Web Application of Classification of Potential Earthquake Region in Central Sulawesi

Jihad Nurul Islami(✉), Rohis Rachman, Yayang Matira, Junaidi, and Fadjryani

Statistics Department, Mathematics and Natural Sciences Faculty, Tadulako University, Palu, Indonesia
jihadnurul16@gmail.com

Abstract. An earthquake is a shaking incident of the earth due to the movement or shift of rocks in the lithospheric part. This occurs suddenly due to the movement of tectonic plates. Central Sulawesi is an area of Sulawesi Island that has a high level of seismicity. This is because Sulawesi Island has many active local faults that can trigger earthquakes. Based on the historical records, there have been several earthquakes that have occurred in Central Sulawesi, and some of that is caused by tsunamis, in 1996 at Toli-toli and Palu and the latest one occurred in Palu, Sigi, Donggala, and Parigi on 28 September 2018. This research aims to group the potential earthquake area in Central Sulawesi using the *K*-Means method by applying the R packages. The results show that we obtained the 2 cluster data. The data characteristic in cluster 1 consists of data that has a low depth so that it is more likely to damage infrastructures but it has no potential for tsunamis. Moreover, the data characteristic in cluster 2 consists of data that has high-depth consequences it does not have any potential to damage the infrastructure, but it could potentially have a tsunami. Finally, the web application for potential grouping earthquake areas in Central Sulawesi using the *K*-Means algorithm is obtained in this research. This is essential to minimize the impacts of the damage due to earthquakes and tsunamis.

Keywords: Earthquake · K-Means · Clustering

1 Introduction

An earthquake is an occurrence that vibrates the earth because of sudden movement or shift of rock layers in the lithosphere. This occurs due to the tectonic plate movement [1]. The territory of Indonesia is located within four plates of the earth's crust, namely the Eurasian Plate (Southeast Asia), Pacific Plate, Australia-India Plate, and Philippine Plate which is often known as the Pacific Ring of Fire. This location makes Indonesia one of the countries that are very vulnerable to potential earthquakes. In addition, the tectonic earthquakes caused by the Pacific Ring of Fire in Indonesia are also often experienced due to many active volcanoes. As a result, the earthquake disaster in Indonesia will be scientifically unavoidable [5].

Central Sulawesi is one of the areas in the central part of Sulawesi Island which has a high level of seismicity. This is because Sulawesi Island has many active local faults

N. I. Inda et al. (Eds.): ISST 2022, APR 7, pp. 312–319, 2023.
https://doi.org/10.2991/978-94-6463-228-6_35

that can trigger earthquakes. One of the most active faults that extend over the north area (Palu City) to the south area (Malili) to Bone Bay is the Palu-Koro Fault in Central Sulawesi [4].

The movement towards the fault which is still very active in the Central Sulawesi region makes this area often rocked by earthquakes, both earthquakes with small magnitudes to earthquakes with large magnitudes. Based on historical records, there have been several earthquakes that have occurred and caused tsunamis. Such as in 1927 which occurred in Palu City, in 1938 which occurred in Parigi, in 1968 which occurred in Tambu, in 1996 which occurred in Toli-toli and Palu City, and most recently occurred in the PASIGALA area (Palu, Sigi, Donggala and Parigi) in 2018 [2].

The series of historical records of earthquakes and many local faults in Central Sulawesi will certainly be a particular concern for the community or migrants that want to visit Central Sulawesi. As a result, it is necessary to have valid data or information about the location to be occupied so that the community or government can take preventive steps.

From the given outlined, it is necessary to develop and create a web application that can facilitate the grouping of areas in Central Sulawesi that are prone to earthquakes and are not prone to earthquakes. In order to do clustering or grouping such as earthquake areas, we can apply the *K*-Means method. This is a method that includes the distance-based clustering algorithm which classifies data into groups or clusters. Moreover, this algorithm can only work on numerical characters. The clusters that have been formed will provide homogeneity of high internal and heterogeneity of high external. The outcome of the cluster as a whole depends on the variables used as the main thing to assess the similarities. The substance of the results will be affected by the addition or subtraction of relevant variables in the cluster analysis. The main reason for the application of the *K*-Means method is that the analysis results of the object size will have a fairly high accuracy. Consequently, the *K*-Means algorithm will relatively be more efficient and measurable to process large numbers of objects [9].

2 Materials

2.1 Earthquake

An earthquake is a violent shock that spreads to the surface of the earth due to disturbances in the lithosphere or skin of the earth. This can occur due to the lithosphere or skin of the earth which has a thickness of 100 km, in which there is an accumulation of energy due to the shift of the lithosphere itself. The lithosphere which has a relative temperature much lower than the mantle layer and core of the earth causes a convective flow, that is, a mass with a high temperature will flow towards an area with a lower temperature. This high-temperature mass is in the asthenosphere layer (the layer in the lithospheric and located above the upper mantle of the earth) which is very viscous in nature that flows slowly. As a result of these movements, the lithosphere fragments into plates that move each other [7].

