. [10] have developed pattern recognition on earthquake seismic data with Gutenberg-Richter law for seismicity prediction of earthquakes in the future and obtained the optimal number of clusters with silhouette index. Shodiq et al. [3], [11]–[13] presented the cluster-based earthquake prediction in Indonesia and provided the multi-dimensional data visualization of the earthquake data distribution. These studies provided prediction systems based on the seismic earthquake data. However, they commonly faced the difficulty to project the prediction into the region adjusted to the earthquake data distribution and to provide an interpretation of the prediction for the region.

II. MATERIAL AND METHOD

This paper presents a new system for the analysis of cluster-oriented seismicity prediction and semantic interpretation of predicted results projected into the region. This system applies our automated grouping algorithm to automatically detect the number of groups depending on the distribution of the earthquake data and create the seismic data group for the seismicity prediction. Semantic interpretation is presented in the system to provide easier information from seismicity prediction analysis. The system consists of four main computing functions: (1) Data acquisition and pre-processing, (2) Automatic clustering of earthquake seismic data distribution, (3) Seismicity prediction from earthquake time period occurs based on cluster with seismic event confidence level using Guttenberg-Richter law, and (4) Region-based seismicity prediction analysis and semantic interpretation of the prediction for each cluster. Fig. 2 shows the computational steps of our proposed system to provide seismicity prediction system in Indonesia.

A. Technical Discussion

1) Earthquake Data

This study uses an earthquake data source from ANSS (Advanced National Seismic System) [14]. The earthquake data is obtained based on the minimum limit and the maximum latitude and longitude of Indonesia. Fig. 3 is an earthquake data search with certain latitude and longitude limits and provides some data attributes. This paper uses the attributes of date, latitude, longitude, depth, and magnitude.

Five selected attributes are stored as binary data streams of vector space.
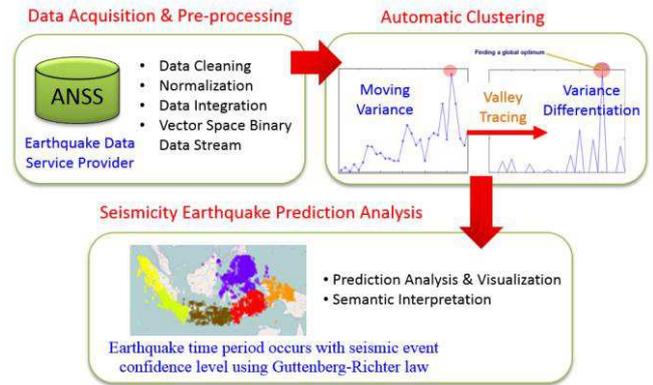

Fig. 2 Our proposed cluster-based seismicity prediction system

Fig. 3 Earthquake data from ANSS

2) Automatic Clustering of Earthquake Seismic Data

Automatic clustering is applied to detect the number of clusters automatically depending on the earthquake data distribution and then create clusters of the earthquake data for the seismicity prediction. In this research, we used Valley Tracing algorithm [15]–[17] for automatic clustering with analyzing the cluster number graphs generated from the clustering process by calculating the variances of each number of clusters created. Variance is commonly used to represent the value of the distribution of the clustering result. Variance is defined in Eq. 1.

$$V = \frac{V_w}{V_b} \tag{1}$$

where:

V_w = variance within cluster
V_b = variance between clusters

V_w expresses the internal homogeneity of the cluster, while V_b expresses the external homogeinity of the clusters. V_w and V_b defined in Eq. 2, Eq. 3 and Eq. 4.

$$v_i^2 = \frac{1}{n_c - 1} \sum_i^{n_c} (d_i - c_i) . v_i^2 \tag{2}$$

$$V_w = \frac{1}{N - k} \sum_{i-1}^k (n_i - 1) . v_i^2 \tag{3}$$

$$V_b = \frac{1}{k - 1} \sum_{i-1}^k n_i (c_i - g)^2 \tag{4}$$

where:

