The Fourth Indonesian-Japanese Conference
on Knowledge Creation dan Intelligent Computing (KCIC) 2015 ISBN : 978-602-72251-0-7

Spatial Analysis of Earthquake Distribution with Automatic Clustering for Prediction of Earthquake Seismicity in Indonesia

Mohammad Nur Shodiq[1], Ali Ridho Barakbah[2], Tri Harsono[3]
Graduate Program on Information Engineering and Computer,
Electronic Engineering Polytechnic Institute of Surabaya
[1]mohnoershodiq@gmail.com, [2]ridho@pens.ac.id, [3]trison@pens.ac.id

Abstract

Many researchers analyzed the earthquakes for predicting the earthquake time period occurrences. This prediction requires the area that has similarity among of the earthquake dataset. However, they commonly faced the difficulty to determine automatic distribution of the high-similarity regions triggered by the spatio-temporal earthquakes. This paper proposes a new approach for determining the area based on earthquake datasets. Its uses automatic clustering with Valley Tracing method to determine the number of optimal earthquake clusters. Then, visualize the clusters based on spatial distribution of cluster. Every clusters are analyzed by the probability of earthquake occurrence with the Gutenberg-Richter law. We made series of experimental studies with earthquake data from 2004 until 2014 in Indonesia. The experimental results performed high accuracy for predicting earthquakes during 1-6 forthcoming years.

Keywords: probability, earthquake, automatic clustering.

1. Introduction

Earthquake is the event of the earth due to release of energy in the earth suddenly. It was caused by the sudden breaking a layer of rock or plate fracture in the earth's crust [1]. The interaction between these plates place Indonesia on the area that has volcanic activity and high seismicity [2]. The high seismic activity could be seen from the results of earthquake recording from 1897 to 2009, there are more than 14.000 seismic events with magnitude M $\geq$ 5.0(SR). These quakes have caused thousands of deaths, destruction and damage to thousand of buildings and infrastructures, as well as substantial funds for rehabilitation and reconstruction [3] [4].

Earthquakes occasionally occur in groups of space and time. So, the scientists are developing a model to explain this grouping pattern recognition. Therefore, it is required modeling of earthquake clustering more accurate to develop a model that explains the pattern or grouping behavior [5]. The result of the clustering process is useful to determine the level of risk of earthquakes in a region in Indonesia.

Each regions has the different historical earthquake datasets. Information in the dataset are used for predicting the probability of earthquake occurrence. One of the empirical relationships that has been used frequently in long-term prediction is the Gutenberg-Richter law [6].

Analysis for earthquake distribution is essential, especially in countries that often occur earthquakes. There are many researchers who studied the field of earthquake, including: Faizah, Wahdi, and Widodo have developed the probability of earthquake in future events using conditional method probability. However, this work was limited to spesific area that directly selected, that is in sumatera fault zone [14]. Moatti, Reza, and Zafarani have developed pattern recognition on earthquake seismic data with Gutenberg-Richter law for prediction of earthquakes in the future, and obtained the optimal number of clusters with silhouette index [6].

2. Proposed Idea

This research proposes a new approach for measuring the risk analysis of earthquake probability events using automatic clustering and visualize the clusters based on spatial distribution of cluster on Indonesian region. We focus this research in Indonesia because it is an archipelago where three plates of the world meet. The interaction between these plates place Indonesia as the region that has volcanic activity and high seismicity. This research applies earthquake dataset from indonesia that is provided by the Agency Meteorologi, Climatology and Geophysics (BMKG), Indonesia.

The automatic clustering in this research consists of two processes. The first process is to find the global optimum of clustering using Valley Tracing [7]. It analyzed the moving variance of clusters for each stage of cluster contruction, then observed the pattern to find the global optimum as well as to avoid the local optima. The second process is to cluster the dataset using Single Linkage Hierarchical K-means clustering [8]. This clustering requires a number of cluster for real clustering seismic catalog. So that, a number of optimal cluster from first process becomes initial clusters for Hierarchical K-Means method [8].

The Fourth Indonesian-Japanese Conference
on Knowledge Creation dan Intelligent Computing (KCIC) 2015 ISBN : 978-602-72251-0-7

The result of the Hierarchical K-Means clustering proces will be visualized on the map based on spatial distribution of clusters on Indonesia region. Every cluster has member of earthquakes data. Then, this data are analyzed the probability of earthquake occurrence with the Gutenberg-Richter law.

