Journal of Theoretical and Applied Information Technology
31st January 2021. Vol.99. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CLUSTER ANALYSIS AND SEISMICITY PARTIONING FOR NORTHERN SUMATERA USING MACHINE LEARNING APPROACH

[1*]EVA DARNILA, [2]KERISTA TARIGAN, [3]SUNARDI, [4]FATI GRANTIANUS NAFIRI LAROSA, [5]MARZUKI SINAMBELA

[1]Department of Informatics, Universitas Malikussaleh, Aceh, Indonesia
[2,3,6]Department of Physiscs, Universitas Sumatera Utara, Medan, Indonesia
[4]Department of Informatics, Universitas Methodist Indonesia, Medan, Indonesia
[3]Medan Aviation Polytechnic, Medan, Indonesia
[5]Badan Meteorologi Klimatologi dan Geofisika, Indonesia

Corresponding Author: [1]eva.darnila@unimal.ac.id
Author: [2]kerista@usu.ac.id, [3]machdisunardi@gmail.com, [4]fatignlarosa@gmail.com, [5]sinambela.m@gmail.com

ABSTRACT

Tectonic activity in the past one year in North Sumatra is quite active as it's located in seismic region on the fault zone accommodates most of the strike-slip movements associated with the sloping convergence between the Indo-Australian and Eurasian plates. In this case, we used clustering approach to see the potential for earthquakes originating from fault activities in Northern Sumatra recorded by seismic network sensors that have been installed either Broadband type or the new mini region. Northern Sumatra based on the source of earthquake activity can be divided into several segments. The main goal of this study to identify the distribution of the Northern Sumatra inland earthquake based on segment activity using the clustering approach. The result shown that the dominance of onshore earthquakes in North Sumatra varies greatly based on magnitude. The magnitude frequency distribution for the period 2019 and January to June 2020 was dominated by earthquakes with magnitudes below 4. The clustering approach in this study illustrates the classification of magnitude of terrestrial earthquakes in North Sumatra in the categories of minor, light and moderate. In 2019 and January-June 2020 period the dominance of the most sources of earthquake was due to the Aceh Central segment.

Keywords: *Cluster, Machine Learning Approach, Earthquakes*

1. INTRODUCTION

The Northern Sumatera is located in seismic region on the ring of fire in the word. In addition to the subduction zone of the island's west coast, Sumatra also has a large slip fault, as known as the Semangko Fault, which stretches across the island [1], [2]. The fault zone accommodates most of the strike-slip movements associated with the sloping convergence between the Indo-Australian and Eurasian plates. Based on the earthquake events that often occur in Sumatra in general [3] cluster based on seismic maps along the fault. In this case the Sumatran Fault Segment, can be described based on the level of the cluster where the maximum magnitude of the earthquake is obtained from the segment and regression.

Tectonic activity in the past one year in North Sumatra is quite active. This can be seen from the results of seismicity in 2019. The dominance of earthquake activity in Northern Sumatra generally occurs on land, as in figure 1. In this case, identification is needed in distributing earthquake events that occur in Northern Sumatra.

The identification of land earthquakes based on Figure 1 shows that variations in magnitude can be classified into 2 colors, namely small and medium magnitudes. Land earthquake events in the 2019 period show that the distribution of earthquake events on land is strongly dominated by Sumatra Fault Zone activities and local segment.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The North Sumatra Seismicity Map in figure 1, 2019, illustrates the source of the earthquake occurrence caused by several local segment activities. The distribution of the earthquake epicentre in North Sumatra is very interesting to study, seeing the frequency of earthquake events increasing, and potentially in the earthquake return period. Earthquake return periods can be described as recurrence intervals that indicate the average time or estimated time the earthquake occurred. In this case, the measurement of earthquake return period is closely related to statistical measurements which are based on historical data over a long period, and are used for risk analysis. The analysis is intended to assume that the probability of an earthquake event occurring does not vary from time to time and does not depend on past events.