c_i = *centroid of cluster i*
g = *grand mean of data*

After calculating the variance V for each number of clusters, we used this variance according to a series of cluster numbers as a moving variance. Fig. 4 shows the illustration of the earthquake-moving variance in Indonesia from 1960-2012 with magnitude ≥ 6.
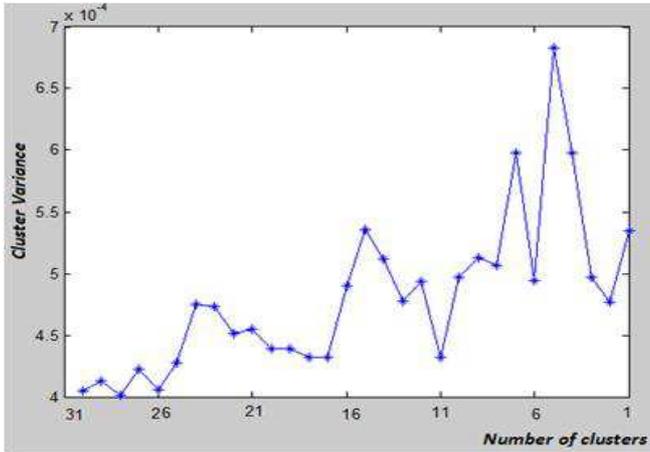

Fig. 4 Moving variance of earthquake data distribution in Indonesia from 1960-2012 with magnitude ≥ 6

The next step in the auto-grouping is to detect the global optimum of the number of clusters. Global optimality is the number of groups to be performed as the optimal number of clusters. To obtain global optimum, we define a series of patterns for the movement of variance from the cluster. Next, we analyze the possibility of optimal global value located in the valley pattern, as shown in Fig. 5.

Fig. 5 Possibility of patterns to be a global optimum

From the pattern analysis in Fig. 5, we can illustrate that the possibility to find the global optimum is on a stage filled with Eq. 5. Fig. 6 shows illustrations of different heights between V_i being met as global optimum.

$$(V_{t-1} \geq V_t) \cap (V_{t+1} > V_t) \tag{5}$$

where:
V_t = variance to t number of clusters, for $t=n..1$, and n is *the* number of clusters when it is same with some data.

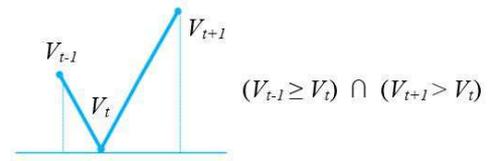$$(V_{t-1} \geq V_t) \cap (V_{t+1} > V_t)$$

Fig. 6. Different value of altitude

We then identify the different altitude values for each stage, as shown in Eq. 6. Fig. 7 shows the differentiation of earthquake data distribution in Indonesia from 1960-2012 with magnitude ≥ 6.

$$\partial_i = (V_{i+1} - V_i) + (V_{i-1} - V_i)$$
$$= (V_{i+1} + V_{i-1}) - (2 \times V_i) \tag{6}$$

The global optimum can be obtained from maximum value of ∂, as shown in Fig. 7. Accuracy (φ) of the valley tracing method can be acquired by defining as follows:

$$\varphi = \frac{\max(\partial)}{\text{closer value to } \max(\partial)} \tag{7}$$
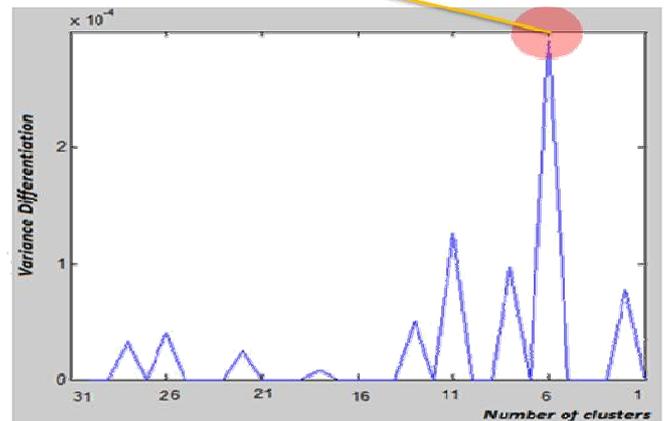
Fig. 7 Variance differentiation of moving variance for earthquake data distribution in Indonesia from 1960-2012 with magnitude ≥ 6
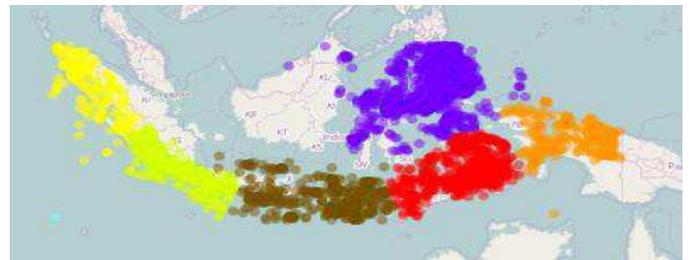
Fig. 8 Visualization of earthquake data distribution after automatic clustering

