cognizancejournal.com
Risna Simatupang *et al*, Cognizance Journal of Multidisciplinary Studies, Vol.3, Issue.5, May 2023, pg. 70-76
(An Open Accessible, Multidisciplinary, Fully Refereed and Peer Reviewed Journal)
ISSN: 0976-7797
Impact Factor: 4.843
Index Copernicus Value (ICV) = 76.35

Seismicity Parameter Study of Nias Island Area Using K-means Clustering Algorithm and Least Square Method

Risna Simatupang[1]; Gatot Yuliyanto[2]; Udi Harmoko[3]
[1-3]Physics Department, Diponegoro University, JL. Prof. Soedarto SH. Semarang, Indonesia, Postal Code: 50271
[2] gatoty@fisika.fsm.undip.ac.id

DOI: 10.47760/cognizance.2023.v03i05.007

Abstract—* *One of the western parts of Sumatra Island that frequently experiences earthquakes is Nias Island. On Nias Island, there have been several earthquakes with large magnitudes recorded by IRIS. The earthquake destroyed facilities and infrastructure and resulted in thousands of deaths. The study of the seismicity level of the Nias Island area is one of the most feasible mitigation strategies. This study aims to classify the earthquakes that occurred into several clusters as well as determining earthquake-prone areas that have the potential to damage. Using earthquake data from 1970 to August 2022 obtained from IRIS, data processing begins by determining the optimal number of clusters using Elbow Method, seismic grouping using K-means Clustering Algorithm, and analysis of earthquake-prone areas based on seismicity parameter calculations a value and b value to each cluster use Least Square Method. Based on Elbow Method and K-means Clustering Algorithm, in this study the Nias area was devided into two clusters. Calculation of the seismicity parameter values for each cluster gives result a value cluster 0 is equal to 3.72 and b value cluster 0 is equal to 0.37 meanwhile a value cluster 1 is equal to 3.60 and b value cluster 1 is equal to 0.44. Based on these data, the areas prone to earthquakes that have the potential to damage are areas that are located on cluster 0.

Keywords—* *Cluster, Seismicity, K-means Clustering Algorithm, Least Square Method, Elbow Method

I. INTRODUCTION

Indonesia is in the confluence area of three active tectonic plates (the Indo-Australian Plate, the Eurasian Plate, and the Pacific Plate) so that the tectonic conditions in Indonesia are very active. The Indo-Australian Plate and the Eurasian Plate meet, causing subduction from the south of Bangladesh and coiling from west to south of Sumatra Island [6].

The existence of this subduction zone causes the western region of Sumatra Island to frequently experience earthquakes. One of the western parts of Sumatra Island that frequently experiences earthquakes is Nias Island. On Nias Island, there have been several earthquakes with large magnitudes recorded by IRIS such as the earthquake in 1974 (M 6.9), the earthquake in 1984 (M 7.1), the earthquake in 2005 (M 6.1), the earthquake in 2006 (M 6.3), earthquake in 2008 (M 6.2).

The danger of an earthquake cannot be avoided, but early mitigation efforts can reduce its impact. One of the mitigation strategies that can be carried out is to conduct a study of the seismicity level of the area concerned. Seismicity is the ratio of seismic activity in an area to seismic activity in other areas. There are two seismicity parameters, namely the degree of fragility of the rock (*b* value) and seismic activity level (*a* value). The seismicity level of an area can be assessed by grouping earthquake data and calculating the seismic parameter values for each group.

Earthquake data grouping can be done by clustering technique. Earthquakes that occur can be grouped into several cluster based on latitude and longitude epicenter position, so that earthquakes with nearby epicenters will be in one cluster. Clustering this earthquake aims to facilitate the analysis of the level of seismicity in the region.

