OPEN ACCESS
Indo-JC
Indonesia Journal of Computing
ISSN 2460-9056
socj.telkomuniversity.ac.id/indojc

Ind. Journal on Computing
Vol. 4, Issue. 3, Dec 2019. pp. 65-76
doi:10.21108/indojc.2019.4.3.359

Clustering of Earthquake Prone Areas in Indonesia Using K-Medoids Algorithm

Fiona Ramadhani Senduk #1, Indwiarti #2, Fhira Nhita #3

#Department of Informatics, Telkom University
Jl. Telekomunikasi No.1 Terusan Buah Batu Bandung, 40257 Indonesia

[1] fionarsenduk@student.telkomuniversity.ac.id
[2] indwiarti@telkomuniversity.ac.id
[3] fhiranhita@telkomuniversity.ac.id

Abstract

Located right above the ring of fire makes Indonesia prone to natural disasters, especially earthquakes. With the number of earthquakes that have occurred, disaster mitigation is very much needed. The use of data mining methods will certainly help in disaster mitigation. One method that can be used is clustering. The clustering algorithm used in this study is k-Medoids, The purpose of this study itself is to analyze the spatial patterns of earthquake distribution in Indonesia. The results of the cluster using k-medoids were also compared with the method previously utilized, namely K-Means. The data used are earthquake data from all regions in Indonesia during 2014-2018 that were recorded by the United State Geological Survey (USGS). From the cluster results obtained, the highest value of silhouette is 0.4574067 with the number k = 6. In addition, the study also found that k-medoids provide better silhouette values than k-means. The analysis of each clustering experiment is presented in this paper.

Keywords: clustering,data mining, earthquake, k-medoid.

Abstrak

Terletak persis diatas ring of fire membuat Indonesia rawan akan bencana alam, terutama gempa bumi. Banyaknya gempa bumi yang terjadi membuat upaya mitigasi bencana sangatlah dibutuhkan. Penggunaan metode data mining tentunya akan membantu dalam penanggulangan bencana gempa bumi. Salah satu metode yang dapat digunakan adalah clustering. Clustering yang dilakukan dalam penelitian ini menggunakan algoritma k-Medoids. Tujuan dari penelitian ini sendiri adalah untuk menganalisa pola spasial dari sebaran gempa di Indonesia. Hasil cluster menggunakan k-medoids juga dibandingkan dengan metode yang telah digunakan sebelumnya yaitu K-Means. Data yang digunakan adalah data titik gempa di seluruh daerah di Indonesia dari tahun 2014-2018 yang dicatat oleh United State Geological Survey (USGS). Dari hasil *clustering* yang didapatkan, nilai silhouette tertinggi ialah sebesar 0.4574067 dengan jumlah k=6. Selain itu, dari penelitian juga didapatkan bahwa k-medoids memberikan nilai silhouette yang lebih baik dibanding k-means. Analisis tiap percobaan clustering disajikan dalam *paper* ini.

Kata Kunci: clustering,data mining, k-medoid, gempa bumi.

I. Introduction

INDONESIA, located right above the ring of fire and the confluence of three large tectonic plates namely Eurasian, Indo-Australian, and Pacific plate, makes this country prone to natural disasters, especially earthquakes. Based on the data from USGS (2016), about 90% of earthquakes that happened in the world including the largest earthquakes occurred along the ring of fire.

Received on Nov 2019. Accepted on Dec 2019

An earthquake is an event where the surface of the earth vibrates. This vibration can be caused by various things including volcanic activity, meteor collisions, explosions that occur under the ground, and the movement of the earth's crust. But from several sources of the earthquake, the movement of the earth's crust became one of the most frequent causes of earthquakes [1]. This type of earthquake is commonly called a tectonic earthquake. Earthquake strength measurements are carried out using seismograph and earthquake strength is called magnitude, using the Richter scale [2]. Unlike other natural disasters, earthquakes cannot be predicted, unleashed within seconds, and happened without warning [3].

Based on the data from the Meteorology, Climatology and Geophysics Agency (BMKG) (2016-2018), in 2016, the number of earthquakes that happened in Indonesia recorded by BMKG was as many as 5,578 earthquakes, in 2017 there were 6,929 earthquakes, and in 2018 there were 11,577 earthquakes. 90% of these earthquakes are small and light earthquakes with magnitude <5.0 on the Richter scale. The 2004 earthquake in Aceh was recorded as one of the most devastating earthquakes in the world where the magnitude reached up to 9.1 Richter scale and caused tsunami waves with wave velocity reached up to 800 km/h [1]. Therefore, it is necessary to analyze spatial distribution of seismicity and the potential of seismogenic sources across the country to begin an effective mitigation of high earthquake risk in Indonesia.