To obtain a reliable clustering process is having a minimum accuracy φ=2, which means that t number of clusters can be considered as a global optimum if it has a candidate with at least half of its value or smaller. After the automatic clustering process is complete, the clustered

seismic data is drawn and displayed into an Indonesian map, as shown in Fig. 8.

3) Seismicity Prediction

After creating clustered seismic data from automatic clustering, we apply Guttenberg-richter law to calculate seismicity prediction probability of recurring time of the earthquake based on the clustered seismic earthquake. The spatial data distribution based on clusters are utilized to investigate three parameters:

(1) rate of seismic productivity for a given area (*a* value)
(2) relative size distribution of events (*b* value)
(3) recurrence time (*Tr*)

The value can be defined as follows [3], [13], [18].

$$a = \log(N_{M \geq Mmin}) + \log(b \ln(10)) + M_{min} * b \qquad (8)$$

where $N_{M \geq Mmin}$ is the cumulative number of earthquakes with magnitude equal or greater than minimum magnitude M_{min}.

b value can be defined as follows [3,13,18,19,20,21]:

$$b = \frac{1}{M_{avg} - M_{min}} \log_{10} e \qquad (9)$$

where M_{avg} denotes the mean magnitude.

Probabilistic recurrence time for a shock with magnitude equal to or greater than *M*, defined as *Tr*, is calculated follows [3], [5], [18]:

$$Tr = \frac{\Delta T}{10^{(a-bM)}} \qquad (10)$$

where ΔT is the length of the observation period.

The probability value of earthquakes for each cluster at a certain time has different results; it is based on historical seismic data that has occurred. The formulation that can be used to calculate the probability of earthquake prediction using the input parameters in the form of magnitude *M* and the period of event *T* as below.

$$P(M,T) = 1 - e^{-N(M) T} \qquad (11)$$

where :

$P(M,T)$ = probability of earthquake in recurrence
time T = observation time

$N(M)$ = cumulative value of the frequency (intensity) of earthquakes every year

The value of this recurring period will be matched to the existing data and then analyzed the success of the method by comparing the accuracy and the data that already exist.

4) Semantic Interpretation

After calculation of probabilistic recurrence time and the probability of an earthquake in recurrence time, we provide the seismicity prediction analysis with a series of magnitudes for each cluster and give the semantic interpretation of the prediction for each cluster. Fig. 9 shows the probabilistic prediction with a series of magnitudes depending on the earthquake cluster.

To give the semantic interpretation for the probabilistic prediction, we set a series of meaning for the dimensional vectors from feature extraction of earthquake data distribution and probabilistic prediction of an earthquake in recurrence time, as shown in Table 1. For applicability of our system for the Indonesian user, we use the semantic interpretation in Indonesian.

Magnitude (SR)	Time Occurences...	Probability (%)	Description
6	1.0	100.0	View
7	18.0	52.0	View
8	554.0	2.0	View
9	Unpredictable	Unknown	View

Fig. 9 Probabilistic prediction with a series of magnitudes depending on the earthquake cluster

TABLE I
A SERIES OF MEANING FOR SEMANTIC INTERPRETATION

Feature Vector	Description	Meaning
T	length of observation period	for observation during T year
c_i	cluster i	in cluster i
M	magnitude	with magnitude M richter scale
Tr	recurrence time	with Tr recurrence time
$P(M,T)$	probability of earthquake in recurrence time	with confidence level $P(M,T)$ %
Semantic Interpretation		
For observation during T year with confidence level $P(M,T)$ %, it will be predicted that the earthquake will hit a region in cluster i with magnitude M richter scale during next T_r years		
Semantic Interpretation in Indonesian		
Dengan waktu observasi selama T tahun, dengan cluster i dengan kekuatan magnitude sebesar M skala Richter diprediksikan akan memiliki waktu perulangan kejadian gempa setiap T_r tahun sekali dengan tingkat keyakinan sebesar $P(M,T)$ %		