2.2 Clustering

Clustering is a technique to group a number of data into certain groups of data (clusters) [8]. Clustering is one of the functionalities methods in data mining. This method is a technique used to cluster data based on the similarity between characteristics possessed by the data. Clustering-based has several types, including partitional clustering, hierarchical clustering, and Fuzzy clustering. Partitional clustering is the division of data into a set (cluster) that does not overlap so that each data is in one cluster only [10].

2.3 K-Means Clustering

K-Means Clustering is a technique used to divide data into several groups. As a result, data that have similarities are in the same group, and data that differ in other groups. The clustering purpose is to minimize the set of objective functions in the clustering process. Basically, this is conducted to minimize the variation in clusters and maximize the variation between clusters. The *K*-Means algorithm is the division of clustering that divides data into k sub-regions [3].

The *K*-Means algorithm is convenient to use because of its ability to cluster large data and outlier data faster. The results of the *K*-Means algorithm require that each data fall into a certain cluster. Moreover, the possibility for each data belongs to a certain cluster of a certain stage, where the next stage moves to another cluster. In the application of the *K*-Means algorithm, the data that can be used is numerical data in the form of numbers. Categorical data can still be used but must first be coded to facilitate the calculation of distances or similarity of characteristics possessed by each object [6].

K-Means is a non-hierarchical method that at the beginning will take some of the population components to be used as the initial center of the cluster. The cluster center is randomly chosen from a data set of the population. After that the *K*-Means will be used to test the component in the data and the component will be marked to one of the cluster centers that have been defined. These are all conducted by regarding the minimum distance between components and each cluster. The center position of the cluster will then be calculated repeatedly until the classification of all of the data components into each cluster center. Finally, we form a new cluster center position [6].

3 Methods

In this study, the data mining process model used is *K*-Means Clustering. The methodology of research is conducted in the following steps.

3.1 Study of Literature

The first step is a literature study, which aims to complete basic knowledge and theories derived from books, journals, and Internet media.

3.2 Data Collection

The data that will be used in this research namely data on earthquake depth, magnitude, and the gap in the vulnerable years from 2014 to 2 June 2022.

3.3 Identification of Problem

This stage is carried out after obtaining the appropriate dataset for the clustering process.

3.4 K-Means Algorithm

Determining the *k* Clusters Number. This step can be conducted by regarding the theory and concept. We can propose to determine the number of clusters.

Generating *k* Cluster Center or Initial Centroids Randomly. Determination of the initial centroid is randomly conducted based on the available objects of k clusters number, where the calculation for the next cluster i-centroid uses the following formula.

$$v = \frac{\sum_{i=1}^{n} x_i}{n}; i = 1, 2, 3, \ldots, n \quad (1)$$

v: the centroid on cluster
x_i: the i-th of x object
n: the objects number of the cluster number

Allocating Each Data to the Nearest Centroid/Average. The two objects' proximity is obtained by considering the distance between two objects. Moreover, the data proximity particular cluster is obtained by considering the distance between the data and the center of the clusters. The calculation of the data distance to each cluster center is necessary to conduct. The closest distance will then determine which data belongs to which cluster. The calculation for the distance is used the Euclidean Distance Space formula as given as follows.

$$D_{L2}(x_2, x_1) = ||x_2 - x_1||^2 = \sqrt{\sum_{j-1}^{P} \left(x_{2j} - x_{1j}\right)^2} \quad (2)$$

x_1: Data Object
x_2: Centroid
p: Data Dimension
D: Distance
L_2: Euclidean Distance Space

Redefine the Center of the New Cluster by Calculating the Average of the Membership of the Current Cluster. Allocate objects to the nearest cluster using the new centroid. If there are members who move clusters, repeat step 4. If no members move clusters, the process is complete.

3.5 Web Application

UI (User Interface)/UX (User Experience) Design. At this stage, the web application design is carried out. The thing to do is to design the UI/UX of the application that is

Table 1. The average of each cluster

Average of Each Cluster					
Cluster	Depth	Magnitude	Gap	Dmin	rms
1	28.0305	4.563125	84.25917	1.736844	0.8644792
2	193.3057	4.349020	103.23529	1.415245	0.769059

tailored to the user's needs. Several things should make in this stage, including doing interface design, mapping the distribution of earthquake points, graphs and data filters.

Server Design. This stage is advanced in the preparation of the application. At this stage, the prediction model that has been performed using *K*-Means clustering is used to predict the potential for earthquakes at the user's location.