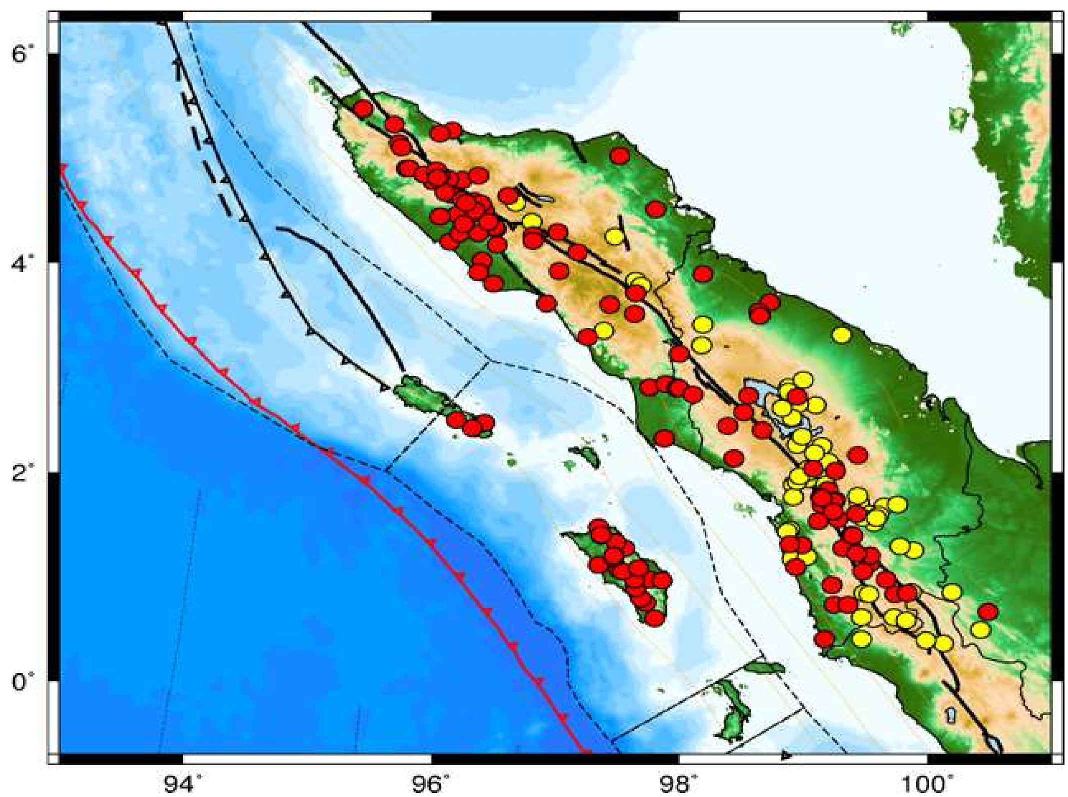


Figure 1. Seismicity of the Northern Sumatra region May Period 2019 - May 2020

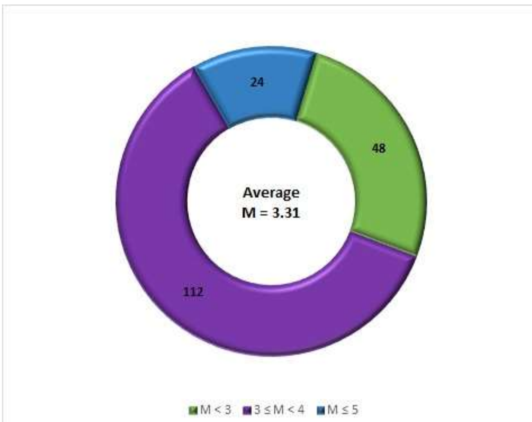
In this study, we focus for clustering of earthquakes occurring on the northern Sumatra. Clustering approach is a main problem to see the potential for earthquakes originating from fault activities in Northern Sumatra recorded by seismic network sensors that have been installed either Broadband type or mini region in Northern Sumatera based on the source of earthquake activity can be divided into several segments. As we know, the new mini region had been installed in Northern Sumatra [4] and had been working to detect the earthquake activity. Based on [2], [5], [6] explained the source of the earthquake segment in North Sumatra was strongly influenced by the Sumatra Fault Zone and several local segments.

In general, the results of this study will be illustrated how the frequency of earthquakes in the Northern Sumatra can be used as a reference in further analysis of seismic sources that can be monitored by network sensors in North Sumatra based on the local fault segment that is the source. This study aims to identify the distribution of the Northern Sumatra inland earthquake based on segment activity using the clustering approach.

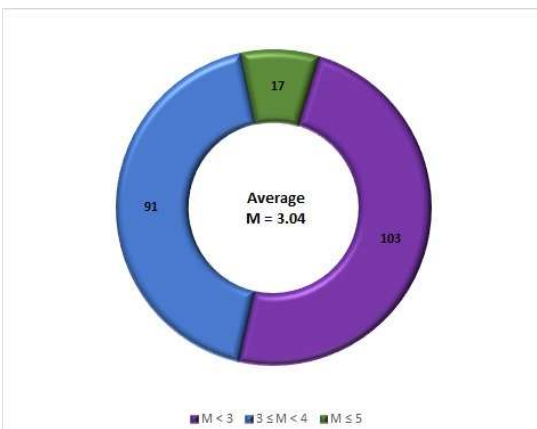
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2. DATA AND METHOD

In this study, we utilized earthquake data from the earthquake catalog of the Deli Serdang Geophysics Station. The recorded earthquake data was observed from 2019 and January-June 2020, with the category of earthquakes on the northern Sumatra land. The selection of earthquakes on land in the period 2019 and January 2020 to June 2020 provides a solid basis for evaluating the performance of sensors that have been installed and reviewed in previous reports. A list of earthquake events on land can be seen in Figure 2.



a



b

Figure 2. Land earthquake events in Northern Sumatra: a. 2019 period, b. period January-June 2020

The number of earthquake events in the 2019 period in Figure 2(a) was 184 with an average magnitude of earthquakes 3.31, and the number of earthquake events in January-June 2020 in figure 2(b) amounted to 211 events with an average earthquake magnitude of 3.04. This condition shows the frequency of the number of earthquake events from 2019 to January-June 2020 has increased.

Earthquake strength ranges from being so weak that it cannot be perceived as violence which can cause serious damage to the entire city. Seismicity or seismic activity of an area refers to the frequency, type and size of earthquake experienced during a certain period of time. Earthquake classification based on magnitude can be seen in Table 1.