III. RESULTS AND DISCUSSION

This experimental study uses the Advanced National Seismic System (ANSS) data. Data collection is obtained based on latitude and longitude of Indonesia with latitude interval: from -11.0 to 6.0, and longitude 95.0 to 141.0. The other parameters are the year (1963-2015) and magnitude (5.0-10.0). We make the experimental study with three earthquakes which are, (1) Earthquake in Nias, (2) Earthquake in Yogyakarta and (3) Earthquake in Padang.

A. Earthquake in Nias

In 2005 the earthquake occurred in Nias with magnitude 8.7 Richter scale. In this experimental study, to calculate the probabilistic prediction, the experiment used earthquake data from 1963 to 2004 (1 year before the earthquake hit Nias in 2005). After applying the automatic clustering, Nias located in cluster 4. Fig. 10 shows the result of probabilistic prediction for cluster 4 with different magnitudes. Fig. 10

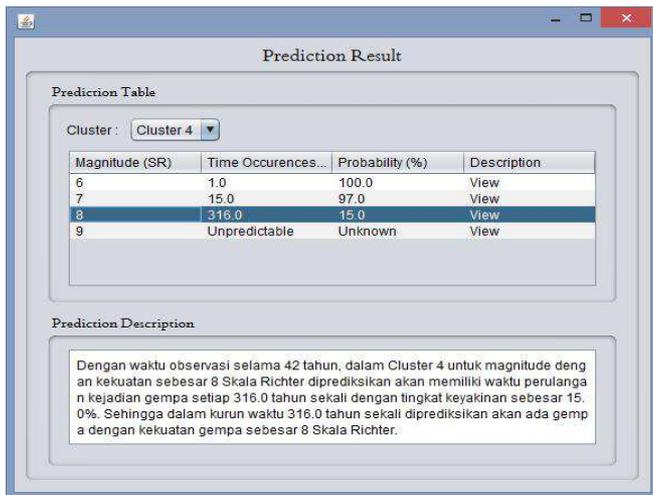also shows the semantic interpretation of the probabilistic prediction.

Fig. 10 Probabilistic Prediction for the earthquake in Nias in 2005

Figure 10 shows that for the time of observation of the 42-year earthquake prediction on the cluster for an earthquake with a magnitude of 6 Richter scale, it had an earthquake recurrence time of at least every two years with a magnitude 6 Richter scale with 100% confidence level. Whereas for magnitude with a strength of 7 Richter scale, it has a minimum earthquake occurrence time every ten years with a magnitude strength of 7 Richter scale occurring with 97% confidence level. For earthquake with magnitude strength of 8 Richter scale, it has an earthquake recurrence time of at least 316 years with magnitude strength of 8 Richter scale occurring with 15% confidence level. For magnitude with strength of 9 Richter scale and more, it has the result of "Unknown" that means the cluster cannot predict the recurrence time of an earthquake. From the probabilistic prediction in Fig. 10, our system cannot anticipate the earthquake in Nias in 2005 with 8.7 Richter scales. However, our system can reach closer anticipation of the future earthquake with providing the probabilistic prediction in 7 Richter during the next 15 years with 97% confidence level.

B. Earthquake in Yogyakarta

In 2006 the earthquake occurred in Yogyakarta with magnitude 6.3 Richter scale. In this experimental study, to calculate the probabilistic prediction, the experiment used earthquake data from 1963 to 2005 (1 year before the earthquake hit Yogyakarta in 2006). After applying the automatic clustering, Yogyakarta located in cluster 4. Fig. 11 shows the result of probabilistic prediction for cluster 4 with different magnitudes. Fig. 11 also shows the semantic interpretation of the probabilistic prediction.