K-means clustering is one of the clustering techniques which is iterative. The data is divided into K-cluster which has been determined by this algorithm. In K-means Clustering Algorithm, the amount cluster required is

cognizancejournal.com
Risna Simatupang *et al*, Cognizance Journal of Multidisciplinary Studies, Vol.3, Issue.5, May 2023, pg. 70-76
(An Open Accessible, Multidisciplinary, Fully Refereed and Peer Reviewed Journal)
ISSN: 0976-7797
Impact Factor: 4.843
Index Copernicus Value (ICV) = 76.35

not fixed to determine the amount of optimal cluster used Elbow Method. The working principle of the Elbow Method is to make a comparison graph between the percentage of variance to the amount of cluster. From the comparison graph of the percentage of variance to the amount of cluster the most obvious and the closest corner shape will be seen elbow and this point is considered the optimum K value [4].

Calculation of the seismicity parameter values in each cluster is done using the Least Square Method. The Least Square Method is a very simple method. This method can be used to estimate the value of a parameter to adapt a function to a set of data and for statistical analysis [3]. The result of the analysis of the seismicity level in each cluster can be used as a basis for determining areas prone to earthquakes that have the potential to damage.

Based on this description, a study of seismicity parameters in the Nias Island area was carried out using K-means Clustering Algorithm and Least Square Method. K-means Clustering Algorithm used to determine the amount of the optimum cluster that can be used for earthquake grouping. To calculate parameters *a* value and *b* value on each cluster used Least Square Method. The results of this study are expected to be the first step to minimize the impact of the earthquake on Nias Island.

II. METHODS

The object used in this research is Nias Island as in Fig. 1. Nias Island is located at coordinates 1°33'00" N-0°33'25" S and 97°00'00" E-98°48'29" E.