Table 1: Earthquake classification based on magnitude [7]

Earthquake Qualification	Magnitude
Minor	< 4.0
Light	4.0 - 5.0
moderate	5.0 < M < 5.9
Strong	6.0 < M < 6.9
Major	7.0 < M < 7.9
Great	>8.0

Earthquake parameters in North Sumatra are very dependent on seismic network sensors installed in North Sumatra in both Broadband type and Mini Regional type. In this case, the earthquake occurrence recorded in the earthquake catalog at Deli Serdang Geophysical Station is based on the seismic network in Figure 3.

Figure 3. is the distribution of mini regional sensors and broadband at Geophysics Station of Deli Serdang. There are 26 types of Broadband and 16 Mini regions which are already operating and support earthquake parameter analysis both on land and at sea. The distribution of these sensors will support BMKG's national operations for earthquake and tsunami early warning.

In this study, we used K-means clustering as the simplest and most popular "unsupervised machine learning algorithms". The purpose of this algorithm is to find groups in the data, with the number of groups represented by the K variable. K variable itself is the number of clusters we want [8], [9]. K-means clustering is a type of unsupervised learning,

Journal of Theoretical and Applied Information Technology
31st January 2021. Vol.99. No 2

ISSN: **1992-8645** www.jatit.org E-ISSN: **1817-3195**

which means that it is used when we have unlabeled data [10], [11].

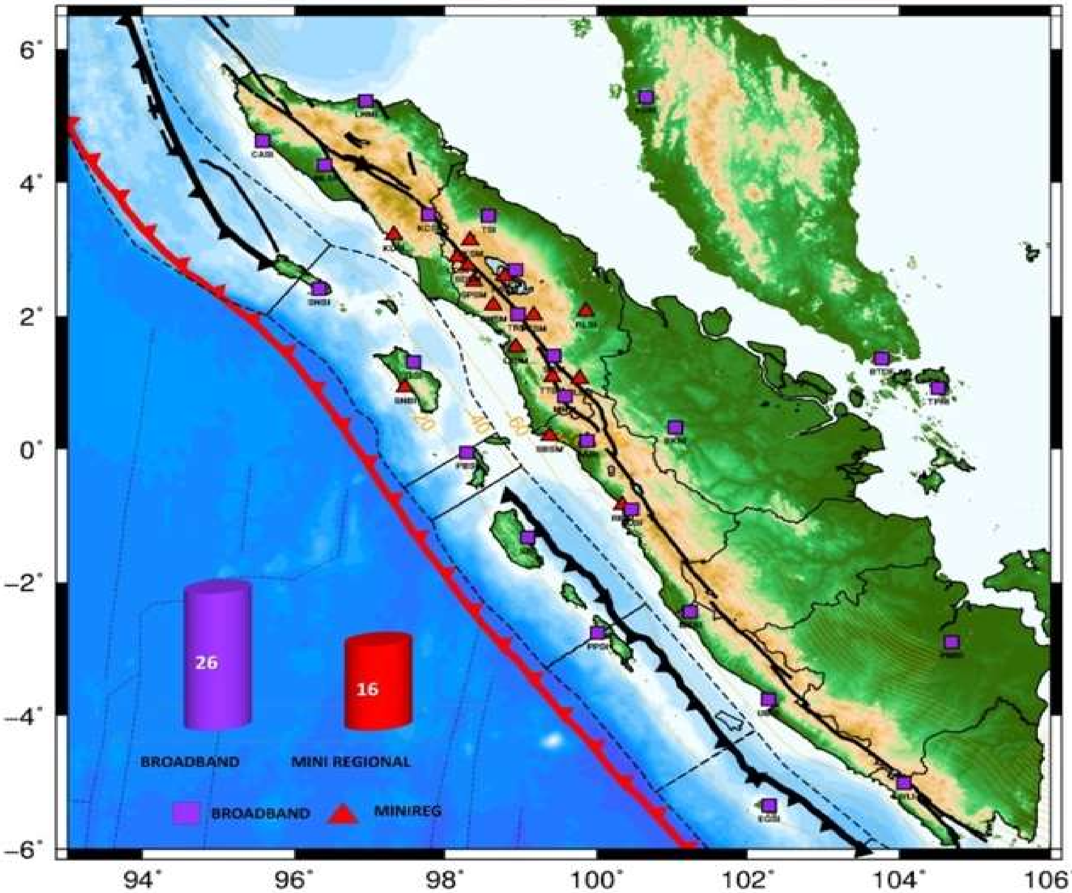


Figure 3. Map of Mini Regional and Broadband Sensor Distribution in Deli Serdang Geophysical Seismic Network

The algorithm works iteratively to find clusters in the data, and the number of groups is represented by the variable K. We use the euclidian distance to find the distance between the data center point and the center of mass, and with this distance we can sort the data points into clusters [12]–[14]. Euclidian distances can be calculated as:

Between two points

If p = (p1, p2) and q = (q1, q2) then the distance is given by

$$d(p,q) = \sqrt{(q1-p1)^2 + (q2-p2)^2} \text{ or } \sqrt{\textstyle\sum_{n}^{i=1}(x_i-y_i)^2} \quad (1)$$

For the three dimensions the formula is,

$$d(p,q) = \sqrt{(q1-p1)^2 + (q2-p2)^2 + (q3-p3)^2} \quad (2)$$

In this case, the steps for grouping K-means as a machine learning approach parth model are shown in Figure 4 by determining the number of clusters (K) for the dataset and selecting k random points in the dataset. This random point k is called a centroid and must be equal to the number of clusters. Calculate the Euclidean distance between each data point and the selected centroid cluster. A point is considered to be in a certain group when the Euclidean distance is minimum. Determine the new centroid for each cluster by taking the average of all points assigned to the cluster. Repeat steps 2 and 3 until the centroid position no longer changes and the cluster assignment remains the same.