4 Results and Discussion

After analyzing the *K*-Means Clustering Algorithm, 2 groups of data were obtained. The obtained data characteristic is described in Table 1.

Table 1 shows that cluster 1 consists of data that has a low depth, consequently, it does not have tsunami potential. Meanwhile, cluster 2 consists of data that has a high depth. As a consequence, it can potentially cause tsunamis.

The results of the Web Application of Classification Potential Earthquake Region in Central Sulawesi project consist of 2 main tabs. The first one is a tab that shows the earthquake distribution points that have been clustered, while the second tab provides a visualization of the earthquake distribution points based on the earthquake magnitude.

Figure 1 shows the clustering tab which describes the visualization of earthquake point mapping in Central Sulawesi that is divided into two clusters. The clustering model used the k-means algorithm. The mapping provides an interactive visual form in exploring earthquake points to provide information related to earthquake-prone areas. Otherwise, the dashboard has a clustering simulation process by changing the variables to see the condition of the earthquake point if viewed under specific conditions. Of course, this is very interesting to study more deeply.

Figure 2 shows the Earthquakes tab. We can see the visualization of earthquake points based on their magnitude. This tab is equipped with the date range filter. This makes it easier to see the distribution of earthquakes that occurred in a specific date range. Moreover, Fig. 3 shows the tab provided earthquake timeline features, histograms, and frequency tables related to date range filters. This feature can make easier to see the distribution of the occurred earthquake.

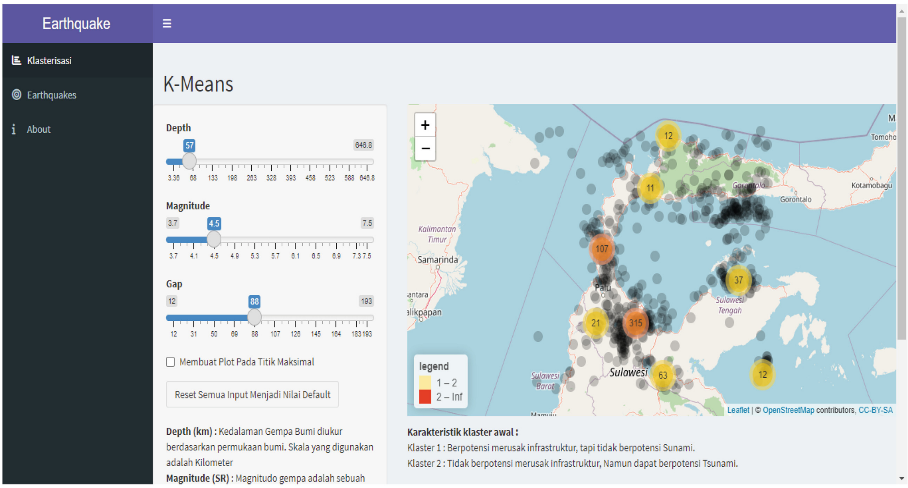

Fig. 1. Tab Clustering (You can access the web dashboard at the following link: https://bulava.shinyapps.io/CentralSulawesiEarthquake/)

