PROC. INTERNAT. CONF. SCI. ENGIN. ISSN 2597-5250
Volume 3, April 2020 | Pages: 3-7 E-ISSN 2598-232X

Cluster Analysis of Earthquake's Data Clustering in Indonesia using Fuzzy K-means Clustering

Ayyubi Ahmad*, Nursyiva Irsalinda
Mathematics Department, Faculty of Applied Science and Technology, Universitas Ahmad Dahlan Yogyakarta,
Jl. Ringroad Selatan, Kragilan, Tamanan, Banguntapan, Bantul, Yogyakarta 55191, Indonesia
Email*: ayyubiahmad1@gmail.com

Abstract. The earthquake is shocks or vibrations in the earth's surface because of shifting layers of rock at the base of the earth's surface. This natural phenomenon is common in Indonesia because it lies between Australian, Eurasian, Pacific plates, and it location surrounded by a ring of fire precisely. Therefore, this study aims to cluster earthquake events in Indonesia and describe the characteristics of each group based on clustering results. The method used is the Fuzzy K-Means Clustering. The clustering results obtained from clustering based on the depth, longitude, and latitude. In this study, the data used is the earthquake's data, which has a magnitude greater than or equal to 5 SR and only clumped by depth. Based on the Davies-Bouldin and Dunn index, the best clustering is 2 clusters which researchers cluster earthquake data in Indonesia into deep and shallow clusters.

Keywords: Cluster Analysis, Earthquake, Fuzzy K-Means Clustering

INTRODUCTION

Indonesia is one of the largest countries in the world. Indonesia is a country that lies between Australian, Eurasian, Pacific plates, and it's location is precisely surrounded by ring of fire. This has resulted in frequent earthquakes in Indonesia. The earthquake is shocks or vibration in earth's surface because of shifting layers of rock at the base of the earth's surface and causing damage even exacting heavy casualties. In Indonesia, at least twice the magnitude of an earthquake in a very short time was June 24th, 2019 in the city of Saumlaki with a magnitude of 7,3 Richter Scale (M_L) and was July 14th, 2019 in the city of Laiwui with a magnitude of 7,2 M_L. In that case, the author wants to cluster the earthquake data by using a cluster analysis. Cluster analysis is a class of techniques that are used to grouping a set of objects or cases in such a way that objects/cases in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters). A cluster must be observed in its homogeneity and heterogeneity. A good cluster should have homogeneous properties in groups where objects in each group have tiny variations and heterogeneous properties between groups which objects that are in different groups have different properties. There are some assumption that correspond to cluster analysis such as normality, linearity, and homoscedasticity. Because there are some difficult cases to meet these assumptions, then it needs another cluster technique that is easy to process. One method that can be used is Fuzzy K-Means Clustering. A Fuzzy K-Means Clustering is a method of analyzing data or data-mining that conducts a modeling process without supervision and it's one of those methods of clustering data with a partition system. A Fuzzy K-Means Clustering is one of the methods of data relating non-hierarchical data or partitional clustering. K-Means Clustering method trying to put data into groups, where the data in the same group is characteristic of each other, and have characteristics that are different from data in other groups. In other words, K-Means Clustering method is aimed at minimizing objective functions gathered by minimizing variations between the data in some clusters and maximize variations with the data in other clusters. A Fuzzy K-Means Clustering could be implemented in earthquake event. This study discusses to cluster of earthquake event in Indonesia using fuzzy k-means clustering to get more information about the earthquake and minimizing the adverse effects.

MATERIALS AND METHODS

The research method used in this research is library research that discusses activities related to the method of collecting library data, reading, storing, and processing research materials.

The Object of Research

The object used in this study is earthquake data in Indonesia from January 1st to September 26th, 2019.

The Place of Research

The place of study: the campus library of the 4th Universitas Ahmad Dahlan Yogyakarta.

The Variable of Research

The variable used is the position of latitude, longitude, depth, and magnitude of earthquakes.

The Stage of Research

Searching and collecting data from various sources both from books and the internet. Then grouping the data using Fuzzy K-Means Clustering.