The process of this study can be seen in detail in Figure 5, which is divided into 3 stages, namely the pre-processing of land earthquake data based on the Northern Sumatra Earthquake sourced from the Deli Serdang Geophysical Station, the second stage by processing and computing time-series earthquake data overland period 2019 and January-June 2020 with a cluster approach. And the last stage is by optimizing the K value and output plotting from the computation results of the Northern Sumatra earthquake earthquakes cluster.

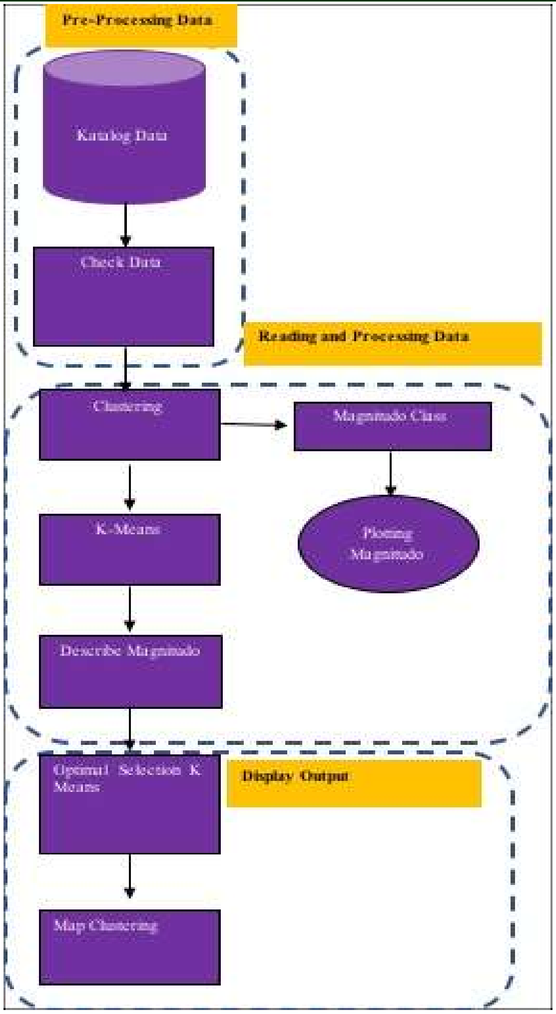
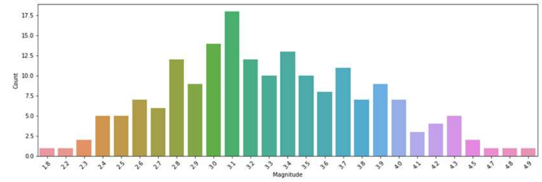


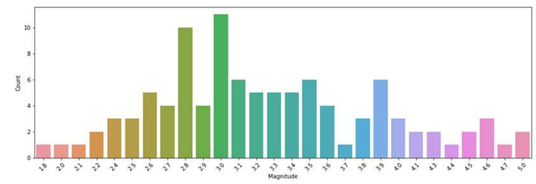
Figure 5. The flowchart of Northern Sumatra Earthquake Clustering

3. RESULTS AND DISCUSSION

The number of land earthquakes that occurred in 2019 was as many as 184 events with different magnitude variations. Based on analytical data the number of earthquakes based on magnitude can be seen in Figure 6a. The number of land earthquakes that occurred in 2020, January to June was 211 events with different magnitudes. Based on analytical data the number of earthquakes based on magnitude can be seen in the figure 6b.



a



b

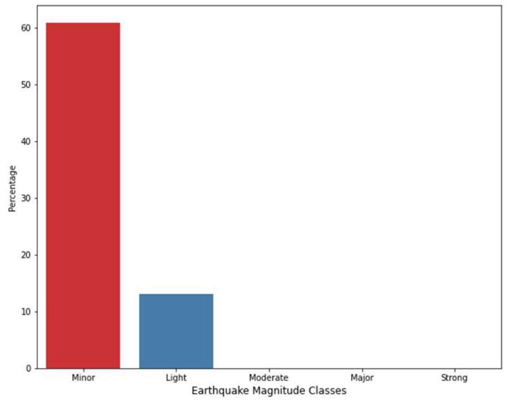
Figure 6 Number of earthquakes based on magnitude: a. 2019, b. January-June 2020

Figure 6 a is an earthquake that occurred on land in the period 2019. Results of computation of time-series on land earthquakes in North Sumatra illustrates the smallest magnitude at 1.8, most 3.1 and largest 4.9. Figure 6b. is an earthquake that occurred on land in the January-June 2020 period. The computational time-series results of land earthquakes in Northern Sumatra illustrate the smallest magnitude at 1.8, most 3.0 and largest 5.0.