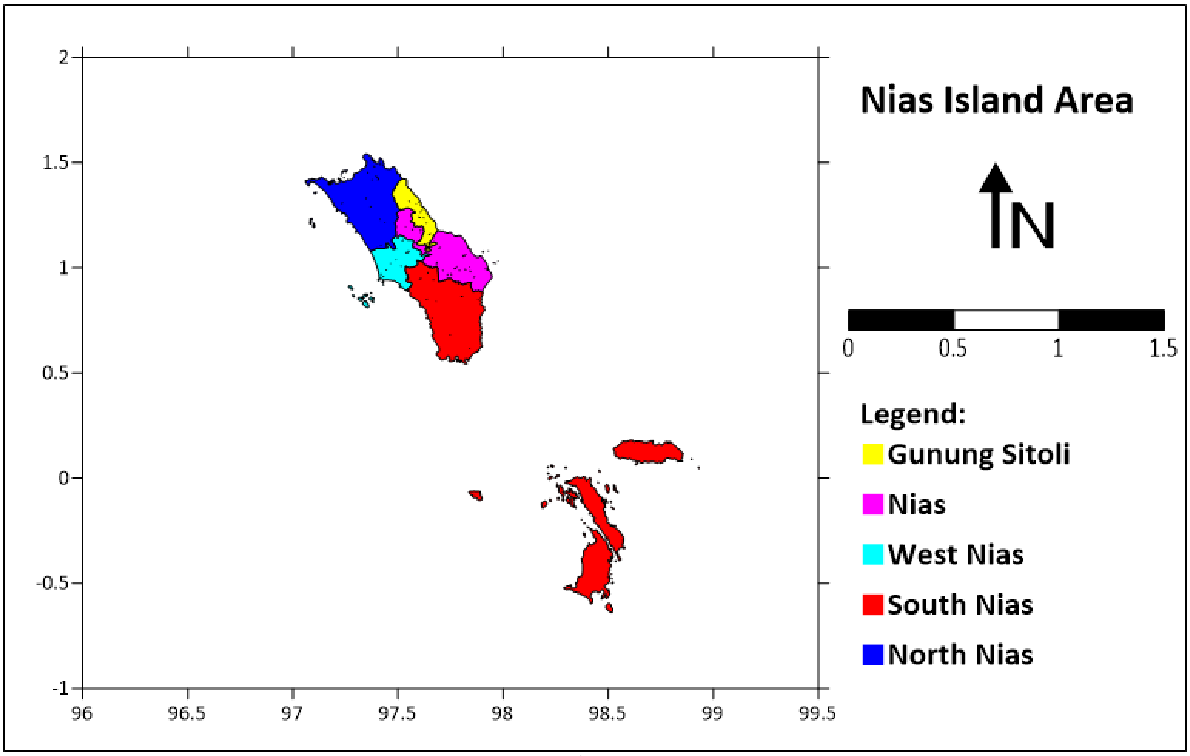


Fig. 1 Map of Nias Island area

The data used in this study is secondary data in the form of catalog of earthquake data on Nias Island from 1970 to August 2022 obtained from IRIS. Earthquake data needed is latitude, longitude, and magnitude data of the earthquake.

The data processing process in this study consists of several stages, namely: determining the amount of optimum cluster, clustering earthquake data, calculation of seismicity parameter values which include *a* value and *b* value and making seismicity maps. The results obtained are analyzed to determine the earthquake-prone areas that have the potential to damage on Nias Island.

ISSN: 0976-7797
Impact Factor: 4.843
Index Copernicus Value (ICV) = 76.35

III. Result and Discussion

A. Research Data

The research data was downloaded from the official IRIS website. The downloaded data is earthquake data in the Nias region with the range from 1970 to August 2022. The downloaded earthquake data consists of 10 variables, namely year, month, date, time of occurrence, latitude, longitude, depth, magnitude, occurrence area, and timestamp. Earthquake occurrence data is selected based on the occurrence area. Earthquakes whose occurrence areas are outside the Nias region are deleted or discarded so that the remaining earthquake data are earthquakes that occurred in the Nias region. This study will utilize 2,200 earthquake data with magnitudes greater than or equal to three ($M \geq 3$) based on the selection results.

B. Clustering Earthquake Data with K-means Clustering Algorithm

1) *Determining the Optimal Number of Clusters:* On K-means Clustering Algorithm required input parameters in the form of optimal K values. The optimal K value can be determined using the Elbow Method. By looking at the percentage of comparison results between the number of K that will form an elbow at a point, Elbow Method determines the best K value. Within Cluster Sum of Squares (WCSS) of each K value can be used to make comparisons between the K values. Fig. 2 is a graph obtained based on the Elbow Method. This graph shows a very large decrease in the WCSS value between the first K value and the second K value and then the WCSS value decreases slowly from the third K value to the last K value. The very large decrease in value results in a broken line resulting in an angle approaching the shape of an elbow at the second K value. The optimal K value obtained using this method is K=2.