RESULTS AND DISCUSSION

Description of Earthquake Events

The earthquake event data obtained from the page https://earthquake.usgs.gov/earthquakes/search/. In this study the researchers observed data based on earthquake magnitude and locations including coordinates, namely longitude, latitude, and depth. Geographically, the area of Indonesia that will be observed based on data obtained by the coordinates of earthquake events are in between 6˚NL - 11˚SL and 95˚EL - 141˚EL. During the period January 1st to September 26th, 2019 there were 196 earthquake events that occurred in Indonesia with magnitudes greater than 5 and depths of up to km. Based on the data obtained, the point of earthquake occurrence in Indonesia during this period can be seen through the following figure:

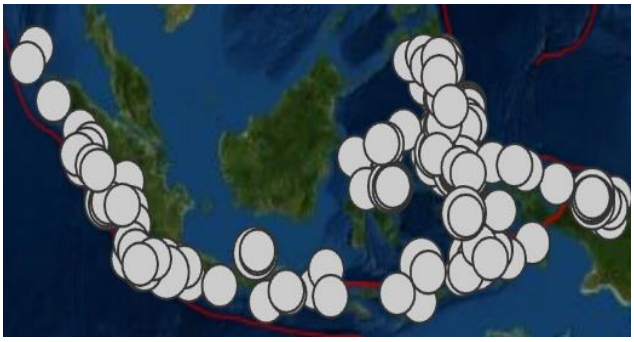

Figure 1. The spread of earthquakes with magnitude ≥ 5 for the period of January 1 – September 26, 2019 based on the page https://earthquake.usgs.gov/earthquakes/search/.

Based on its depth, it's carried out on earthquake events with a depth of 0 km to 626,1 km. In this case the researchers divides the depth into three equal parts. A total of 190 earthquakes occurred at a depth of 0 km to 208,7 km, a total of 2 events at a depth of 208,7 km to 417,4 km, and as many as 4 events at a depth of 417,4 km to 626,1 km. Based on data obtained that the minimum depth occurred on September 14th, 2019 in Laiwui with a depth of 4,5 km.

Earthquake Data Grouping Using Fuzzy K-Means Clustering

Earthquake data grouping uses fuzzy k-means clustering based on the position of the epicenter, namely latitude, longitude, and depth. In this study there are several steps in clustering using k-means clustering.

The algorithm of K-means clustering:

1. Specify *k* as the number of clusters you want to form
2. Allocate data into clusters randomly
3. Determine the cluster center of the data that exist in each cluster with the equation:

$$C_{kj} = \frac{x_{1j} + x_{2j} + \cdots + x_{nj}}{n}$$

where

C_{kj} = the center of the *k*-cluster on the variable-*j*th $(j = 1,2,..,p)$

n = the number of data in the *k*-cluster

4. Determine the distance of each object to each centroid by calculating the distance of each object to each centroid using Euclidean distance

$$d(X_i, X_g) = \sqrt{\sum_{j=1}^{p} (X_{ij} - X_{gj})^2}$$

5. Calculate objective functions with formulas

$$J = \sum_{i=1}^{n} \sum_{j=1}^{k} a_{ij} d(x_i, C_{kj})^2$$

6. Allocate each data to the nearest centroid/average formulated as follows

$$a_{ij} = \begin{cases} 1, & s = min\{d(x_i, C_{kj})\} \\ 0, & otherwise \end{cases}$$

a_{ij} is the membership value of point x_i to cluster center C_{kj}, s is the shortest distance from data x_i to cluster center C_{kj} after comparison.

7. Repeat steps 3-6 until there is no object movement or there is no change to the objective function.

Grouping of Earthquake Data with 2 Clusters

By using 25 iterations it's obtained clustering as recommended by the following figure:

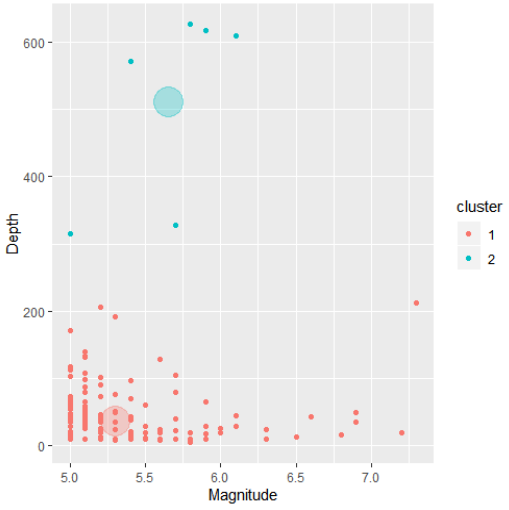

Figure 2. The Clustering Earthquake Data with 2 Cluster.