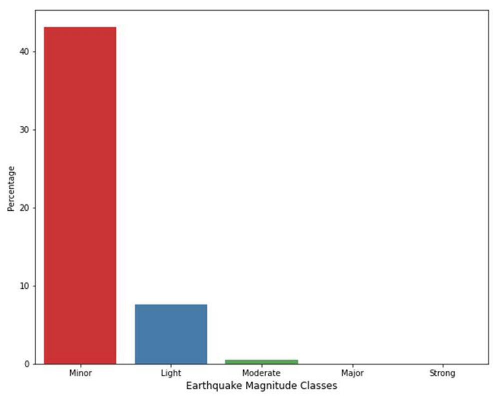
In this study the magnitude classification recorded in the Deli Serdang geophysical data catalog is categorized into 5 classes, namely in the minor, light, moderate and strong categories. Percentage of magnitude of inland earthquakes in North Sumatra from 2019 can be seen in Figure 7a. Based on analytical data in Figure 7a, the earthquake in the 2019 period is divided into 2 categories, namely magnitude with the categories of minor and light. In the figure 7b the magnitude classification recorded in the Deli Serdang geophysical data catalog is categorized into 5 classes, namely in the minor, light, moderate and strong categories. Percentage of magnitude of inland earthquakes in North Sumatra from January to June 2020.

Journal of Theoretical and Applied Information Technology
31st January 2021. Vol.99. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



a

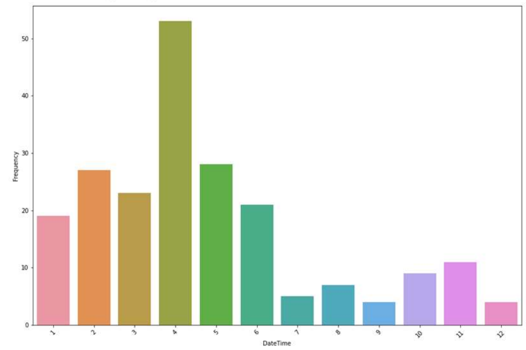


b

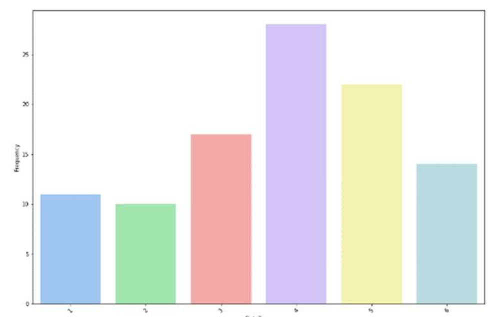
Figure 7. Percentage of groundwater grade by magnitude classification: a. 2019, b. January-June 2020

Figure 8 (a) is the frequency distribution of terrestrial earthquakes in North Sumatra in 2019. The plot results in Figure 8 illustrate the highest frequency of terrestrial earthquakes in April 2019 and in September 2019 the frequency of events was minimal. Figure 8b. is the frequency distribution of land earthquakes in North Sumatra in the periods January-June, 2020. The plot results in Figure 14 illustrate the most frequency of land earthquakes in April 2020.

In this case, an analysis study of earthquake clusters in Northern Sumatra for the 2019 period was conducted. The magnitude cluster of the North Sumatra Earthquake in the 2019 period was conducted to classify the Earthquake on the segment in North Sumatra. Land earthquakes both on a small and large scale will have an impact, the earthquake magnitude cluster is based on damage estimates. Figure 9 is an export map that illustrates damage estimation based on the magnitude of the 2019 period. The results of the plotting obtained 2 categories, namely limited damage (blue dot) and minor damage (green dot).



a



b

Figure 8 Distribution of the frequency of the North Sumatra earthquake based on magnitudo per month: a. 2019, b. 2020

The cluster analysis of earthquake events in Northern Sumatra in January to June 2020 was conducted. Clustering is done to classify land earthquakes against segments in North Sumatra. Land earthquakes both on a small and large scale will have an impact. This study also conducted an earthquake magnitude cluster based on estimated damage potential. Figure 15is an export map depicting damage estimates based on the magnitude of the strength of the January-June 2020 period. From the ploting results, 3 categories are obtained, namely limited damage (blue dot), minor damage (green dot) and Slight Damage.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

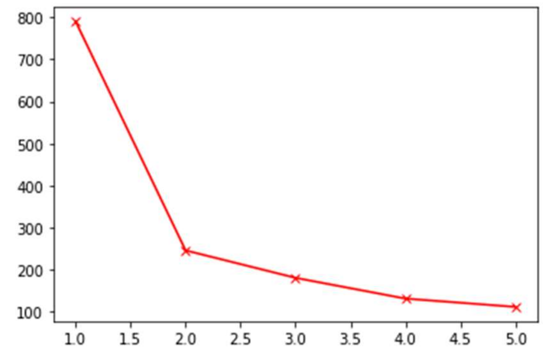
a

b

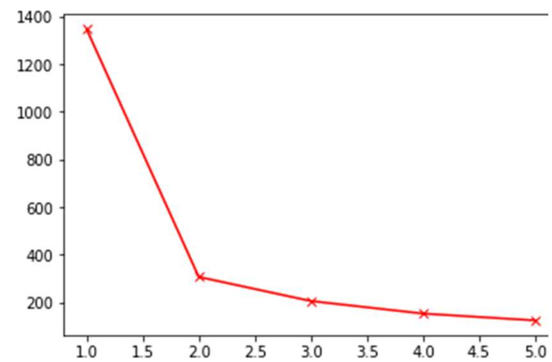
Figure 9. Estimated damage potential based on the magnitude level: a. 2019, b. 2020

With the k-mean cluster approach, the characteristics of the North Sumatra earthquake in the 2019 period are divided into 4 classes. From the results of k-means clustering the k value can be seen in Figure 10a.