Fig. 11 shows that for the time of observation of the 43-year earthquake prediction on the cluster for an earthquake with a magnitude of 6 Richter scale, it had an earthquake recurrence time of at least every 1 year with a magnitude 6 Richter scale with 100% confidence level. Whereas for magnitude with a strength of 7 Richter scale, it has a minimum earthquake occurrence time every 15 years with magnitude strength of 7 Richter scale occurring with 97% confidence level. For earthquake with magnitude strength of

8 Richter scale, it has an earthquake recurrence time of at least 347 years with magnitude strength of 8 Richter scale occurring with 13% confidence level. For magnitude with strength of 9 Richter scale and more, it has the result of "Unknown" that means the cluster cannot predict the recurrence time of an earthquake. From the probabilistic prediction in Fig. 11, our system can give a warning in which the earthquake will hit a region in cluster 4 (where Yogyakarta located in) during next 1 year with magnitude 6 Richter and 100% confidence level, and then the earthquake occurred in Yogyakarta in 2006 with magnitude 6.3 Richter.
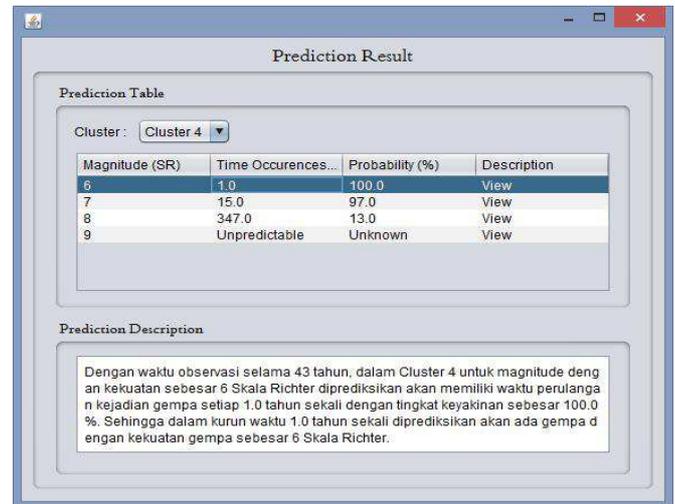

Fig. 11 Probabilistic Prediction for the earthquake in Yogyakarta in 2006

C. Earthquake in Padang

In 2009 the earthquake occurred in Padang with a magnitude 7.6 Richter scale. In this experimental study, to calculate the probabilistic prediction, the experiment used earthquake data from 1963 to 2008 (1 year before the earthquake hit Padang in 2009). After applying the automatic clustering, Padang located in cluster 6. Fig. 12 shows the result of probabilistic prediction for cluster 6 with different magnitudes. Fig. 12 also shows the semantic interpretation of the probabilistic prediction.
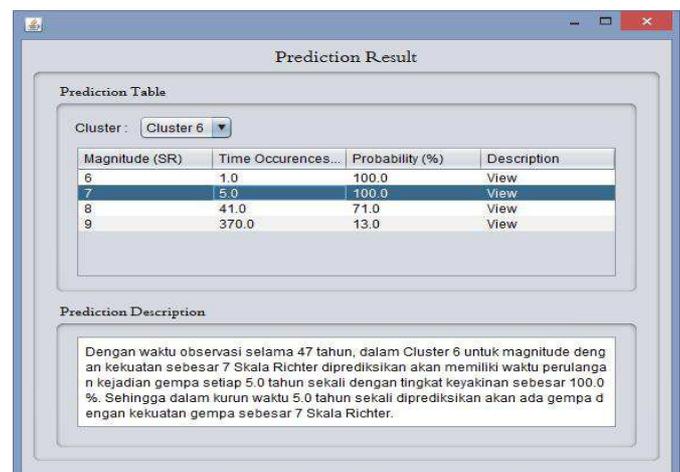