It can be seen from the picture above that cluster 1 consist of 190 members with centroids at 3,687958 °S, 122,3467 °E, and with a depth of 36,93526 km, while cluster 2 consist of 6 members with centroids at 5,561167 °S, 117,6063 °E, and with a depth of 511,38333 km.

Table 1. The Spread of Earthquake Events with 2 Clusters.

Magnitude	Number of Clusters	
	1	**2**
>=7	2	0
6-7	9	1
5-6	179	5

In cluster 1, earthquake events are scattered at each magnitude interval and most with magnitudes between 5 M_L to 6 M_L. However, in cluster 2 earthquake events with magnitudes greater than or equal to 7 M_L never occurred. As many as 96,94% of the earthquake events occurred in cluster 1 which is the number of earthquakes in cluster 1 more than cluster 2.

Table 2. The Centroid Earthquake Events with 2 Clusters.

Cluster	Latitude	Depth	Longitude
1	3,687958	36,93526	122,3467
2	5,561167	511,38333	117,6063

If seen from the centroid, cluster 1 contains shallow earthquakes and cluster 2 contains deep earthquakes.

Grouping of Earthquake Data with 3 Clusters

By using 25 iterations it's obtained clustering as recommended by the following figure:

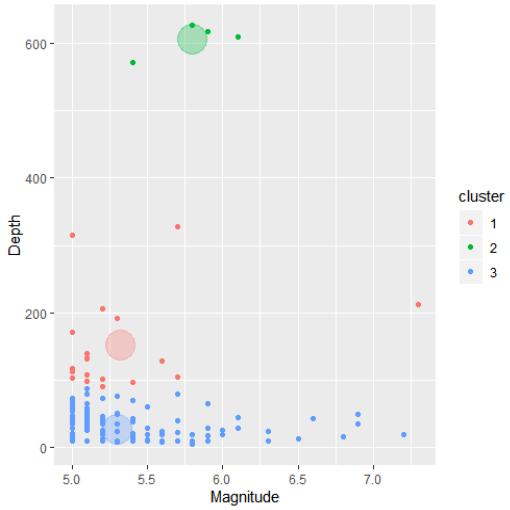

Figure 3. The Clustering Earthquake Data with 3 Cluster.

It can be seen from the picture above that cluster 1 consist of 19 members with centroids at 5,816947 °S, 127,6147 °E, and with a depth of 151,95789 km, cluster 2 consist of 4 members with centroids at 6,345 °S, 113,2858°E, and with a depth of 606,425 km, and cluster 3 consist of 173 members with centroids at 3,457671 °S, 121,8132 °E, and with a depth of 27,59017 km.

Table 3. The Spread of Earthquake Events with 3 Clusters.

Magnitude	Number of Clusters		
	1	**2**	**3**
>=7	1	0	1
6-7	0	1	9
5-6	18	3	163

In cluster 3, earthquake events are scattered at each magnitude interval and most with magnitudes between 5 M_L to 6 M_L. However, in cluster 2 earthquake events with magnitudes greater than or equal to 7 M_L never occurred and in cluster 1 earthquake events with magnitudes between 6 M_L to 7 M_L never occurred. As many as 88,27% of the earthquake events occurred in cluster 3 which is the number of earthquakes more than any other cluster.

Table 4. The Centroid Earthquake Events with 3 Clusters.

Cluster	Latitude	Depth	Longitude
1	5,816947	151,95789	127,6147
2	6,345000	606,42500	113,2858
3	3,457671	27,59017	121,8132

If seen from the centroid, cluster 1 and 3 contains shallow earthquakes and cluster 2 contains deep earthquakes.

Grouping of Earthquake Data with 4 Clusters

By using 25 iterations it's obtained clustering as recommended by the following figure:

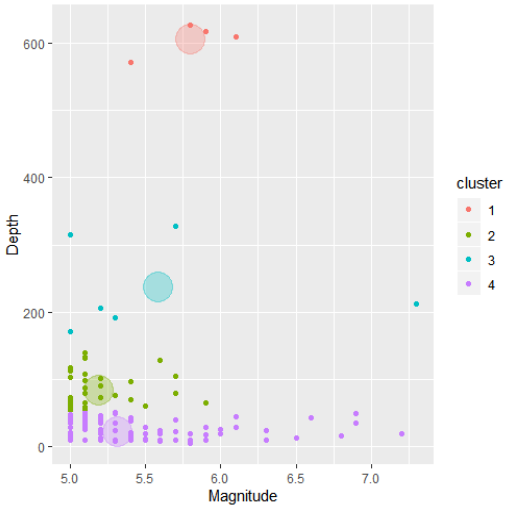

Figure 4. The Clustering Earthquake Data with 4 Cluster.