With the k-mean cluster approach, the characteristics of the North Sumatra earthquake in the January-June 2020 period are divided into 4 classes. From the results of k-means clustering the k value can be seen in Figure 10b. In the dataset without labels find the optimal value of K. In this case, the number of clusters needed in the classification of the Northern Sumatra earthquake is uncertain. Therefore, a heuristic approach with the Elbow method is used to identify the optimal value for K. In this study identifying values for K uses multiple learning. We plot the sum of squared distances on running k-means for various K values.



a



b

Figure 10. K-value of the Northern Sumatra earthquake: a. 2019, b. January to June 2020

In the dataset without labels find the optimal value K. In this case, the number of clusters needed is uncertain. Therefore, a heuristic approach with the Elbow method is used to identify the optimal value for K. In this study identifying values for K uses multiple learning. We plot the sum of squared distances on running k-means for various K values. there are limitations to using the elbow method because it is a heuristic approach.

Using the optimal number (K) as with the elbow method, earthquakes can be grouped from the dataset into four different groups. The number or earthquake with a magnitude of 1.5 or higher is much higher in the red group. A total of 92 earthquakes formed the red cluster, and 33 earthquakes formed the blue cluster. The North Sumatra region closest to the local segment is most affected by the earthquake in this 2019 period. The 2019 earthquake and aftershocks explain why we have such a cluster in the Northern Sumatra region. A total of 194 earthquakes with a magnitude of 3 or more occurred in 2019 and a total of 42 earthquakes with a magnitude of 4. In Figure 11 b, the number or earthquake with a magnitude of 1.5 or higher is much

Journal of Theoretical and Applied Information Technology
31st January 2021. Vol.99. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

higher in the red group. A total of 92 earthquakes formed the red cluster, and 33 earthquakes formed the blue cluster. The North Sumatra region closest to the local segment was most affected by the earthquake in the January-June 2020 period.

a

b

Figure 11. Northern Sumatera Earthquake Cluster based on K-Means: a. 2019, b. January to June 2020

The January-June 2020 earthquake and aftershocks explain why we have such a cluster in the Northern Sumatra region. A total of 194 earthquakes with magnitude below 4 January-June 2020 period.

The map in Figure 12a also shows a yellow centroid epicenter for two groups. Clusters are generated by calculating the euclidian distance between centroids and data points. One interesting observation on the map is that the earthquake epicenter in the North Sumatra region caused by both small and large earthquakes explains why the epicenter is mostly in the North Sumatra fault area and local segment. Figure 12 b. is an earthquake cluster in the January-June 2020 period which shows a blue centroid epicenter for two groups. Clusters are generated by calculating the euclidian distance between centroids and data points. One interesting observation on the map is that the earthquake epicenter in the Northern Sumatra region caused by both small and large earthquakes explains why the epicenter is mostly in the North Sumatra fault area and the local segment and there is an increase in the activity of the North Sumatra earthquake. from the 2019 period to the January-June 2020 period.

a

b

Figure 12. Northern Sumatra Earthquake Cluster: a. 2019, b. 2020

Segment cluster analysis in this study is divided into two categories, the first is the 2019 period and the January-June 2020 period. In the 2019 period, land earthquakes can be clustered in Figure 13.

Based on Figure 14, the Batee A segment is very dominant in the occurrence of land earthquakes that occurred in the 2019 period in the Aceh region, followed by the activities of the Aceh central segment and the Angkola segment. Whereas in the area of North Sumatra the source of earthquake activity was dominated by the Renun A segment. A complete description of the occurrence of land earthquakes in the 2019 period can be seen in Figure 14. the total number of earthquake events caused by segment sources is detailed in Figure 14.

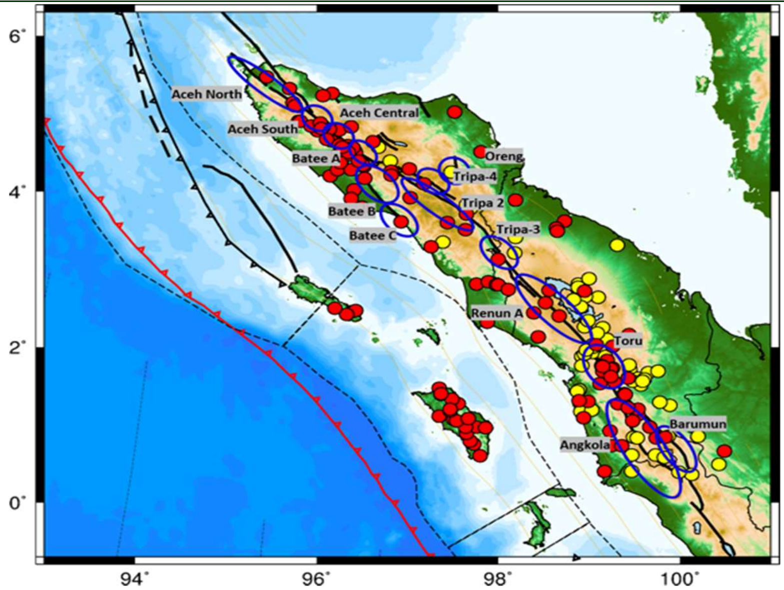


Figure 13. Inland earthquake clusters in Northern Sumatra based on the 2019 segment.