3. System Design

3.1 Procedural Concept of Proposed System

The procedural concept of our proposed system can be seen in Figure 1.

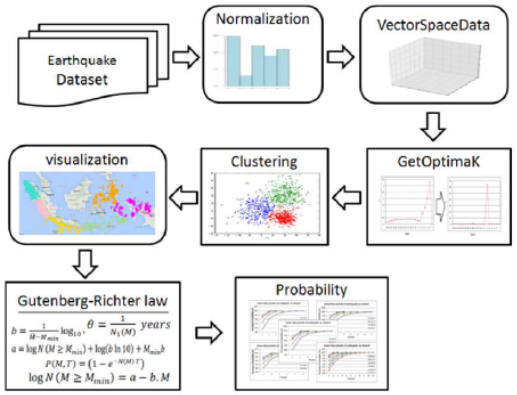

Figure 1. Design system probability of earthquake occurrence based automatic clustering

The earthquake dataset parameters consist of longitude and lattitude. Those parameters will be normalized for scaling of earthquake dataset attribute values in same range values between 0-1. Afterwards the datasets are stored in the VectorSpaceData. GetOptimaK is a process for determining the optimal number of clusters. The number of cluster will be initial cluster to the process of real clustering the datasets. The result of this clustering will be visualized on the map based on spatial distribution of cluster. And then, The last process is to analyze the probability of an earthquake return time period of each cluster using Gutenberg-Richter law.

3.2 Earthquake Dataset

The earthquake dataset in this research come from catalogue of Agency Meteorolgi, Climatology and Geophysics (BMKG) from January 2004 to July 2014. The number of earthquake occurence is 2665 data and becomes 1262 data after filtering the datasets. The earthquake data covers the entire territory of Indonesia, which is the boundary 6 north latitude - 11 south latitude and east longitude 95 - 141 east longitude, and depth of 0-650 km.

Spatial earthquake distribution map of seismicity in Indonesia in this research could be seen in Figure 2, the x-axis represents the longitude, the y-axis represents the lattitude, colors represents the hypocenter, and the diameter of the dot represents the magnitude. Each parameter has a different scale, hypocenter consists of nine categories that represent the color depth, and diameter of a point consists of nine levels that represent the magnitude scale.

The magnitude of completeness (Mc) in this research uses 5.1 [9]. Mc is required for probabilistic analysis. Incompleteness of seismic data will give the result in seismic risk parameters resulting into overestimated or underestimated. Mc could be a function as a threshold magnitude, so the event of an earthquake under the Mc value will be eliminated [9] [4].

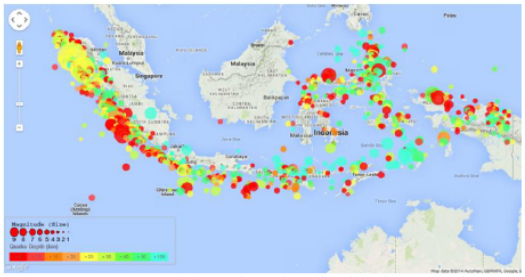

Figure 2. Spatial earthquake distribution map in indonesia from catalogue of BMKG during 2004 - 2014

3.3 Automatic Clustering

3.3.1 GetOptimalK

GetOptimalK is a process for determining the optimal number of clusters. This prosess uses cluster analysis for analyzing the cluster. Cluster analysis constructs good cluster when the members of a cluster have a high degree of similarity to each other (internal homogeneity) and are not like members of other clusters (external homogeneity) [7]. Each cluster has their own variance of cluster (V_c). Variance of cluster can be referred as an identity of each cluster. The variance of cluster can be calculated as:

$$v_c^2 = \frac{1}{n_c - 1} \sum_{i=1}^{n} \left(d_i - \overline{d_c}\right) \quad (1)$$

where : v_c^2 = variance of cluster c
c = 1... k,
n_c = total numbers of each clusters
d_i = the data member-i in a cluster
$\bar{d}_c$ = the centroid of cluster c

while the variance within clusters (v_w) can be defined as:

The Fourth Indonesian-Japanese Conference
on Knowledge Creation dan Intelligent Computing (KCIC) 2015 ISBN : 978-602-72251-0-7

$$v_w^2 = \frac{1}{N-k}\sum_{i=1}^{k}(n_i - 1).v_i^2 \quad (2)$$

where: N = total numbers of members in all clusters
n_i = The amount of data in cluster i
k = number of clusters
v_i = variance of cluster *i*

then, variance between cluster (v_b) can be defined as:

$$v_b^2 = \frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{d}_i - \bar{d})^2 \quad (3)$$

where: v_b = variance between cluster
$\bar{d}$ = average from $\bar{d}_i$

The ideal cluster has minimum V_w (variance within cluster) and a high maximum V_b (variance between cluster). The following formula is to calculate variance V.

$$v = \frac{v_w}{v_b} x100\% \quad (4)$$

However, finding the ideal cluster is very difficult, because we can not directly apply minimal (V) to find the global optimum cluster. So, to find the global optimum as the ideal cluster requires identifying the moving variance that has been introduced by Barakbah 7 [2]. The following Table 1 shows the identifying pattern of moving variance.

Table 1. Pattern of moving variance

Pattern	Possible?	Pattern	Possible?
	√		x
	√		x
	√		x
	x		√
	x		x
	x		x
	x		x
	x		

The possibility to find the global optimum by hill-tracing resides in stage i, using pattern of moving variance like Figure 3 and defined as below:

$$v_i = \alpha.v_{i+1} \quad (5)$$

where α is altitude value

Figure 3. Pattern of hill-tracing

While, the possibility to find the global optimum by valley-tracing resides in stage i, using pattern of moving variance like the Figure 4 and defined as below:

$$(v_{i-1} \geq v_i) \cap (v_{i+1} > v_t) \quad (6)$$

where i = 1..n,

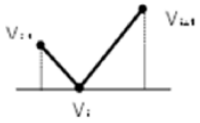

Figure 4. Pattern of valley-tracing

Furthermore, both the approach valley-tracing and hill-climbing method to identify high value difference (∂) at each stage, which is defined by the following formula:

$$\partial = (v_{i+1} - v_i) + (v_{i-1} - v_i)$$
$$= (v_{i+1} + v_{i-1}) - (2.v_i) \quad (7)$$

∂-value is used to avoid local optima, where this equation is obtained from the maximum ∂ filled as in equation (7) above.

To get maximum ∂-value, put λ as a threshold, then it will contruct automatic clustering. To construct automatic clustering, put λ as a threshold by the following formula:

$$max(\partial) \geq \lambda \quad (8)$$

The more complex clustering case need smaller l to set as more precise as possible.

To determine the accuracy of the automatic clustering, can be defined as:

$$\varphi = \frac{max(\partial)}{closer\ valuetomax(\partial)} \quad (9)$$

where: φ = accuracy value closer value to max(∂) is a candidate value max (∂) previously.

the result of φ value greater than 2, will show that the clusters are well-separated clusters.

3.3.1 Hierarchical K-Means Clustering

Hierarchical K-means clustering is a combination of K-Means and hierarchical clustering [8].. The Hierarchical K-means clustering algorithm is as follows:

The Fourth Indonesian-Japanese Conference
on Knowledge Creation dan Intelligent Computing (KCIC) 2015 ISBN : 978-602-72251-0-7

1. Set as each data of A, where 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 data set, with K as predefined number of clusters
10. Apply hierarchical algorithm
11. Record the centroids of clustering result as $D\{d_i|i = 1, \dots, K\}$
12. Then$D$ $\{d_i|i = 1, \dots, K\}$ as initial cluster centers for K-means clustering.

3.4 Visualize the Spatial Data Distribution based on Cluster

Sometimes, spatial data and magnitude are used to analyze the earthquake on many researches that related to the earthquake [4] [10] [11] [12]. Thus, earthquake dataset in this research which is used for clustering is epicenter parameter that consist of longitude and lattitude.

Clustering algorithm in this research uses hierarchical kmeans clustering. It has four models, namely single linkage, centroid linkage, complete linkage, and average centroid. This model is used to determine the optimal number of clusters that has a high accuracy rate. The results of the data process can be seen in Table 2.