It can be seen from the picture above that cluster 1 consist of 4 members with centroids at 6,345 °S, 113,2858 °E, and with a depth of 606,425 km, cluster 2 consist of 33 members with centroids at 5,387788 °S, 121,3542 °E, and with a depth of 84,73939 km, cluster 3 consist of 6 members with centroids at 5,743833 °S, 128,2653°E, and with a depth of 237,3 km, and cluster 4 consist of 153 members with centroids at 3,244699 °S, 122,3796 °E, and with a depth of 22,48431 km.

Table 5. The Spread of Earthquake Events with 4 Clusters.

Magnitude	Number of Clusters			
	1	**2**	**3**	**4**
>=7	0	0	1	1
6-7	1	0	0	9
5-6	3	33	5	143

In cluster 4, earthquake events are scattered at each magnitude interval and most with magnitudes between 5 M_L to 6 M_L. In cluster 1 and 2 earthquake events with magnitudes greater than or equal to 7 M_L never occurred. Whereas, in cluster 2 and 3 earthquake events with magnitudes between 6 M_L to 7 M_L never occurred. Also in cluster 2 only earthquakes occurred with a magnitude of between 5 M_L to 6 M_L. As many as 78,06% of the earthquake events occurred in cluster 4 which is the number of earthquakes more than any other cluster.

Table 6. The Centroid Earthquake Events with 4 Clusters.

Cluster	Latitude	Depth	Longitude
1	6,345000	606,42500	113,2858
2	5,387788	84,73939	121,3542
3	5,743833	237,30000	128,2653
4	3,244699	22,48431	122,3796

If seen from the centroid, cluster 2 and 4 contains shallow earthquakes, cluster 3 contains medium earthquakes, and cluster 1 contains deep earthquakes.

The Validity of Clustering

If the earthquake data has been clustered, researchers need to test the validity of the clustering. Cluster validity test aims to measure the accuracy of data clustering. In this case, the validity test used is the Davies-Bouldin and Dunn index.

The Davies-Bouldin Index (DBI)

The Davies-Bouldin Index (DBI) was introduced by David L. Davies and Donald W. Bouldin in 1979 and was used to evaluate clustering algorithms. Using the Davies-Bouldin validity index can identify compact clusters that are well separated from other clusters as well as the Dunn validity index.

Let us denote by $ì_k$ the mean distance of the points belonging to cluster C_k to their barycenter $G^{\{k\}}$:

$$\mu_k = \frac{1}{n_k}\sum_{i\in I_k}\left\|M_i^{\{k\}} - G^{\{k\}}\right\|$$

Let us also denote by

$$\Delta_{kk'} = d\left(G^{\{k\}}, G^{\{k'\}}\right) = \left\|G^{\{k'\}} - G^{\{k\}}\right\|$$

the distance between the barycenters $G^{\{k\}}$ and $G^{\{k'\}}$ of clusters C_k and $C_{k'}$.

One computes for each cluster k, the maximum M_k of the quotients $\frac{\mu_k+\mu_{k'}}{\Delta_{kk'}}$ for all indices $k' \neq k$. The Davies-Bouldin index is the mean value, among all the clusters of the quantities M_k:

$$\omega = \frac{1}{K}\sum_{k=1}^{K} M_k = \frac{1}{K}\sum_{k=1}^{K} \underset{k'\neq k}{max}\left(\frac{\mu_k+\mu_{k'}}{\Delta_{kk'}}\right)$$

A smallest Davies-Bouldin index indicates the best clustering.

The Dunn Index (DI)

The Dunn index (DI) was introduced by J. C. Dunn in 1974 and it's a metric for evaluating clustering algorithms. This's part of a group of validity indices including the Davies-Bouldin index, in that's an internal evaluation scheme, where the result's based on the clustered data itself.

Let us denote by ψ_{min} the minimal distance between points of different clusters and ψ_{max} the largest within-cluster distance.