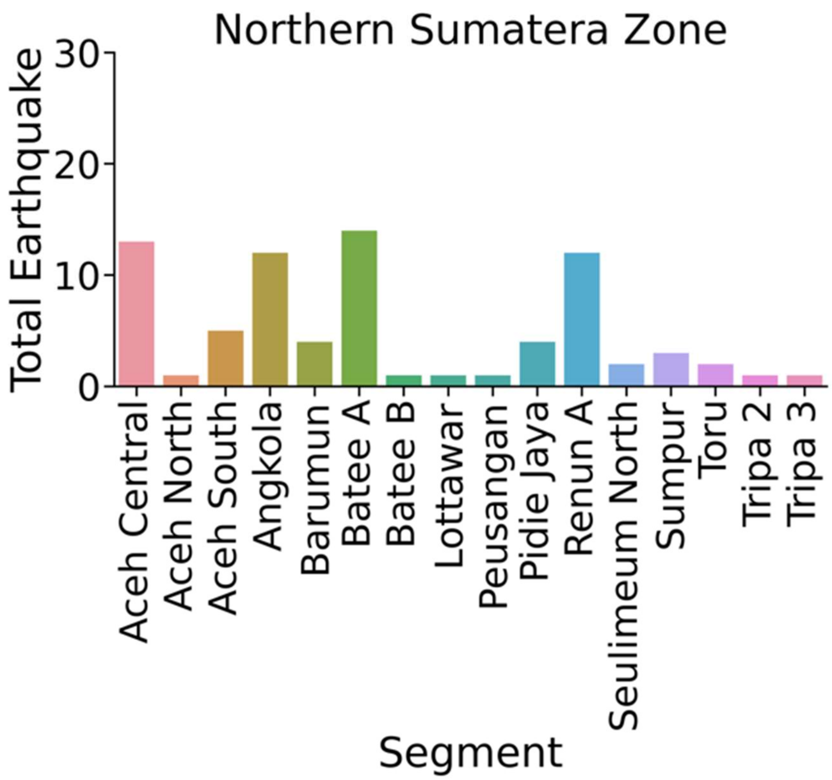


Figure 14. Total inland earthquakes in North Sumatra based on 2019 segment clusters

Figure 15 is a map of the earthquake clusters for the January-June 2020 period. The dominance of the source of earthquake events caused by the Aceh Central segment is more dominant than other segments.

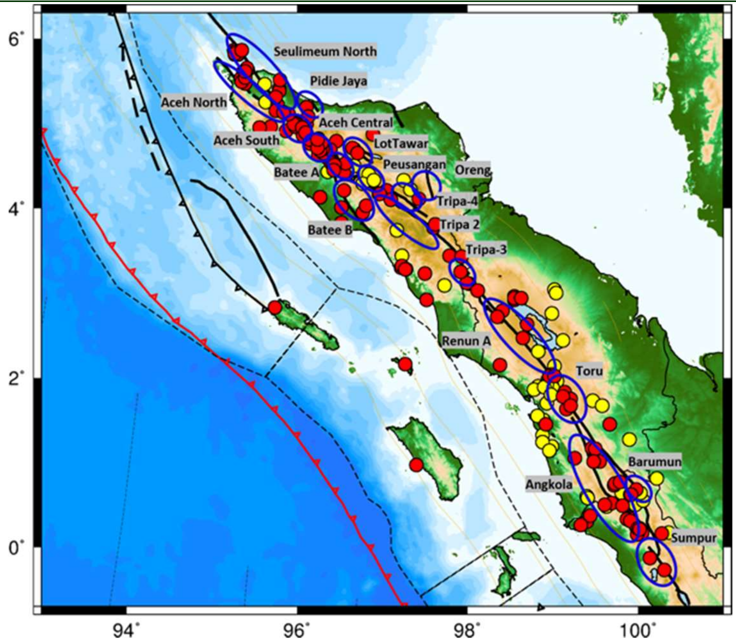


Figure 15. Earthquake inland clusters in Northern Sumatra by segment from January to June 2020

This condition illustrates the activities of the Aceh segment from 2019 to the period January-June 2020 which experienced a fairly high increase and dominated the occurrence of land earthquakes in Northern Sumatra.

In Batee A segment, the activity of apex in this fault has decreased from 2019 to January-June 2020. The Seimelue North segment which in the previous period in 2019 had minimal seismic activity, in the January-June 2020 period increased. The high activity of the earthquake in the January-June 2020 period due to several segments made special attention in this study, especially when viewed from the role of seismic sensors installed in North Sumatra. Whether or not the results of the analysis will greatly depend on the performance and density of existing sensors.