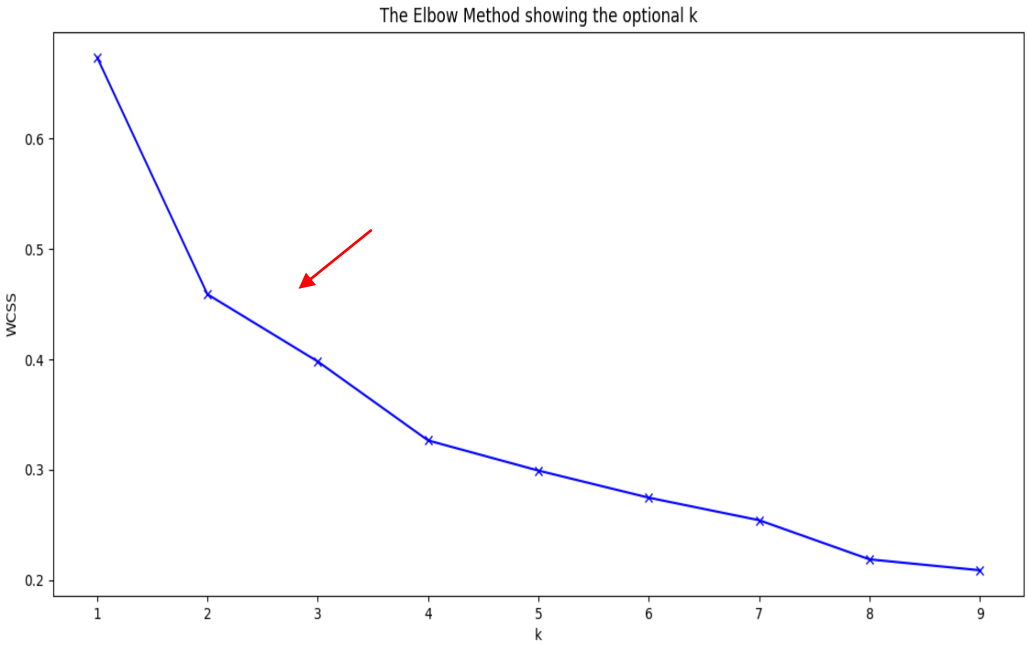


Fig. 2 Elbow graph (arrows indicate optimal K values)

cognizancejournal.com
Risna Simatupang *et al*, Cognizance Journal of Multidisciplinary Studies, Vol.3, Issue.5, May 2023, pg. 70-76
(An Open Accessible, Multidisciplinary, Fully Refereed and Peer Reviewed Journal)
ISSN: 0976-7797
Impact Factor: 4.843
Index Copernicus Value (ICV) = 76.35

2) *Results of Formed Clusters:* Earthquake data plot based on latitude and longitude before clustering process as in Fig. 3

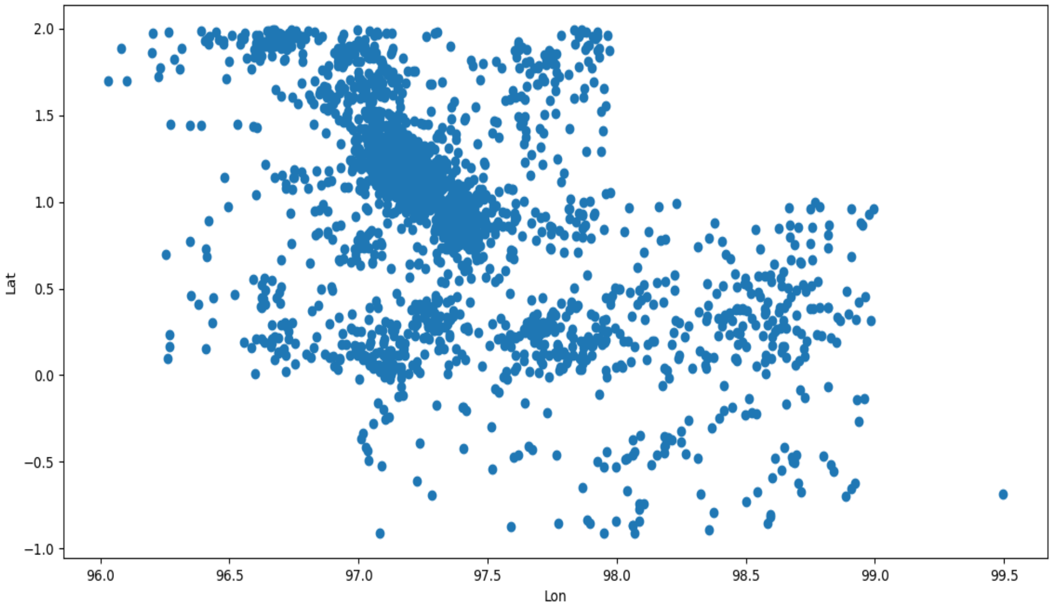


Fig. 3 Plot of earthquake data before clustering

K-means Clustering Algorithm in this study using the input value K=2 so that the earthquake data in this study will be divided into two clusters. This algorithm will predict the most appropriate cluster for each of the earthquake data based on latitude and longitude of the epicenter. Data frame in the form of cluster prediction for each earthquake data is used to create an earthquake data plot based on cluster division as in Fig. 4. On Fig. 4 the earthquake data is divided into two clusters that is cluster 0 and cluster 1. Cluster 0 is the part that is colored green and cluster 1 is the part colored in red.