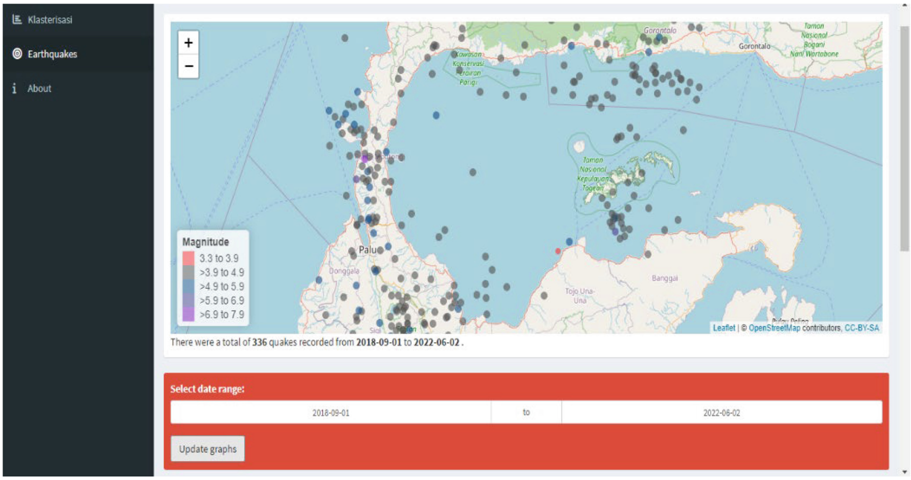

Fig. 2. Tab Earthquakes by visualization of the earthquake distribution points

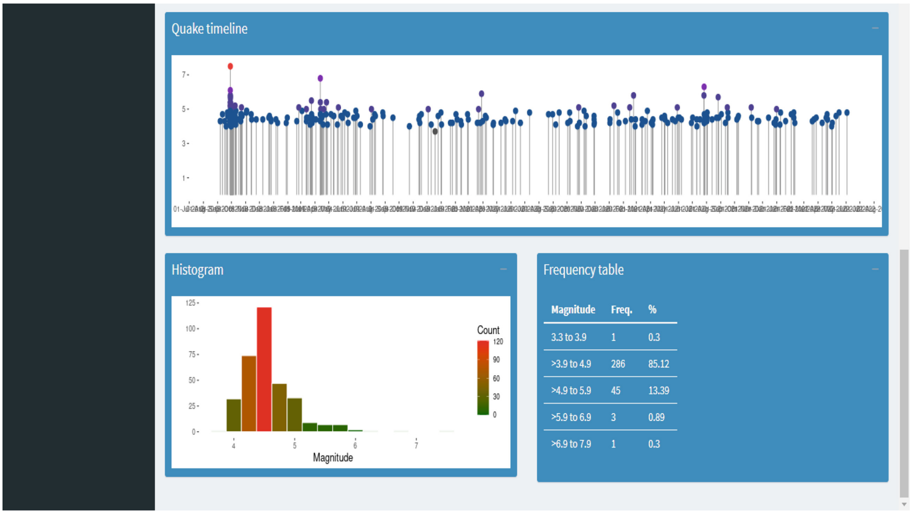

Fig. 3. Tab Earthquakes by visualization of the earthquake frequency by the magnitude

5 Conclusion

The Central Sulawesi Earthquake Potential Area Classification Web application has been created according to the stated objectives. The K-means algorithm is used to group data on earthquake points in Central Sulawesi into two clusters. The first cluster is located in an area that has the potential to damage infrastructure but does not have the potential for a tsunami. Meanwhile, the second cluster is located in an area that does not have the potential for infrastructure damage but can potentially cause tsunamis. This application can definitely help people to see the potential for earthquakes in several areas based on predetermined clusters and can help the community and the government to take preventive steps in minimizing casualties and losses due to infrastructure damage in realizing the optimal sustainable development goals.

Acknowledgements. We would like to thank the Mathematics and Natural Sciences Faculty, Tadulako University for the research funding.

References

1. Atmojo, S., Muhandis, I.: Sistem Informasi Geografis Bencana Gempa Bumi dengan Pendekatan PGA untuk Mitigasi Bencana. Jurnal Ilmiah Edutic 6(1), 10–14 (2019).
2. Daryono: Tataan Tektonik dan Sejarah Kegempaan Palu, Sulawesi Tengah. In Badan Meteorologi Klimatologi dan Geofisika, Palu (2011).
3. Ediyanto, Mara, M. N., Satyahadewi, N.: Pengklasifikasian Karakteristik dengan Metode K-Means Cluster Analysis. Buletin Ilmiah Matematika Statistika dan Terapannya (Bimaster), 2(2), 133–136 (2013).

4. Gautama, A. A., Ngr Wisnu, Purwanto, Y., Purboyo, T. W.: Analisis Pengaruh Penggunaan Manhattan Distance pada Algoritma Clustering Isodata (Self- Organizing Data Analysis Technique) untuk Sistem Deteksi Anomali Trafik, E-Proceeding Eng., 2(3), 7404–7411 (2015).
5. Kaharuddin, Hutagalung, R., Nurhamdan: Perkembangan Tektonik dan Implikasinya Terhadap Potensi Gempa dan Tsunami di Kawasan Pulau Sulawesi. In: Proceedings JCM, Makassar (2011).
6. Marisa, Sembiring, U. A., Margaretha, H.: Earthquake Probability Prediction in Sumatra Island Using Poisson Hidden Markov Model (HMM). In: AIP Conference Proceedings (2019).
7. Metisen, Benri M., Sari, H. L.: Analisis Clustering Menggunakan Metode K-Means dalam Pengelompokan Penjualan Produk pada Swalayan Fadhila, Jurnal Media Infotama, 11(2) (2015).
8. Mustafa, B.: Analisis Gempa Nias dan Gempa Sumatera Barat dan Kesamaannya yang Tidak Menimbulkan Tsunami. Jurnal Ilmu Fisika Universitas Andalas, 2(1), 44–50 (2010).
9. Sibuea, Mustika L., Safta, A.: Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering, Jurteksi (Jurnal Teknologi dan Sistem Informasi), 4(1), 85–92 (2017).
10. Witten, Ian H., Frank, E.: Data Mining Practical Machine Learning Tools and Techniques 2nd Edition, Morgan Kanfmann, San Fransisco (2005).