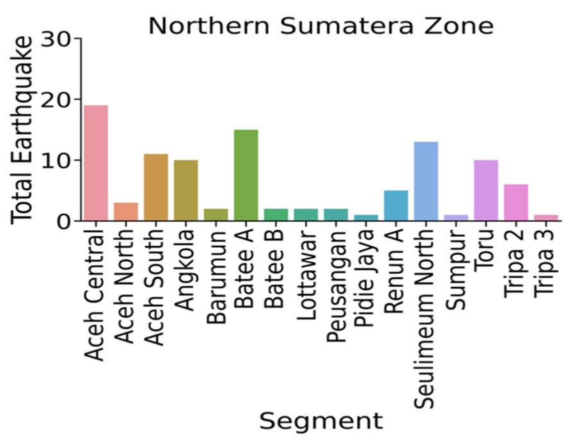


Figure 16. Total inland earthquakes in North Sumatra based on cluster segments in the January-June 2020 period.

In detail, the total earthquake in the January-June 2020 period can be seen in Figure 15 an increase in earthquake activity due to faults or local segments prompts special attention for earthquake analysis on land and the potential for future return periods.

4. CONCLUSION

Based on the results of the classtering of the North Sumatra onshore earthquake, the dominance of onshore earthquakes in North Sumatra varies greatly based on magnitude frequency distribution for the period 2019 and January to June 2020 was dominated by earthquakes with magnitudes below 4.

In 2019 and January-June 2020 period the dominance of the most sources of earthquake was due to the Aceh Central segment. The distribution of the frequency of the earthquake in the northern part of Sumatra, which increased from 2019 to January-June 2020, was strongly influenced by the density of new sensors of mini region already installed in the North Sumatra region. The results of the sensor performance study in the previous study had a significant impact on the recording of special earthquake in the North Sumatra region. Future we can identify the earthquake activity in local segment area in North Sumatra.

ACKNOWLEDGEMENTS

This research is supported by Universitas Malikussaleh and Agency of Indonesia Meteorology, Climatology and Geophysics (BMKG).

REFERENCES:

[1] K. Sieh and D. Natawidjaja, "Neotectonics of the Sumatran fault, Indonesia," *J. Geophys. Res. Solid Earth*, vol. 105, no. B12, pp. 28295–28326, 2000, doi: 10.1029/2000jb900120.

[2] T. Rahayu, "Wajah Tektonik Sumatera Bagian Utara," *https://books.google.co.id/books/about?id=yyPsDwAAQBAJ&redir_esc=y*, 2020. .

[3] P. W. Burton and T. R. Hall, "Segmentation of the Sumatran fault," *Geophys. Res. Lett.*, vol. 41, no. 12, pp. 4149–4158, 2014, doi: 10.1002/2014GL060242.

[4] M. Sinambela, M. Situmorang, K. Tarigan, S. Humaidi, and T. Rahayu, "Design of solar power system for the new mini region of broadband seismometer shelter in Tiganderket, Karo, North Sumatera, Indonesia," *Case Stud. Therm. Eng.*, vol. 22, p. 100747, Dec. 2020, doi: 10.1016/j.csite.2020.100747.

[5] D. H. NATAWIDJAJA and W. TRIYOSO, "the Sumatran Fault Zone — From Source To Hazard," *J. Earthq. Tsunami*, vol. 01, no. 01, pp. 21–47, 2007, doi: 10.1142/s1793431107000031.

[6] C. Subarya *et al.*, "Plate-boundary deformation associated with the great Sumatra-Andaman earthquake," *Nature*, vol. 440, no. 7080, pp. 46–51, 2006, doi: 10.1038/nature04522.

[7] USGS, "Earthquakes Magnitude Scale and Classes," *http://www.geo.mtu.edu/UPSeis/magnitude.html*, 2020. .

[8] Z. Fan and X. Xu, "Application and visualization of typical clustering algorithms in seismic data analysis," in *Procedia Computer Science*, Jan. 2019, vol. 151, pp. 171–178, doi: 10.1016/j.procs.2019.04.026.

[9] R. Joaquín Pérez-Ortega, Nelva Nely Almanza-Ortega, Andrea Vega-Villalobos, "The K-Means Algorithm Evolution | IntechOpen," *https://www.intechopen.com/books/introduction-to-data-science-and-machine-learning/the-em-k-em-means-algorithm-evolution*, 2018. .

[10] S. F. Hussain and M. Haris, "A k-means based co-clustering (kCC) algorithm for sparse, high dimensional data," *Expert Syst. Appl.*, vol. 118, pp. 20–34, Mar. 2019, doi: 10.1016/j.eswa.2018.09.006.

[11] E. Bedalli, E. Mançellari, and O. Asilkan, "A Heterogeneous Cluster Ensemble Model for Improving the Stability of Fuzzy Cluster Analysis," *Procedia Comput. Sci.*, vol. 102, no. August, pp. 129–136, 2016, doi: 10.1016/j.procs.2016.09.379.

[12] P. Novianti, D. Setyorini, and U. Rafflesia, "K-means cluster analysis in earthquake epicenter clustering," *Int. J. Adv. Intell. Informatics*, vol. 3, no. 2, pp. 81–89, 2017, doi: 10.26555/ijain.v3i2.100.

[13] E. Irwansyah and E. Winarko, "Spatial data clustering and zonation of earthquake building damage hazard area," *EPJ Web Conf.*, vol. 68, pp. 0–5, 2014, doi: 10.1051/epjconf/20146800005.

[14] A. Ahmad and N. Irsalinda, "Cluster Analysis of Earthquake ' s Data Clustering in Indonesia using Fuzzy K-means Clustering," vol. 3, no. April, pp. 3–7, 2020.