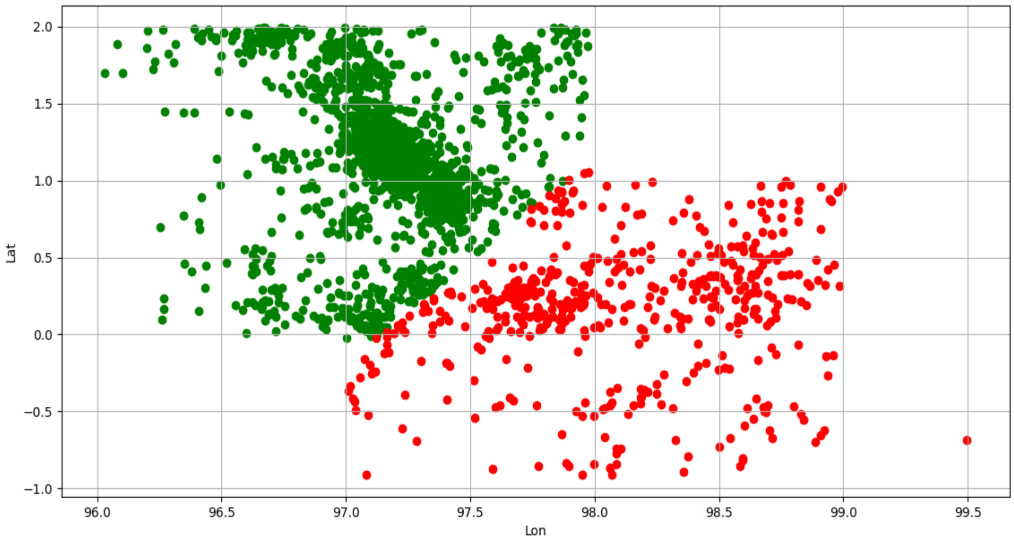


Fig. 4 Plots of cluster results formed using latitude and longitude data

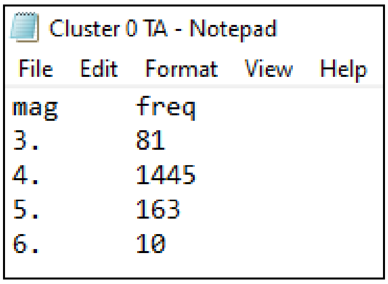
C. Parameter Seismicity

The seismicity pattern of an area can be known based on the analysis of the magnitude-frequency relationship with the following equation:

$$\text{Log N} = a - b\text{M} \quad (1)$$

with N is the number of earthquakes, M is the magnitude of the earthquake, whereas *a* and *b* are a seismicity parameter [2].

1) Seismicity Parameters in Cluster 0: Earthquake data on cluster 0 are sorted by magnitude from smallest to largest using Microsoft Excel software and arranged in the form of statistical data with txt format as shown in Fig. 5. The Least Square Method is used to calculate the value of the seismic parameter based on statistical data, so that the results are obtained as in Fig. 6. It shows that a value for cluster 0 is equal to 3.72 and b value for cluster 0 is equal to 0.37.



mag	freq
3.	81
4.	1445
5.	163
6.	10

Fig. 5 Earthquake statistical data cluster 0

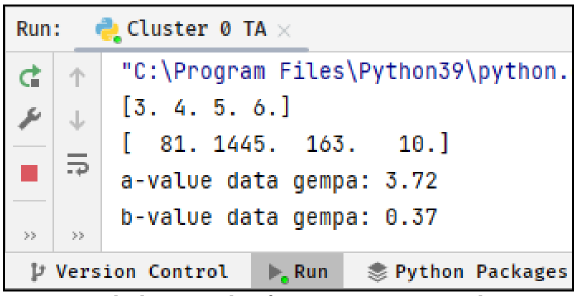
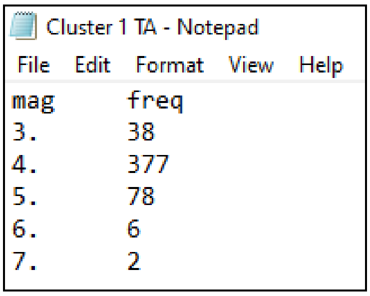


