

# Grouping of Flood Victim Data Based on Damage Rate Using K-Means Algorithm Case Study: Binjai City Social Service

Melisa Tri Yanti<sup>1\*</sup>, Yani Maulita<sup>2</sup>, Kristina Annatasia Br Sitepu<sup>3</sup>

<sup>1,2,3</sup> STMIK Kaputama, Indonesia

[melisatriyanti919@gmail.com](mailto:melisatriyanti919@gmail.com)<sup>1\*</sup>, [yanimaulita26@gmail.com](mailto:yanimaulita26@gmail.com)<sup>2</sup>, [kannatasia88@gmail.com](mailto:kannatasia88@gmail.com)<sup>3</sup>

## Abstract

This study aims to classify flood disaster victim data in Binjai City based on three main variables: sub-district or location, damage level, and type of aid received. The data were obtained from the Binjai City Social Service in 2024 and processed using the K-Means Clustering method with the Matlab R2014b application. The stages include data transformation, determining the number of clusters, selecting initial centroids, calculating Euclidean distance, and evaluating the results. Tests were conducted with configurations of 2, 3, 4, and 5 clusters. In the 2-cluster configuration, the distinction was observed between areas with low damage and limited to moderate aid, and areas with medium damage and more extensive aid. In the 3-cluster configuration, the second test produced the most optimal cluster in Kartini Sub-district with light damage and limited food aid. In the 4-cluster configuration, the most compact cluster was found in Setia Sub-district with medium damage and aid in the form of food and blankets. In the 5-cluster configuration, the most specific result was obtained in Rambung Barat Sub-district with medium damage and aid in the form of food and blankets. These findings indicate that the 5-cluster configuration provides more detailed and targeted classification, serving as a strategic reference for aid distribution.

**Keywords:** K-Means Clustering, Flood Disaster, Sub-district, Damage Level, Type of Aid, Matlab, Binjai City

## 1. Introduction

Floods are one of the most frequent natural disasters in Indonesia, and their impact is not only felt in terms of the environment, but also causes significant social and economic losses. Binjai City as an area with high rainfall and a drainage system that is not optimal, is included in flood-prone areas. Data from the Binjai City BPBD shows that in the last five years, there has been an increase in the frequency and scale of floods, with many affected areas experiencing light to severe damage to infrastructure and residents' settlements [1].

To improve the effectiveness of disaster management, a data-driven approach is needed to understand the characteristics of floods and the extent of damage caused. One approach that can be used is to apply the Clustering or data grouping. With Clustering, flood data can be grouped based on the level of damage caused such as light, medium, to severe damage. The results of this grouping can help in the preparation of more targeted risk handling, recovery, and mitigation strategies [2]

## 2. Literature Review

### 2.1. Flood

In addition to natural factors and human behavior, the increase in population is also a significant cause of flooding. The expansion of residential land that reduces the water catchment area makes the rainwater flow not well accommodated, causing inundation and flooding. In areas such as Bekasi, the problem of flooding is a complex problem and requires 34 collective handling, not only from the government but also from the active participation of the community. Therefore, the use of technology such as the grouping of flood-prone areas with the K-Means algorithm is one of the solutions to support a more precise and targeted disaster mitigation system ([3]).

### 2.2. Data Mining

According to the journal [4] "Algorithm Implementation K-Means Clustering In the Data Grouping of House Damage Due to Natural Disasters in Cirebon Regency", data mining is a data exploration method that aims to build a model to identify data patterns that are not visible in the data that has been stored. This process also includes grouping data to understand these patterns and take follow-up actions according to the findings, so as to provide support in evaluation activities as expected. In general, data mining is an automated or semi-automated process to find useful knowledge from large data sets, and is included in the stages Knowledge Discovery in Databases (KDD).

### 2.3. Clustering

According to [5] method Clustering is a method of grouping data into multiple Cluster or groups where the data in Cluster have maximum similarity and data between Cluster have minimal similarities. Clustering used for grouping or identifying data or information that has certain characteristics.