Table 2. Results of the number of cluster and accuracy

	single	centroid	complete	average
Number of optimal cluster	6	3	6	3
Accuracy	1.90	2.68	100	2.13
Execution time (s)	35.47	46.30	44.90	45.28

From Table 2, the accuracy result of single linkage is 1.90, centroid linkage is 2.68, complete linkage is 100, and centroid linkage is 2.13.The result of accuracy value greater than 2 will show that the clusters are well-separated clusters. In Table 2, there are 3 algorithms that have accuracy more than 2. So, in this research will use algorithm that has highest accuracy. The best accuracy results obtained is complete linkage. This algorithm has accuracy 100. Therefore, the optimal number of cluster comes from this process, that has number of cluster is 6.

After a predetermined number of clusters, the next process is the process of real clustering on earthquake dataset. Real clustering method in this research is hierarchical kmeans clustering. This method has four models, namely, the single linkage, centroid linkage, complete linkage, and average linkage. To get good method for cluster this earthquake dataset, we use cluster analysis. cluster analysis is used to measure the value of the spread of data clustering, there are sum of squared error (SSE), and variance cluster (V). Cluster analysis of measurement results can be seen in Table 3 below.

Table 3. Cluster analysis earthquake dataset

	Single	Centroid	Complete	Average
SSE	0.98454	0.98347	0.98347	0.98454
Variance	3.4147×10^{-4}	3.4131×10^{-4}	3.4131×10^{-4}	3.4147×10^{-4}
time(s)	1.424	6.116	6.259	6.068

Based on Table 3, the results of cluster analysis measurement using SSE, the best value are centroid kmeans, that is 0.98347 and complete kmeans, that is 0.98347. While the best value of variance cluster measurement at centroid kmeans and complete kmeans, the variance has same value, that is 3.4131×10^{-4}. Thus, in this research uses centroid algorithm on hierarchical kmeans clustering, because it has time less than complete kmeans, that is 6.116 second.

Spatial data distribution of visualization based on cluster uses PHP/JavaBridge tools and google map on Indonesia map. This visualization can be seen in Figure 5 below.

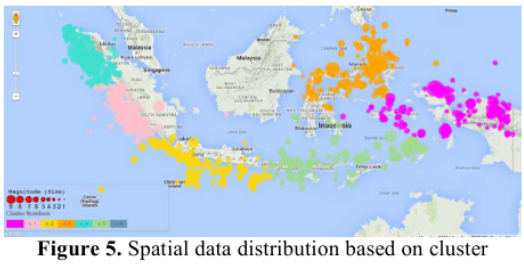

Figure 5. Spatial data distribution based on cluster

In Table 5, there are 6 clusters. Every cluster has different colorsS. The cluster0 colored by orange, cluster1 colored by yellow, cluster2 colored by pink, cluster3 colored by red, cluster4 colored by blue, and cluster5 colored by green. Whereas, the amount of data each cluster can be seen in Table 4.

Figure 2 and Figure 5 have equal data distribution. Figure 2 describes the distribution of earthquake without any clustering of the earthquake. Most earthquakes occur in offshore. While Figure 5 describes the results of a grouping character of the earthquake data. The results clustering is 6 clusters. The regions that are members of the cluster can be seen in Table 5. Naming the area of the earthquake is based on the distance offshore with the nearby area. Thus, the name of the region can be classified as more than one cluster.

The Fourth Indonesian-Japanese Conference
on Knowledge Creation dan Intelligent Computing (KCIC) 2015 ISBN : 978-602-72251-0-7

Tabel 4. The amount of data member of cluster

Cluster	amount of data	Average magnitude	minimal magnitude
Cluster0	286	5.60	5.1
Cluster1	169	5.56	5.1
Cluster2	243	5.67	5.1
Cluster3	179	5.70	5.1
Cluster4	238	5.60	5.1
Cluster5	147	5.76	5.1

Based on Table 4 above, cluster5 has the fewest number, that is 147 data and cluster0 has the most number, that is 286 data. Cluster1 has lowest Magnitude average, that is 5.56, while cluster5 has the highest in that is 5.76. All clusters have a minimum magnitude 5.1 (SR). The number of cluster members as shown in Table 4 will be a reference in the analysis of the probability of an earthquake.