Along with the times and the development of technology, many earthquake data have been obtained and can be studied. Data mining can be used to process and analyze these data. Data mining is a part of computer science that has the purpose of extracting information from a data set and transforming that information into a new information structure that can be understood for further use [4]. The pattern or trend of a data set can be studied or analyzed and the results of the analysis will be useful for decision making in the future. Data mining is an extraction of unique patterns from a data set [5]. There are several methods used in data mining such as classification, clustering, associations, etc. [6].

Clustering method is a process to group the data into several clusters or groups so that data in one cluster has a maximum level of similarity and the data between clusters has a minimum similarity [6]. Clustering is done when there is no information for each class to be predicted but the data must be divided into groups. [7] The clustering method has several algorithms that can be utilized for its implementation including k-means and k-medoids.

Many studies using k-means have been conducted to process earthquake data such as research conducted by Kamat & Kamath [8], Savaş et al. [9] and Novianti et al. [10]. However, according to research by Soni & Patel [11] which compares the k-means algorithm and k-medoids, k-medoids are more efficient than the k-means algorithm. Accuracy of clustering with k-medoids reached 92%, while k-means was only 88.7%. Whereas in research done by Arora & Varshney [12] and Selvi & Çağlar [13], k-medoids is superior in all aspects compared to k-means. The k-medoids is an algorithm that uses medoid (data point) in a cluster as the centroid. Unlike k-means, the k-medoids algorithm is not sensitive to outliers [4]. The basic strategy of the k-Medoids is to find k clusters in n objects by first randomly finding representative objects (medoid) for each cluster [14].

This research utilizes a k-medoids cluster analysis approach for the main purposes of (a) finding and identifying spatial patterns of seismic activities, which can be used to form a basis for delineating active tectonic areas that generate earthquakes and also (b) to compare the cluster results with a method that have been used in the previous studies, namely K-Means.

II. Literature Review

A. K-Medoids

K-medoids algorithm is a clustering algorithm that related to both k-means and medoidshift algorithm [15]. This algorithm is the development of the k-means algorithm. Both the k-means and k-medoids algorithms are partitional (breaking the dataset into several clusters) and the two algorithms aim to minimize the distance between the points in a cluster and a point that is the midpoint of the cluster [16]. Unlike k-means, k-medoids are not sensitive to noise or outliers. The steps of k-medoids are as follows [17]:

1. Determine the desired number of clusters
2. Select k data as centroid or medoid initialization, one centroid for each cluster
3. Calculate the distance between the data and the initial medoid (using Euclidean distance)
4. Allocate data to the cluster with the closest medoid and calculate the cost.
5. Update the medoid from each cluster with the category values that often appear in each cluster. Compare the cost value generated. If the cost value is smaller then replace the medoid with the new medoid value, if it is larger then there is no need to change the medoid.

Repeat the processes until all data has become medoid. Fig. 1 is the flowchart of k-medoids algorithm:

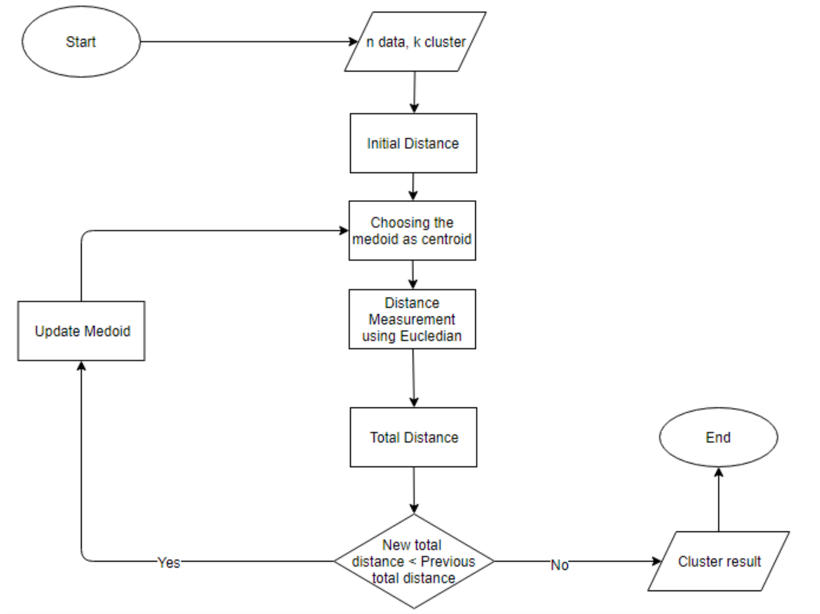

Fig. 1. Flowchart of K-Medoids Algorithm

B. K-Means

K-means algorithm is one of the most commonly used clustering methods in data mining. This algorithm aims to make the average value of a cluster as the centroid of the cluster [5]. K-Means clustering method groups the data based on their closeness to each other according to the Euclidean distance. K-means cluster analysis is an iterative process since it is a hard partitioning algorithm [10]. First, data are initially partitioned. Each group is calculated its mean and then the data partitioned again by allocating each data to its nearest means cluster position. In its simplest form, this algorithm consists of these steps [5] :