Fig. 6 Calculation results of seismicity parameters cluster 0

2) Seismicity Parameters in Cluster 1: Earthquake data on cluster 1 are sorted by magnitude from smallest to largest using Microsoft Excel software and arranged in the form of statistical data with txt format as shown in Fig. 7. The Least Square Method is used to calculate the value of the seismic parameter based on statistical data, so that the results are obtained as in Fig. 8. It shows that a value for cluster 1 is equal to 3.60 and b value for cluster 1 is equal to 0.44.



mag	freq
3.	38
4.	377
5.	78
6.	6
7.	2

Fig. 7 Earthquake statistical data cluster 1

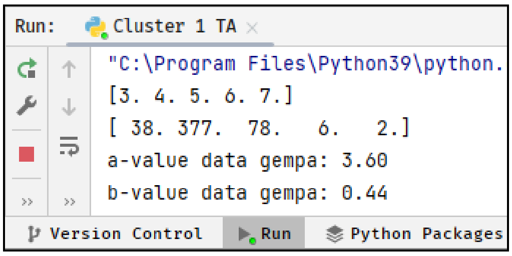


Fig. 8 Calculation results of seismicity parameters cluster 1

cognizancejournal.com
Risna Simatupang *et al*, Cognizance Journal of Multidisciplinary Studies, Vol.3, Issue.5, May 2023, pg. 70-76
(An Open Accessible, Multidisciplinary, Fully Refereed and Peer Reviewed Journal)
ISSN: 0976-7797
Impact Factor: 4.843
Index Copernicus Value (ICV) = 76.35

D. Analysis and Visualization of Final Research Results

Visualization of the seismicity map for the Nias Island area as shown in Fig. 10 obtained based on the plot clustering data on Fig. 9. Based on Fig. 10, the area of Nias Island is devided into two clusters, that is cluster 0 and cluster 1. Cluster 0 covers the area of Gunung Sitoli City, North Nias Regency, West Nias regency, most of Nias Regency, and a small part of South Nias Regency. Cluster 1 covers most of the area of South Nias Regency and a small part of Nias Regency.