While the Indonesian territory belonging to the cluster can be seen in Table 5.

Table 5. Member cluster based on region

Cluster	Region	Cluster	Region
cluster0	Gorontalo Kalimantan Timur Maluku Maluku Utara Papua Barat Sulawesi Utara Sulawesi Barat Sulawesi Selatan Sulawesi Tengah Sulawesi Tenggara	cluster3	Ambon Irian Barat Irian Jaya Jayapura Maluku Maluku Selatan Maluku Tenggara Maluku Utara Papua Papua Barat Sulawesi Tenggara Sulawesi Utara
cluster1	Bali Banten Bengkulu DKI Jawa Barat Jawa Tengah Jawa Timur Lampung NTT Yogyakarta	cluster4	Aceh Sumatera Barat Sumatra Utara
cluster2	Bengkulu Jambi Lampung Sumatra Barat Sumatra Selatan Sumatra Utara	cluster5	Bali Maluku Maluku Tenggara Nusa Tenggara Barat Nusa Tenggara Timur Sulawesi Selatan Sulawesi Tenggara

Probability of earthquake occurrence each cluster at a certain time has different results, its based on historical earthquake data. The equation to determine probability of earthquake event with a magnitude M and time period T is as follows [13]:

$$P(M,T) = \left(1 - e^{-N(M)T}\right) \quad (13)$$

where:

$P(M,T)$ is the probability of an earthquake with a magnitude M and time period T.
T is the time (year/s).
$N(M)$ is the number of cumulative frequency of earthquakes per year or index seismicity.

The average value of the return period of earthquake occurrence can be calculated by as follows [14]:

$$\theta = \frac{1}{N_1(M)} years \quad (14)$$

where:
θ is return time periode

Validation model of equation (13) on probability of earthquake occurrence in this study uses holdout method. This method splits dataset into two groups, first group is data learning, it used to train the classifier, and second group is data test, it used to estimate the error rate of the trained classifier. In this study, there are three experiments with data sets as Table 6 follows :

Table 6. Dataset for validation model

DataSet	data learning	data test
Q1	Earthquake dataset 2004-2008	Earthquake dataset 2009-2014
Q2	Earthquake dataset 2004-2010	Earthquake dataset 2011-2014
Q3	Earthquake dataset 2004-2012	Earthquake dataset 2013-2014

Performance analysis value on dataset Q1, Q2, and Q3 uses true, false, and unknown. Where, true value is used when there is an earthquake in datatest, false value is used when there is no an earthquake in datatest, and unknown value is used when probability value in outside range datatest. In Q1 dataset as shown in Table 7, true value amount to 10 (42%), false value amount to 2 (8%), and unknown value amount to 12 (50%). In Q2 dataset as shown in Table 8, true value amount to 9 (38%), false value amount to 5 (21%), and unknown value amount to 10 (42%). In Q3 dataset as shown in Table 9, true value amount to 6 (25%), false value amount to 4 (17%), and unknown value amount to 14 (58%).

The Fourth Indonesian-Japanese Conference
on Knowledge Creation dan Intelligent Computing (KCIC) 2015 ISBN : 978-602-72251-0-7

Table 7. Validation model result from Q1 dataset

	Magnitude	P(M,T), T=6 years	return time period: 0 years	Datatest							value
				2009	2010	2011	2012	2013	2014	total	
Cluster 0	6	100%	1	9	6	6	4	2	4	31	True
	7	89%	3	1				1		2	True
	8	27%	20	0						0	Unknown
	9	4%	139	0						0	Unknown
cluster1	6	100%	1	3	3	6	2	1	1	16	True
	7	49%	10	2						2	True
	8	7%	87	0						0	Unknown
	9	1%	838	0						0	Unknown
cluster2	6	100%	1	8	5	2	1	1	0	17	True
	7	99%	2	0		3		0		3	False
	8	57%	8	0						0	Unknown
	9	15%	37	0						0	Unknown
cluster3	6	100%	1	11	9	3	7	1	2	33	True
	7	59%	7	7						7	True
	8	11%	50	0						0	Unknown
	9	2%	362	0						0	Unknown
cluster4	6	100%	1	0	3	2	2	2	2	11	True
	7	91%	3	2			0			2	False
	8	28%	19	0						0	Unknown
	9	4%	133	0						0	Unknown
cluster5	6	100%	1	7	3	3	2	6	0	21	True
	7	93%	3	1				1		2	True
	8	46%	10	0						0	Unknown
	9	13%	44	0						0	Unknown