Fig. 12 Probabilistic Prediction for the earthquake in Padang in 2009

Fig. 12 shows that for the time of observation of the 46-year earthquake prediction on the cluster for an earthquake

with a magnitude of 6 Richter scale, it had an earthquake recurrence time of at least every 1 year with a magnitude 6 Richter scale with 100% confidence level. Whereas for magnitude with a strength of 7 Richter scale, it has a minimum earthquake occurrence time every 5 years with a magnitude strength of 7 Richter scale occurring with a 100% confidence level. For earthquake with magnitude strength of 8 Richter scale, it has an earthquake recurrence time of at least 41 years with magnitude strength of 8 Richter scale occurring with 71% confidence level. For magnitude with strength of 9 Richter scale and more, it has an earthquake recurrence time of at least 370 years with magnitude strength of 9 Richter scale occurring with 13% confidence level. From the probabilistic prediction in Fig. 12, our system can give a warning in which the earthquake will hit a region in cluster 6 (where Padang located in) during next five years with magnitude 7 Richter and 100% confidence level, and then the earthquake occurred in Padang in 2009 with magnitude 7.6 Richter.
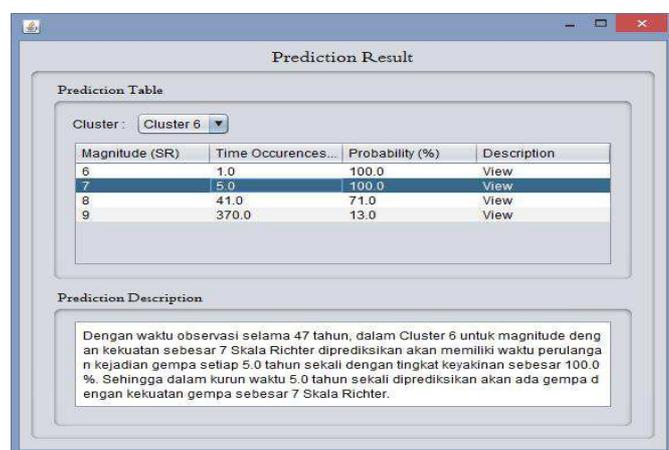

Fig. 12 Probabilistic Prediction for the earthquake in Padang in 2009

IV. CONCLUSION

In this paper, we have presented a seismicity prediction analysis based on automatic clustering for earthquake data distribution in Indonesia. This system has 4 main features: (1) Data acquisition and pre-processing, (2) Automatic clustering of earthquake seismic data distribution, (3) Seismicity prediction from earthquake time period occurs based on cluster with seismic event confidence level using Guttenberg-Richter law, and (4) Region-based prediction analysis and semantic interpretation of the prediction for each cluster.. For applicability of our proposed system, we made a series of an experimental study using ANSS earthquake data of the year 1963-2015 in 3 locations: Nias in 2005 with 8.7 Richter, Yogyakarta in 2006 with 6.3 Richter, and Padang in 2009 with 7.6 Richter. From the experimental result of each earthquake location, our system cannot anticipate the earthquake in Nias, but it can anticipate the earthquake in Yogyakarta and Padang. For the earthquake in Nias in 2005 with 8.7 Richter scales, our system can reach closer anticipation of a future earthquake with providing the probabilistic prediction in 7 Richter during the next 15 years with 97% confidence level. For the earthquake in Yogyakarta in 2006 with 6.3 Richter, our system gave the probabilistic prediction during next 1 year with magnitude 6 Richter and 100% confidence level. For the earthquake in Padang in 2009 with 7.6 Richter, our system gave the probabilistic prediction next 5 years with a magnitude 7 Richter and 100% confidence level. Our proposed system also gave a straightforward interpretation of prediction probability by providing semantic interpretation of the prediction for each cluster.

REFERENCES

[1] Cincin Api Pasifik, Wikipedia Indonesia, https://id.wikipedia.org/wiki/Cincin_Api_Pasifik accessed on July 8, 2015.

[2] Gempa Bumi, Badan Meteorologi, Klimatologi, dan Geofisika, www.bmkg.go.id/BMKG_Pusat/Gempabumi_Tsunami/Gempabumi. bmkg, accessed on July 8, 2015.

[3] Mohammad Nur Shodiq, Ali Ridho Barakbah, Tri 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) 2015, Surabaya/Bali, Indonesia, March 24-26, 2015.

[4] R. Sadeghian, G.R. Jalali-Naini, "A New Probability Density Function in Earthquake Occurrences," Journal of Industrial Engineering International, Islamic Azad University, Vol. 4, No. 6, pp. 1-6, 2008.