The distance between clusters C_k and $C_{k'}$ is measured by the distance between their closest points:

$$\psi_{kk'} = \underset{\substack{i\in I_k \\ j\in I_{k'}}}{min}\left\|M_i^{\{k\}} - M_j^{\{k'\}}\right\|$$

and ψ_{min} is the smallest of these distances $\psi_{kk'}$:

$$\psi_{min} = \underset{k\neq k'}{max}\,\psi_{kk'}$$

For each cluster C_k, let us denote by D_k the largest distance separating two distinct points in the cluster:

$$D_k = \underset{\substack{i\in I_k \\ j\in I_{k'}}}{max}\left\|M_i^{\{k\}} - M_j^{\{k'\}}\right\|$$

Then ψ_{max} is the largest of these distances D_k:

$$\psi_{max} = \underset{1\leq k\leq K}{max}\, D_k$$

The Dunn index is defined as the quotient of ψ_{min} and ψ_{max}:

$$\omega = \frac{\psi_{min}}{\psi_{max}}$$

A highest Dunn index indicates the best clustering.

Based on the data, we calculated the DB and Dunn index values for each cluster and obtained as follows:

Table 7. The DBI and DI for Several Clusters.

Number of Clusters	DBI	DI
2	0,330	0,333
3	0,456	0,076
4	0,530	0,013
5	0,611	0,034
6	0,635	0,028
7	0,634	0,015
8	0,582	0,015
9	0,724	0,014
10	0,533	0,051
11	0,543	0,034
12	0,588	0,061
13	0,566	0,034
14	0,555	0,041
15	0,562	0,042

Based on the DBI, the smallest value is 0,330 which means that clustering earthquake data into two clusters is the best clustering. Whereas based on the DI, the largest value is 0,333 which means that clustering earthquake data into two clusters is the best clustering. It can be seen from the two indices that it has the same decision that the best clustering is to divide the earthquake data into two clusters.

Discussion

In this study, there are 4 attributes including: magnitude, latitude, depth, and longitude. But, the clustering results show that the clustering was formed based on depth only. At the beginning the researchers determined 2, 3, and 4 clusters to see which cluster was the best, so that it could be used to cluster earthquake data in Indonesia. From the results obtained that using 2 clusters is the best. Because 2 clusters have the smallest Davies-Bouldin and the largest Dunn index value among the other cluster index values. Then researchers cluster the earthquake's data in Indonesia into deep and shallow clusters.

CONCLUSIONS

In this study, researchers aims to cluster earthquake events in Indonesia and describe the characteristics of each group based on clustering results. Clustering is only based on depth and uses the Fuzzy K-means Clustering. From some clustering, the best cluster is obtained based on the validity of the cluster by using the Davies-Bouldin and Dunn index, which is clustering earthquake data into 2 clusters. Then the researchers set these 2 clusters into deep and shallow clusters.

REFERENCES

Afrin, Fahmida and Md. Al-Amin *et al*. 2015. Comparative Performance of Using PCA with K-Means and Fuzzy C Means Clustering for Customer Segmentation. IJSTR, Vol. 4, ISSUE 10, 70-74.

Ahmar, Ansari Saleh and Darmawan Napitupulu *et al*. 2018. Using K-Means Clustering to Cluster Provinces in Indonesia. Journal of Physics: Conference Series: 1028, 1-6.

Cebeci, Zeynel and Figen Yildiz. 2015. Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures. JAI, Vol. 6, No.3, 13-23.

EMC Education Services. 2015. Data Science & Big Data Analytics. John Wiley & Sons, Inc., U.S.A.

Hot, Elma and Vesna Popović-Bugarin. 2016. Soil Data Clustering by using K-means and Fuzzy K-means Algorithm. Telfor Journal, Vol. 8, No.1, 56-61.

Kalaivani, E. Rama and G. Suganya *et al*. 2017. Review on K-Means and Fuzzy C Means Clustering Algorithm. IJIR, Vol-3, Issue-2, 1822-1824.

O., Enikuomehin A. and Rahman M. A. *et al.* 2016. Fuzzy K-means Application to Semantic Clustering for Image Retrieval. Journal of Advances in Computing, 6(1), 1-5.

Ross, Timothy J. 2010. Fuzzy Logic with Engineering Applications Third Edition. John Wiley & Sons, Inc., U.S.A.

USGS Science for A Changing World. 2019. *Earthquake's Data in Indonesia*, November 2019. https://earthquake.usgs.gov/earthquakes/search/.

Wu, Junjie. 2012. Advances in K-means Clustering A Data Mining Thinking. Springer, Verlag Berlin Heidelberg.

Xu, Jinglin and Junwei Han *et al*. 2016. Robust and Sparse Fuzzy K-Means Clustering. IJCAI-16, 2224-2230.

THIS PAGE INTENTIONALLY LEFT BLANK