Table 8. Validation model result from Q2 dataset

	Magnitude	P(M,T), T=4 years	return time period: 0 years	Datatest					value
				2011	2012	2013	2014	total	
Cluster 0	6	100%	1	6	4	2	4	16	True
	7	91%	2	0		1		1	True
	8	29%	12	0				0	Unknown
	9	5%	81	0				0	Unknown
cluster1	6	100%	1	6	2	1	1	10	True
	7	65%	4	1				1	True
	8	13%	30	0				0	False
	9	2%	236	0				0	False
cluster2	6	100%	1	2	1	1	0	4	True
	7	97%	2	0		0		0	False
	8	49%	6	0				0	Unknown
	9	12%	32	0				0	Unknown
cluster3	6	100%	1	3	7	1	2	13	True
	7	93%	2	1		1		2	True
	8	43%	8	0				0	Unknown
	9	11%	35	0				0	Unknown
cluster4	6	100%	1	2	2	2	0	6	True
	7	86%	3	0			0	0	False
	8	24%	15	0				0	Unknown
	9	4%	104	0				0	Unknown
cluster5	6	100%	1	3	2	6	0	11	True
	7	94%	2	1		0		1	False
	8	46%	7	0				0	Unknown
	9	13%	29	0				0	Unknown

The result of return time period each clucter could be seen on Table 10, while probability calculation result earthquakes could be seen in Figure 6 through Figure 9 as follows.

Level of earthquake risk with a magnitude M≥6 each cluster has high level, it could be seen in Table 10 above, each cluster has return period of 1 year, it means that the occurrence of this earthquake will occur every year. While, the earthquake with magnitude M≥7 has return period of 2 years, except on cluster1. Cluster1 has earthquake level of risk lower than the other clusters, that is 4 years. Level of earthquake risk with a magnitude M≥8 has different level of risk each cluster. on cluster0, cluster2, cluster3, cluster4, and cluster5 have a return period between 5 to 15 years, whereas in cluster1 has a return period of 33 years. While, level of earthquake risk with a magnitude M≥9 on cluster5 has an earliest return period, that is 25 years, while the cluster that has the longest return period is cluster1 which has 283 years of return period.

As in Figure 6, percentage of probability on cluster0, cluster2, cluster3, cluster4, and cluster5 are 100% at the time is 2 years, its means that there will occurrence one or more earthquakes with magnitude greater than M ≥ 6 within a time period of 2 years, whereas in cluster1 will occur in the 3 years. Based on the return time period at Table 10, This earthquake will be repeated every 1 year.

The Fourth Indonesian-Japanese Conference
on Knowledge Creation dan Intelligent Computing (KCIC) 2015 ISBN : 978-602-72251-0-7

Table 9. Validation model result from Q3 dataset

	Magnitude	P(M,T), T=2 years	return time period: θ years	Datatest			value
				2013	2014	Total	
Cluster 0	6	100%	1	2	4	6	True
	7	72%	2	1		1	True
	8	16%	12	0		0	Unknown
	9	2%	83	0		0	Unknown
cluster1	6	99%	1	1	1	2	True
	7	41%	4	0		0	Unknown
	8	6%	33	0		0	Unknown
	9	1%	271	0		0	Unknown
cluster2	6	100%	1	1	0	1	False
	7	84%	2	0		0	Unknown
	8	28%	7	0		0	Unknown
	9	6%	34	0		0	Unknown
cluster3	6	100%	1	1	2	3	True
	7	80%	2	1		1	True
	8	29%	6	0		0	Unknown
	9	7%	28	0		0	Unknown
cluster4	6	100%	1	2	2	4	True
	7	63%	2	0		0	False
	8	13%	15	0		0	Unknown
	9	2%	109	0		0	Unknown
cluster5	6	100%	1	6	0	6	False
	7	77%	2	0		0	False
	8	28%	6	0		0	Unknown
	9	7%	27	0		0	Unknown

Table 10. Return time period on clucter

Magnitude	Cluster0	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
6	1	1	1	1	1	1
7	2	4	2	2	2	2
8	12	33	7	7	14	6
9	86	283	39	36	102	25

As in Figure 7, percentage of probability on cluster0, cluster2, cluster3, cluster4, and cluster5 are 100% at the time is 10 years, its means that there will occurrence one or more earthquakes with magnitude greater than M ≥ 7 within a time period of 10 years, whereas in cluster1 will occur in the 20 years. Based on the return time period at Table 10, This earthquake will be repeated every 2 year, except in cluster1. Cluster1 has a return period of 4 years.