1. Determine the number of k-clusters (randomly)
2. Generate random values for the cluster centroid as many as k-clusters.
3. Calculate the distance of each input data to each centroid using Euclidean Distance to find the closest distance from each data with the centroid.
4. Classify each data based on its proximity to the centroid (the smallest distance).
5. Update centroid values. The new centroid value is obtained from the average cluster
6. Repeat the steps until the members of each cluster have nothing to change.
7. If the above steps have been fulfilled, then the average value of the cluster center in the last iteration will be used as a parameter to determine the data classification

C. Distance Measurement

In this study Euclidean distance measurement will be carried out for both k-medoids and k-means algorithms. This type of distance measurement is commonly used in data mining. In euclidean distance, the distance between two points defined as a straight line. The Euclidean distance calculated by (1) [18]:

$$D_{Euclidean}(x_i, x_j) = \sqrt{(x_i - x_j)^2} = \sqrt{\sum_{m=1}^{n}(x_{im} - x_{jm})^2} \quad (1)$$

$D_{Euclidean}(x_i, x_j)$	: Euclidean Distance
x_i	: Data -i
x_j	: Data -j
x_{im}	: Data -i atribute -m
x_{jm}	: Data -j atribute -m

In research conducted by Mohibullah et al. [19] who tested euclidean and manhattan distance using k-medoid, it was concluded that the two distance calculations gave the same good results but were still dependent on the inputted data. Small classified data is better if using Euclidean and large classified data (big data) is better to use Manhattan.

D. Silhouette Score

In clustering, determining the number of clusters is one of the most important steps. Silhouette coefficient is a method used to evaluate clusters and see the quality of placement of data in a cluster. Silhouette score is important to see whether the cluster produced is of good quality. The stages of calculating silhoutte coefficients are as follows [16]:

1. Calculate the average distance of the i object to all objects in the group. We call the average distance a (i).
2. Calculate the average distance of the i object to all objects in another cluster we call b (i), and take the smallest value.
3. The silhoutte coefficient value is obtained by (2):

$$S(i) = \frac{b(i) - a(i)}{\max(b(i), a(i))} \quad (2)$$

And can be written with (3):

$$S(i) = \begin{cases} 1 - \frac{a(i)}{b(i)} & , if\ a(i) < b(i) \\ 0 & , if\ a(i) = b(i) \\ \frac{b(i)}{a(i)} - 1 & , if\ a(i) > b(i) \end{cases} \quad (3)$$

Where:

S(i) :Silhouette score

a(i) :average distance between data i and all objects in the cluster.

b(i) :average distance between data i to all objects in another cluster

The range of values from silhoutte coefficient is -1 to 1. If the silhoutte coefficient value is close to 1 then the object is in the right cluster, if it is around 0 then the object can be between 2 clusters, and if the result is negative then the object may be in the wrong cluster [17]. The best number of clusters or the optimum bumber of clusters is the number of clusters with the highest average silhouette score where the average is taken from the value of silhouette of each cluster.

III. RESEARCH METHOD

TABLE I.
SYSTEM DESIGN

System Design
Input: USGS earthquake dataset (2014-2018) **Output:** Cluster analysis using k-medoids algorithm Comparison between k-medoids and k-means algorithm
Preprocessing Data: Missing data imputation Normalization using min-max **Algorithm:** K-Medoids algorithm K-Means algorithm **Validation:** Silhouette Average using k=2,3,4,5,6,7,8,9,10

From the Table I, it can be seen that there are several stages of the system. First step is to accumulate the earthquake data from United State Geological Survey (USGS). Then the preprocessing needs to be done to prepare the data before being used for clustering. The next step is the clustering stage where two types of analysis will be carried out, namely cluster analysis using k-medoids to find spatial pattern of seismic acitivity and also comparing the cluster results with the k-means algorithm.