[5] Rusnardi Rahmat Putra, Junji Kiyono, "Seismic Hazard Analysis for Indonesia", Journal of Natural Disaster Science, Vol. 33, No. 2, pp. 59-70, 2012.

[6] Das S., C. Henry, "Spatial Relation between Main Earthquake Slip and Its Aftershock Distribution," Department of Earth Sciences, University of Oxford, 2003.

[7] Masryur Irsyam, Donny T Dangkua, "Proposed Seismic Hazard Maps of Sumatra and Java Islands and Microzonation Study for Jakarta City, Indonesia," U.S Geological Survey USGS-MS966, 2008.

[8] Fujiwara Hiroyuki, Sinichi Kawai, "National Seismic Hazard Maps of Japan," University of Tokyo, 2006.

[9] Restu Faizah, Habib, Widodo, "Probabilitas Kejadian Gempabumi Pada Masa Mendatang Di Zona Sesar Sumatra", Seminar Nasional Statistika dalam Managemen Kebencanaan, Fakultas MIPA, UII Yogyakarta, 15 Juni 2013

[10] Adel Moatti, Amin, Zafarani, "Pattern Recognition on Seismic Data for Earthquake Prediction Purpose", Proceedings of International Conference on Environment, Energy, Ecosystems and Development, Industrial Engineering Tarbiat Modares University Tehran, Iran, 2013.

[11] Mohammad Nur Shodiq, Ali Ridho Barakbah, Tri Harsono, "Cluster Oriented Spatio Temporal Multidimensional Data Visualization of Earthquake in Indonesia", EMITTER International Journal of Engineering Technology, Vol. 3, No. 1, pp. 53-67, June 2015.

[12] Mohammad Nur Shodiq, Dedy Hidayat Kusuma, Mirza Ghulam Rifqi, Ali Ridho Barakbah, Tri Harsono, "Spatial Analysis of Magnitude Distribution for Earthquake Prediction using Neural Network Based on Automatic Clustering in Indonesia", The International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC) 2017-IEEE co-sponsored conference, Surabaya, Indonesia, September 26-27, 2017.

[13] Mohammad Nur Shodiq, Dedy Hidayat Kusuma, Mirza Ghulam Rifqi, Ali Ridho Barakbah, Tri Harsono, "Neural Network for Earthquake Prediction Based on Automatic Clustering in Indonesia," International Journal on Informatics Visualization, Vol. 2, No. 1, 2018.

[14] NECDC. ANSS Catalog Search. Accessed on 20 December 2015.

[15] Ali Ridho Barakbah, Kohei Arai, "Identifying moving variance to make automatic clustering for normal data set," IECI Japan

Workshop 2004 (IJW 2004), Musashi Institute of Technology, Tokyo, May 22, 2004.

[16] Ali Ridho Barakbah, Kohei Arai, "Determining constraints of moving variance to find global optimum and make automatic clustering," Proc. Industrial Electronics Seminar (IES) 2004, pp.409-413, Surabaya, Indonesia, October 12, 2004.

[17] Ali Ridho Barakbah, Kohei Arai, "Reversed pattern of moving variance for accelerating automatic clustering," EEPIS journal, pp.15-21, Vol. 9, No. 2, December 2004.

[18] Bambang Sunardi, "Analisa Fraktal dan Rasio Slip Daerah Bali-NTB Berdasarkan Pemetaan Variasi Parameter Tektonik", Jurnal meteorologi dan Geofisika, Vol. 10, No.1, 2009.

[19] Ahmad Zamani, Reza M. Sorbi, "Application of neural network and ANFIS model for earthquake occurrence in Iran," Earth Sci Inform. Springer-Verlag Berlin Heidelberg, 2013.

[20] E. Florido, F. Martínez-Álvarez, A. Morales-Esteban, J. Reyes, J.L. Aznarte-Mellado, "Detecting precursory patterns to enhance earthquake prediction in Chile ", Computers and Geosciences, 2014.

[21] Supriyanto Rohadi, "Studi Variasi Spatial Seismisitas Zona Subduksi Jawa", Jurnal Meteorologi Dan Geofisika, Vol. 8 No.1, Juli 2007.