As in Figure 8, percentage of probability on cluster0, cluster2, cluster3, cluster4, and cluster5 are 100% at the time is 100 years, its means that there will occurrence one or more earthquakes with magnitude greater than M ≥ 8 within a time period of 100 years, whereas in cluster1 has a percentage of 96% at the time is 100 years. Based on the return time period at Table 10, This earthquake has a different return period, in cluster0 has a return period of 12 years, cluster1 has a return period of 33 years, cluster2 and cluster3 have the same return period of 7 years. whereas in cluster5 has a return period of 6 years

As in Figure 9, percentage of probability on cluster0, cluster2, cluster3, and cluster5 are 100% at the time is 500 years, its means that there will occurrence one or more earthquakes with magnitude greater than M ≥ 9 within a time period of 500 years, whereas in cluster1 has a percentage of 83% and in cluster4 has a percentage of 99% at the time is 500 years. Based on the return time period at Table 10, cluster5 has fastest return period that is 25 years, while the cluster1 has the longest return period that is 283 years.

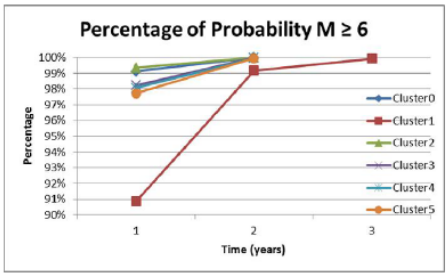

Figure 6. Percentage of probability at magnitude M ≥ 6

 ISBN : 978-602-72251-0-7

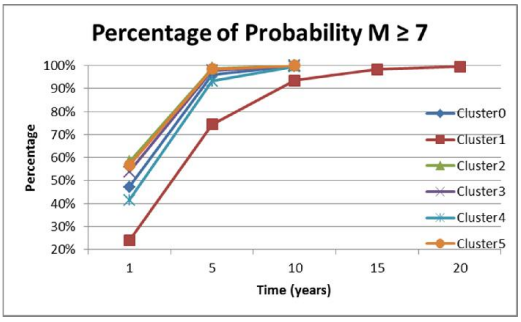

Figure 7. Percentage of probability at magnitude M ≥ 7

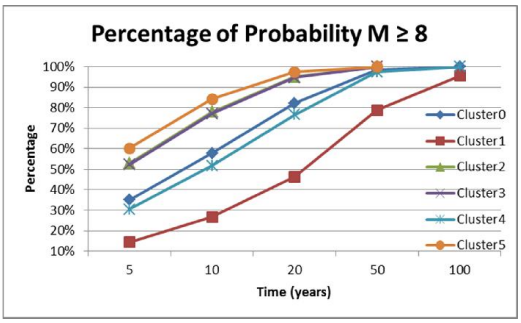

Figure 8. Percentage of probability at magnitude M ≥ 8

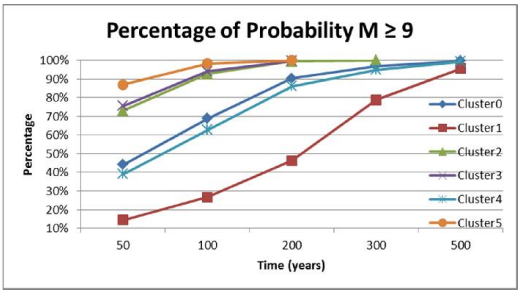

Figure 9. Percentage of probability at magnitude M ≥ 9

4. Conclusion

This paper proposes an approach for measuring the optimal number of cluster using valley tracing and hill climbing method. While, the proximity measurement data using complete linkage that has an accuracy value of 100 and the optimal number clusters is 6 clusters. While, clustering data algorithm uses centroid linkage on hierarchical kmeans clustering, it has a SSE value 0.98347, variance value is 3.4131×10^{-4}, and the time required is 6.116 seconds.