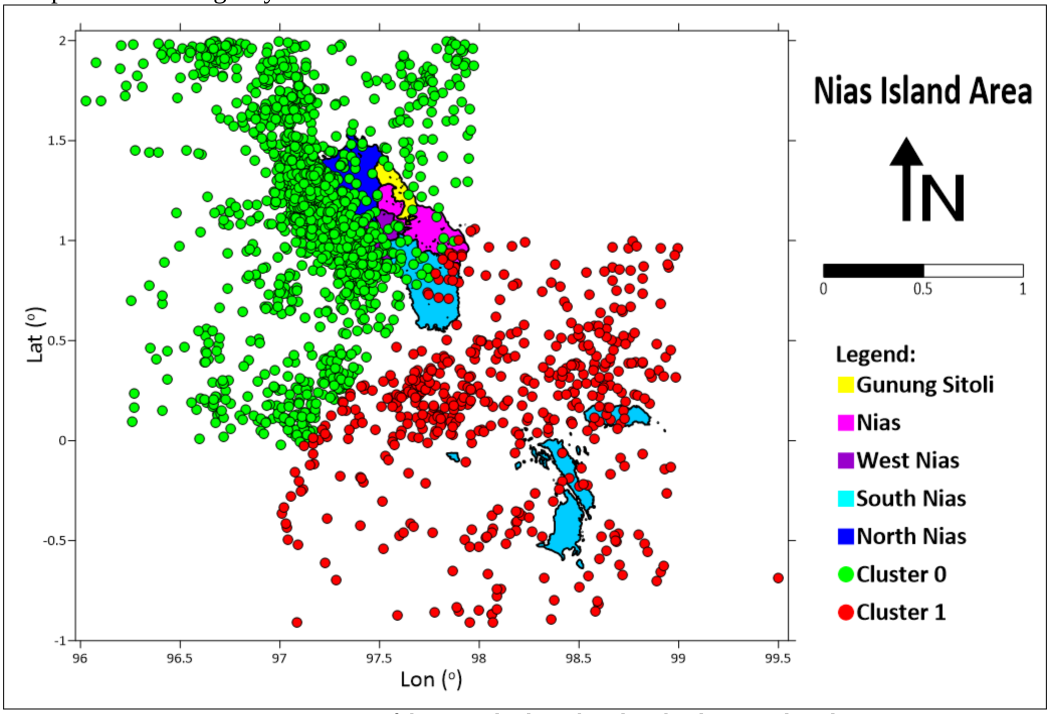


Fig. 9 Seismicity map of the Nias Island area based on the clustering data plots

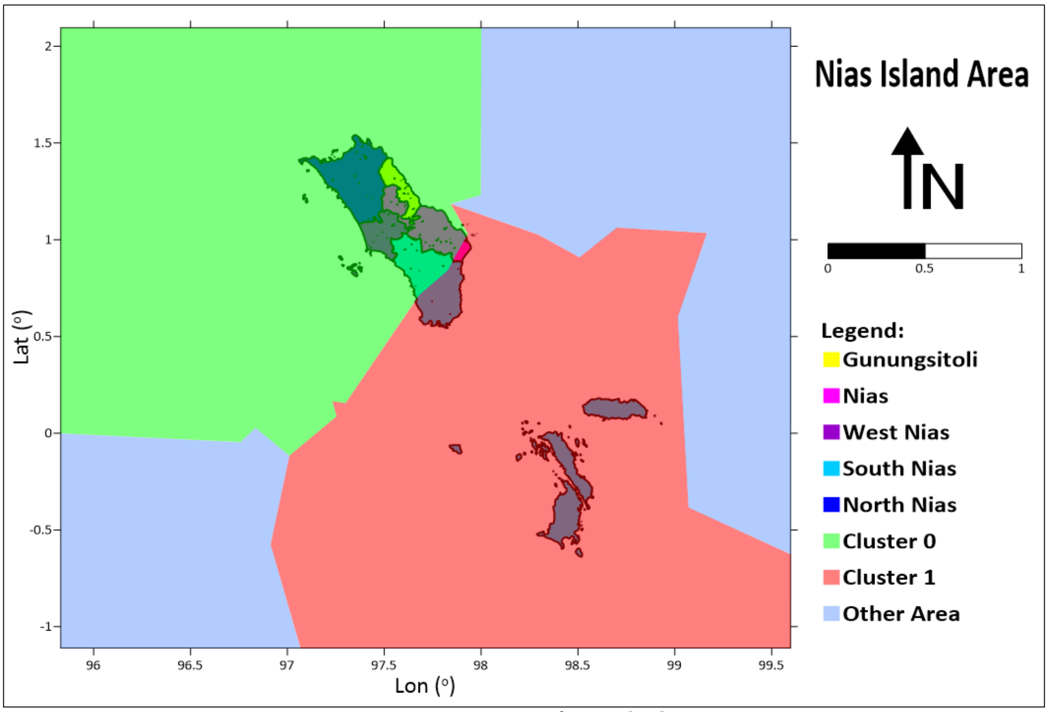


Fig. 10 Seismicity map of Nias Island area

Based on the calculation of seismicity parameters using the Least Square Method on each cluster obtained *a* value for cluster 0 is equal to 3.72 and *b* value for cluster 0 is equal to 0.37 meanwhile *a* value for cluster 1 is equal to 3.60 and *b* value for cluster 1 is equal to 0.44.