A. Dataset

The dataset used is earthquake data throughout 2016-2018 obtained from the United States Geological Survey (USGS). The earthquake dataset has 5 variables: latitude, longitude, magnitude, depth, and dmin. Latitude and longitude are spatial data that show the location of the earthquake epicenter. Magnitude is a value that shows the strength of an earthquake on the Richter scale. In this study, only magnitude above 5 Ms are considered for clustering. The depth of the earthquake used in this study ranged from 0-700 km. Whereas dmin is the minimum distance of the station (nearest station) to the epicenter's center (in degrees). One degrees is approximately 111.2 km. The smaller the dmin number, the calculated depth of the earthquake will be more reliable. Table II is a sample dataset (un-normalized) used:

TABLE II
DATASET BEFORE NORMALIZATION

Latitude	Longitude	Depth (km)	Mag (SR)	Dmin (km)
1.6507	126.3645	40.15	5.1	1.329
-2.677	102.349	166	5.7	1.778
1.0774	97.3382	25.91	5.2	0.327
5.9998	126.9477	86.65	5.3	1.725
5.8983	126.9209	60.21	7	1.769
-1.4498	134.0858	41	5.8	2.345
1.1186	126.4533	35	5.4	0.976
-0.4765	99.732	98.03	5	1.531
-8.3086	116.7887	10	5.1	2.613

A total of 962 earthquakes with magnitudes above 5 Ms and depths of 0-700 km occured in Indonesia throughout 2014-2018, where Indonesia span from 6º N - 11º S and 95º W - 141º E. From the results of plotting the spatial data, it can be seen that the earthquake occurred in almost all parts of Indonesia except Kalimantan Island. This earthquake data will then be clustered so that the areas that are most prone to earthquakes will be seen. The earthquake distribution is shown in the Fig. 2:

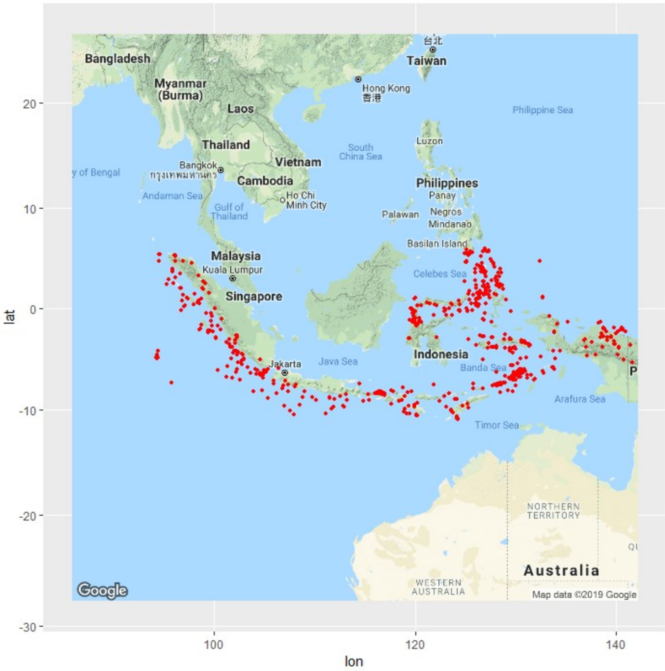

Fig. 2. Earthquakes that occur in Indonesia and its surroundings with magnitude ≥ 5.0 Richter scale during 2014-2018

B. *Preprocessing*

Data preprocessing is often neglected while it is an important process in data mining [20]. In earthquake data used as the dataset, there are some missing values and each variable has a different range of values so preprocessing needs to be done. In this study, the preprocessing carried out is missing data imputation and also min max normalization given by (5):

$$X_i' = \left(\frac{X_i - Min_x}{Max_x - Min_x}\right) \qquad (5)$$

Where :

X_i' : data after normalization (results)

X_i : data before normalization

Min_x : minimum value in variable x

Max_x : maximum value in variable x

IV. RESULTS AND DISCUSSION

A. Cluster Analysis Using K-Medoids

Clustering using k-medoids was conducted on 962 earthquake data in Indonesia and the number of clusters used was 2 to 10 clusters. After determining the number of clusters and also the centroid of each cluster, the distance from the centroid to the nearest non-centroid earthquake data is calculated using Euclidean Distance, which formula can be seen in Equation (1). The Fig.3 is the results of the silhouette average obtained for each cluster:

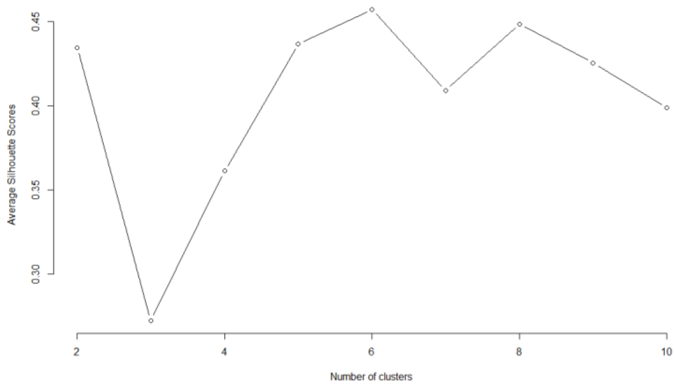

Fig. 3. Silhouette Average Graph from K-Medoids Algorithm

In Fig. 3, it is seen that the lowest value of silhouette average using k-medoids is below 0.30 with k=3 and the highest value is above 0.45 with k= 6. The results of the silhouette average for k-medoids clustering using all data shown in the Table III:

TABEL III
SILHOUETTE AVERAGE USING ALL DATA

No	Number of Clusters	Silhouette Average Using K-Medoids
1	2	0.4345131
2	3	0.2723350
3	4	0.3613861
4	5	0.4368691
5	6	0.4574067
6	7	0.4092505
7	8	0.4486897
8	9	0.4256167
9	10	0.3988550

In Table III, it is known that the highest silhouette average value is 0.4574067 with k=6. The second best silhouette value obtained is 0.4486897 with the number k equal to 8. Fig. 4. is the result of cluster mapping using k-medoids with number of clusters equal to 6 as the cluster with the highest silhouette average:

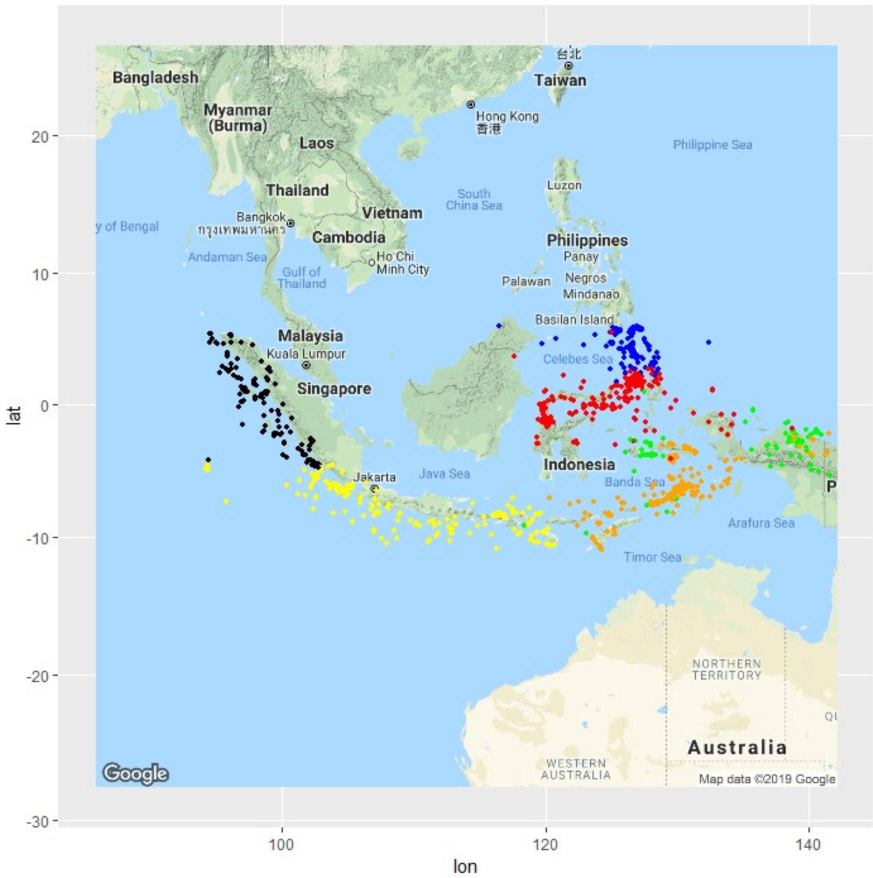

Fig. 4. Results of Mapping K-Medoids Clusters with k=6

Fig. 4. is the result of cluster mapping using k-medoids where the color of cluster one is yellow, cluster two is blue, cluster three is red, cluster four is orange, cluster five is green and cluster six is black.