Based on the probability of earthquake occurrence at Figure 6 through Figure 9, there is no earthquake damage within a period of 5 to 10 years. In other words, there is no earthquake damage with magnitud more than 7 (SR) in 2020. While, earthquake occurrence with magnitude M ≥ 8 and , M ≥ 9 will accur within the 50 and more 200 years.

cluster5 has a high level of earthquake risk , its mean that return period on this cluster is more short time than the others. In this cluster would accur earthquake with magnitud 6 (SR) every year. Whereas, magnitud 7 (SR), which can be categorized as highly damaging seismic event, would accur earthquake every 2 years. And also, it has a return time period 6 years on magnitude M ≥ 8 and 25 years on magnitude M ≥ 9. cluster5 has member of region, there are bali, maluku, maluku tenggara, nusa tenggara barat, nusa tenggara timur, sulawesi selatan, and sulawesi tenggara.

In further research, it will develop a model clustering with epicenter and time parameters, as well as its magnitude.

References

[1] BMKG Sanglah Denpasar, 2013, " Geodinamika Informasi Meteorologi Klimatologi dan Geofisika Vol.2 No.11 ", BMKG Sanglah Denpasar, Denpasar

[2] Anonim. 2007. "Analisis Potensi Rawan Bencana Alam di Papua dan Maluku (Tanah Longsor – Banjir – Gempa Bumi - Tsunami)", Laporan Akhir, Deputi Bidang Pembinaan Sarana Teknis dan Peningkatan Kapasitas, Kementerian Negara Lingkungan Hidup, Jakarta

[3] Irsyam, Masyhur. Dkk. 2010, "Ringkasan Hasil Studi Tim Revisi Peta Gempa Indonesia 2010", Tim Revisi Peta Gempa Indonesia. Bandung

[4] Sunardi, Bambang. 2009. " Analisa Fraktal dan Rasio *Slip* Daerah Bali-NTB Berdasarkan Pemetaan Variasi Parameter Tektonik". Jurnal meteorologi dan geofisika vol. 10 no.1 tahun 2009.

[5] Ace Tempest Re. 2009. " cat 360 making sense of earthkuaqe cluster". Newsletter. http://www.acegroup.com/bm-en/assets/cat3601q1010makingsenseofearthquakeclusters.pdf

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

[7] A. R. Barakbah, K. Arai, 2004," Determining Constraints of Moving Variance to Find Global Optimum and Make Automatic Clustering ", IES, Politeknik Elektronika Negeri Surabaya, Surabaya.

[8] Kohei Arai, Ali Ridho Barakbah, "Hierarchical K-means: an algorithm for centroids initialization for K-means", Reports of the Faculty of Science and Engineering, Saga University, Japan, Vol. 36, No. 1, 2007.

 ISBN : 978-602-72251-0-7

[9] Rohadi, Supriyanto. 2009. "Studi Seismotektonik Sebagai Indikator Potensi Gempabumi di Wilayah Indonesia" Jurnal Meteorologi dan Geofisika Volume 10 Nomor 2 Tahun 2009 : 111 – 120
[10] Sunardi, Bambang. 2008. "Studi Potensi Seismotektonik Sebagai Precursor Tingkat kegempaan di wilayah sumatera". Jurnal meteorologi dan geofisika vol. 9 no.2 tahun 2008.
[11] Sunardi, Bambang. 2007. "Studi Variasi Spatial Seismisitas Zona Subduksi Jawa". Jurnal meteorologi dan geofisika vol. 8 no.1 tahun 2007.
[12] Lilik Wahyuni Purlisstyowati, dkk. "Analisis Tingkat Resiko Gempa Bumi Tektonik di Papua pada Periode 1960-2010 ". Jurnal Fisika. Volume 02 Nomor 02 Tahun 2013
[13] Rohadi, Supriyanto, dkk "Studi Variasi Spatial Seismisitas Zona Subduksi Jawa". Jurnal Meteorologi Dan Geofisika, Vol. 8 No.1 Juli 2007
[14] 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