Seismic activity level or *a* value closely related to the number of earthquake events in the region. Constant *a* or *a* value that is low correlates with relatively low seismic activity, which means that there is an accumulation of energy in the region, thus vice versa for the region with high *a* value [5]. Based on the calculation results of the seismicity parameter values, *a* value of cluster 0 is higher than *a* value of cluster 1. This shows that the level of seismic activity at cluster 0 is higher than the seismic activity level of cluster 1. Based on statistical data, the number of earthquake occurrences in cluster 0 is more than the number of earthquakes in cluster 1. The results of this calculation are in accordance with the results of research that regions that have *a* value those with higher altitudes tend to experience more frequent earthquakes [5].

The degree of fragility of rock or *b* value is closely related to stress level of the rocks in the region. Constant *b* or *b* value that is low indicates the tall one stress level. High rocks stress level result in the accumulation of energy in the rocks in the area so that in that area there is a chance of a big earthquake. On the contrary the high *b* value indicates the low stress level; this indicates a major earthquake has occurred [1]. The calculation results show that *b* value of cluster 0 is lower than *b* value of cluster 1. This indicates that the rocks stress level in cluster 0 is higher than rocks stress level in cluster 1. Based on the results of the calculations that have been obtained, the probability of a large earthquake occurring in cluster 0 is higher than the probability of a large earthquake occurring in cluster 1. Based on statistical data, the magnitude range of earthquake occurrence data at cluster 0 is lower than the magnitude range of earthquake occurrence data at cluster 1. Constant *b* or *b* value which is higher in cluster 1 indicates that there has been a major earthquake on cluster 1. The results of this calculation are in accordance with the results of research that regions that have *b* value those with higher altitudes tend to experience earthquakes with larger magnitudes [1].

IV. CONCLUSIONS

K-means Clustering Algorithm classifies earthquake data from 1970 to August 2022 in the Nias Island area into two clusters that is cluster 0 and cluster 1. Cluster 0 covers the area of Gunung Sitoli City, North Nias Regency, West Nias regency, most of Nias regency, and a small part of South Nias Regency. Cluster 1 covers most of the area of South Nias Regency and a small part of Nias Regency. Based on the results of calculations and analysis of seismicity parameters in the Nias Island area, the areas prone to earthquakes that have the potential to damage are the areas located on cluster 0.

ACKNOWLEDGEMENT

We thank the Incorporated Research Institutions for Seismology (IRIS) for their use of the dataset. We also thank Diponegoro University for its continuous support. We appreciate anonymous reviewers and editors for useful comments that help improve the quality of articles.

REFERENCES

[1]. Alabi, A. A., Akinyemi, O. D., Adewale, A., 2013, Seismicity Pattern in Southern Africa from 1986 to 2009, *Earth Science Research*, Vol 2, No 2, 2013.
[2]. Gutenberg, B., and Richter, C.F, 1954, *Seismicity of Earth and Associated Phenomenon*, Princeton University Press, Princeton.
[3]. Herve, A., 2003, *Least Square*, The University of Texas, Dallas.
[4]. Madhulatha, T. Soni, 2012, An Overview on Clustering Methods, *IOSR Journal of Engineering*, Vol 2, No 4, 719-725.
[5]. Rohadi, S., 2009, Studi Seismotektonik Sebagai Indikator Potensi Gempa Bumi di Wilayah Indonesia, *Jurnal Meteorologi dan Geofisika*, 10, 2, 111-120.
[6]. Thea, A.S., and Mutiara, I.A., 2014, Analisa Anomali Gayaberat Terhadap Kondisi Tatanan Tektonik Zona Subduksi Sunda Megathrust di Sebelah Barat Pulau Sumatera, *Geoid*, Vol 10, No 01, 75-80.