The first cluster consists of 184 earthquakes that occurred along Lampung, Java, West Nusa Tenggara,East Nusa Tenggara and surrounding. This cluster has an earthquake depth average of 77.91173 km which is a medium depth earthquake (0-70 km). About 83% of earthquakes in this cluster have depth between 6-70 km. The average of earthquake magnitude in this cluster is 5.30434 ms. The earthquake with the highest magnitude in this cluster is 7.8 Ms which occurred in the Mentawai Islands in 2016.

Cluster two consists of 110 earthquakes that occured in the Sulawesi sea and its surroundings. The average depth of earthquake in cluster two is 77.91173 km which is intermediate earthquakes (70-300 km). The average magnitude of this cluster is 5.256364 ms with the largest earthquake recorded was 7 Ms which occured around the Philippine archipelago.

Cluster three has the most members. There are 257 earthquakes occured in North Sulawesi, Central Sulawesi, North Maluku, and the surrounding areas. About 82% of the earthquakes that occur in this cluster are shallow earthquakes with a depth average of 46.05502 km. This cluster has the highest magnitude average compared to other clusters which is equal to 5.337354 ms. The major earthquake that occurred in this cluster was the Palu earthquake with the magnitude of 7.4 ms in 2018. Shallow earthquakes generally tend to be more damaging than deeper quakes. Seismic waves from deep quakes have to travel farther to the surface, losing energy along the way. Quoted from the geological magazine managed by the Ministry of Energy and Mineral Resources of Indonesia (2016), the shallow earthquake has the potential to cause disaster even though its magnitude is small. The example of shallow earthquake is the Aceh earthquake in 2004 with magnitude 9.1 ms and the Japanese earthquake in 2011 with a magnitude of 9 ms. Both earthquakes happened under a 60 km deep (shallow earthquake).

Cluster four has as many as 200 earthquakes that occured along the Banda Sea and its surroundings. This cluster has an average of intermediate earthquake depth (70-300 km) which is 83.64965 km. The average of earthquake magnitude in this cluster is 5.2325 ms.

While cluster five has the fewest members. There are 87 earthquakes occured around Maluku, Papua and its surroundings. This cluster has the lowest depth average of 30.86644 km and has the second largest average of earthquake magnitude after cluster 3 which is equal to 5.322989 Ms.

And the last cluster is cluster six that has 124 earthquakes that occur along the island of Sumatra. This cluster has an average of magnitude of 5.287903 ms with a depth average of 47.15347 km. A summary of cluster analysis is in the Table IV:

TABLE IV
K-MEDOIDS CLUSTER RESULT WITH K=6

No. Clus	Average of depth (Km)	Average of mag (Ms)	Average of dmin (km)	Total of Events
1	45.91935	5.304348	2.411092	184
2	77.91173	5.256364	2.597464	110
3	46.05502	5.337354	1.675405	257
4	83.64965	5.2325	2.09162	200
5	30.86644	5.322989	5.567034	87
6	47.15347	5.287903	1.625855	124

B. K-Medoids vs K-Means Analysis

In this study, clustering with k-medoids and k-means was carried out using different number of clusters. The number of clusters used is 2 to 10. In addition, variations in the number of datasets are also used as input data for clustering. Table V are the results of the clustering obtained:

TABLE V
HIGHEST SILHOUETTE AVERAGE USING K-MEDOIDS AND K-MEANS ALGORITHM

No	Numbers of Data	Highest Silhouette Average Using K-Medoid	Highest Silhouette Average Using K-Means
1	300	0.43645877	0.3951920
2	550	0.4945051	0.3555890
3	750	0.5121741	0.3670343
4	900	0.5375306	0.3736186
5	962	0.4588860	0.3750746

In Table V, a silhouette average calculation for both k-medoids and k-means clustering results is performed by using different numbers of data (300,550,750,900, and 962 (all dataset)). The number of clusters tested is two to ten clusters. It can be seen that the silhoutte average obtained using k-medoids is better than the silhouette average using k-means in all numbers of data. The highest silhouette average using k-medoids for all data obtained at 0.4588860, while for k-means only at 0.3750746. The following are the complete results of clustering of all datasets using k-means:

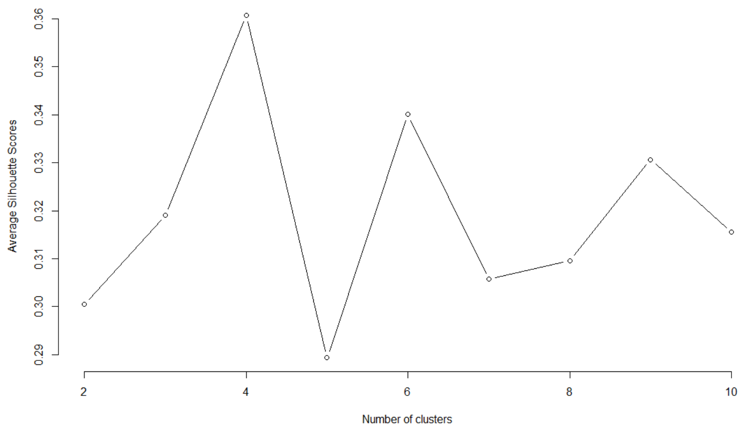

Fig. 5. Silhouette Average Graph from K-Means Algorithm Using All Data

In Fig 5, it is seen that the lowest value of silhouette for k-means is around 0.29 with k= 5 and the highest value is around 0.36 for k= 4. The results of the silhouette average for all data using k-means shown in the Table VI:

TABEL VI
SILHOUETTE AVERAGE USING ALL DATA

No	Number of Clusters	Silhouette Average Using K-Means
1	2	0.3005488
2	3	0.3190718
3	4	0.3607622
4	5	0.2893766
5	6	0.3400689
6	7	0.3057094
7	8	0.3096252
8	9	0.3306172
9	10	0.3155021

If we compare Table VI with Table III, it can be seen that the silhouette average value obtained for all cluster numbers is higher when using k-medoids. The results show that for k-means, the highest value of silhouette is 0.3607622 with k=4. While the highest silhouette value for k-medoids is 0.4574067 with the number of clusters equal to 6 (Table III). Fig.5 showing the plot results from clustering using both K-Medoids and K-Means method:

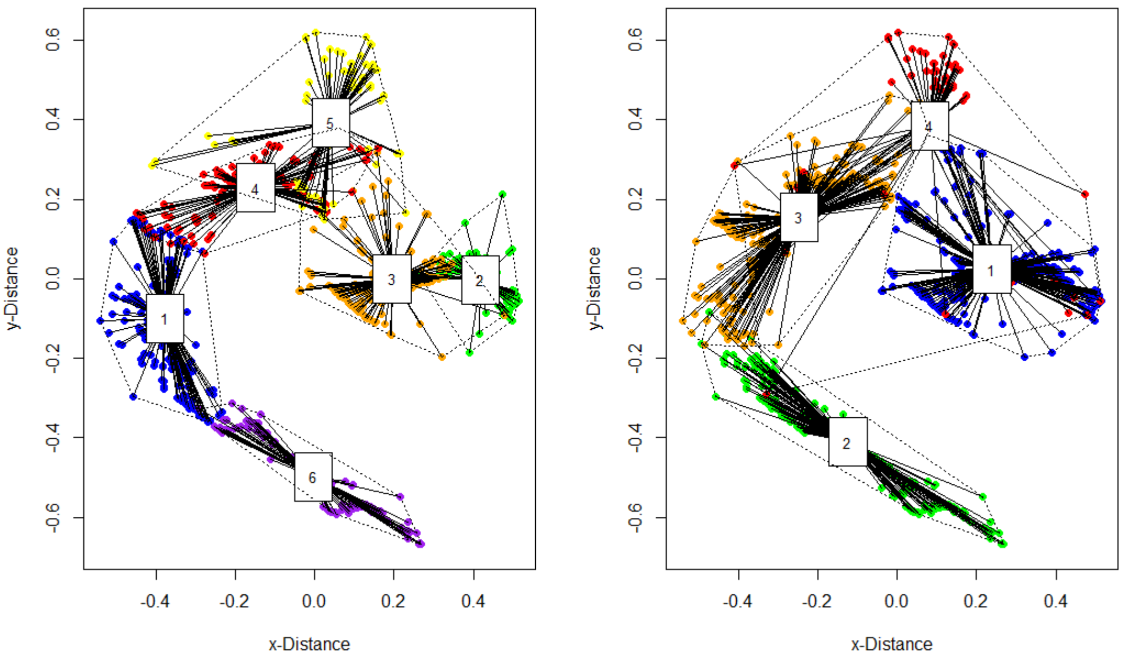

Fig. 6. Plot results of k-medoids clusters (left) and k-means clusters (right)

Fig. 6 is the result of k-medoids and k-means plotting using the number of clusters with the highest silhouette value for each method, where k-medoids uses k=6 and k-means uses k=4. The K-Means algorithm's clusters results presented in Fig. 5 (right) showing the overlapping of clusters while the results of K-Medoids in Fig. 5 (left) showing less overlapping compared to K-Means. From Fig. 5, it can also be seen how some data in the k-means cluster are not in the right cluster. As illustrated by cluster 4 on k-means, several cluster members are far from the center of the cluster. This causes a decrease in the value of the silhouette average.

V. Conclusion

From this study, it can be concluded that:

1. Clustering using k-medoids was successfully performed. By using 962 earthquake data, the best number of clusters is 6 clusters where the silhouette average value obtained was 0.4574067. From the results of the cluster analysis, cluster 3 is the area that prone of earthquake the most because it has an average of depth below 70 km (shallow earthquake), the highest magnitude average compared to other clusters which is equal to 5.337354 ms, and has the most earthquake event data throughout 2014-2018.
2. From this study, it was also found that k-medoids performed better than k-means. Cluster results obtained are better when using k-medoids. The value of the silhouette average is higher using k-medoids than using k-means. With this result, it can be concluded that k-medoids is better than k-means as a tool for earthquake cluster analysis.

Earthquake is indeed unpredictable, but with this research it is expected that the areas prone to earthquakes can improve their disaster mitigation efforts. Future studies are expected to use larger data to increase the value of the silhouette average obtained so that it can produce a better cluster and also compare it with other clustering techniques.

REFERENCES

[1] Sugito, T., Nanin (2008). Tsunami (31-32).Universitas Pendidikan Indonesia.

[2] Gongqian, Xu., Hanghang, He., Ying ,Tao., Yang, Wang. (2005) Earthquake. Roskilde Universitetscenter's Digitale Arkiv.

[3] Malizia, D. (2018) A Visual Guide to Volcanoes and Earthquake. The Rosen Publishing Group, Inc. New York.

[4] Clifton, Christopher (2010). Encyclopædia Britannica: Definition of Data Mining.

[5] Han, J., Kamber, M., Pei, J. (2012) Data Mining Concept and Techniques, 3rd ed. Morgan Kaufmann-Elsevier, Amsterdam.

[6] Bharati, M., & Ramageri, M. (2010). Data Mining Techniques an Applications. Indian Journal of Computer Science and Engineering. 1(4), 301-305.

[7] Aggarwal, C., Reddy, C.: Data Clustering: Recent Advances and Applications. Data Mining and Knowledge Discovery Series. Chapman and Hall/CRC, Taylor & Francis Group, Boca Raton (2013)

[8] Kamat, R.,K., Kamath, R.,S. (2017) Earthquake Cluster Analysis: K-Means Approach. Journal of Chemical and Pharmaceutical Sciences 10 (1)

[9] Savaş, C., Yıldız, M. S., Eken, S., İkibaş, C., & Sayar, A. (2019). Clustering Earthquake Data: Identifying Spatial Patterns From Non-Spatial Attributes. In Big Data and Knowledge Sharing in Virtual Organizations (pp. 224-239). IGI Global.

[10] Novianti, Pepi, Dyah Setyorini, & Rafflesia, Ulfasari. (2017) "K-Means cluster analysis in earthquake epicenter clustering." International Journal of Advances in Intelligent Informatics 3.2. 81-89.

[11] Soni, K. G., & Patel, A. (2017). Comparative Analysis of K-means and K-medoids Algorithm on IRIS Data. International Journal of Computational Intelligence Research, 13(5), 899-906.

[12] Arora, P., & Varshney, S. (2016). Analysis of k-means and k-medoids algorithm for big data. Procedia Computer Science, 78, 507-512.

[13] Selvi, H. Z., & Çağlar, B. (2016). Using K-Means and K-Medoids Methods for Multivariate Mapping. International Journal of Applied Mathematics, Electronics and Computers, 4(Special Issue-1), 342-345.

[14] Kaur, N. K., Kaur, U., & Singh, D. D. (2014). K-Medoid clustering algorithm-a review. International Journal of Computer Application and Technology (IJCAT), 1(1), 2349-1841.

[15] Larose, Daniel T. 2005. Discovering Knowledge in Data : An Introduction to Data Mining. John Willey & Sons, Inc

[16] Kaufman, L. and Rousseeuw, P.J. (1987), Clustering by means of Medoids, in Statistical Data Analysis Based on the Norm and Related Methods, edited by Y. Dodge, North-Holland, 405–416.

[17] Prasetyo, Eko. (2014) Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab. Yogyakarta: ANDI

[18] Hertz, T. (2006). Learning Distance Functions: Algo Rithms and Applications. Hebrew University.

[19] Mohibullah, M., Hossain, M. Z., & Hasan, M. (2015). Comparison of euclidean distance function and manhattan distance function using k-mediods. International Journal of Computer Science and Information Security, 13(10), 61.

[20] García, S., Luengo, J., & Herrera, F. (2015). Data preprocessing in data mining (pp. 195-243). Switzerland: Springer International Publishing.