### 2.4. Matlab

According to [6] MATLAB is a numerical computing tool that is widely used to solve mathematical problems efficiently, especially in the form of data analysis and visualization Numerical. In the context of this study, MATLAB is used as a tool to analyze the problem of ordinary differential equations (GDP) of order one using the numerical method. MATLAB is preferred for its ability to process data quickly and accurately and supports many numerical methods including Runge-Kutta, Euler, and modification. In addition, MATLAB also has a graphical interface (GUI) feature that can make it easier for users, both students and practitioners, to analyze and visualize the results of mathematical calculations.

## 3. Research Methods

Process stages Knowledge Discovery in Database or KDD is explained in the book It can be seen in Figure 2.2 namely:

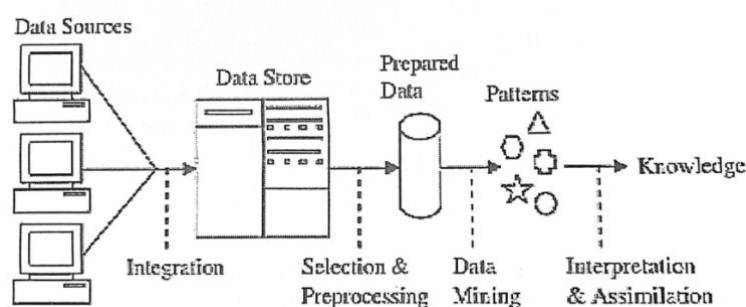


Fig. 1: Stages of Data Mining [7]

1. Data Cleanup – Remove inconsistent or irrelevant noise and data.
2. Data Integration – Combining data from various databases into a single unit.
3. Data Selection – Selects only relevant data for analysis.
4. Data Transformation – Changing or adjusting the format of the data to match the data mining method.
5. Process Mining – The core stage to discover hidden knowledge from data.
6. Pattern Evaluation – Assessing the resulting pattern or model to fit the hypothesis.
7. Knowledge Presentation – Presenting data mining results visually and informatively for users.

### 3.1. Data Analysis

According to [5] K-Means Clustering is a method Non-hierarchical clustering which divides the data into one or more groups, where data that has similar characteristics is grouped together, and data with different characteristics is placed in another group. The term K refers to the number of clusters desired, while means refers to the average value of a data group which in this case is called a cluster. Thus, K-Means Clustering is a data analysis or data mining method that works unsupervised and uses a partition system to perform data clustering. This method is useful for separating data into groups that have different characteristics from each other.

There are various ways to measure the degree of similarity between objects, one of which is using weighted Euclidean distance. This method calculates the distance between two points by paying attention to the value of each attribute at both points. The Euclidean distance formula is:

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad (1)$$

Information:

d = distance between x and y

x = kluster center data

y = data on the attribute

i = each data

n = amount of data

$x_i$  = data at the center of the cluster i

$y_i$  = data on each data to i

## 4. Results and Discussion

The analysis process of grouping flood data based on the level of damage is carried out by applying the K-Means algorithm, thus facilitating the data classification process. With this method, the number of groups can be identified based on the category of damage level, assistance received, as well as various other supporting factors.

### 4.1. Discussion of Input Data

The data is input into the system in the form of data on flood victims in 2023-2024 obtained from the Binjai City Social Service. The data will be transformed based on the transformation value of each variable used. The following is an explanation of the input data, the variables used, and the value of the transformation applied, namely:

1. Data Input

- Amount of data : 600
- Variable : X Neighborhoods
- Y Damage Rate
- Z Assistance Received

2. Clustering Clusters : 3 Clusters

3. Flood Victim Data and Transformation Value

**Table 1:** Data Transformation Value

NO	VARIABLE	TRANSFORMATION	VALUE TRANSFORMATION
		Because	1
		Binjai Estate	2
		Faithful	3
		Kartini	4
		Rambung Dalam	5
		Tribute	6
		West Rambung	7
		Mencirim	8
		Berngam	9
		Binjai Town	10
		Tangsi	11
		The city of senembah	12
		Limes Mungkur	13
		Lemonade	14
		São Paulo, Sã	15
1.	Neighborhoods	Likes to be progressive	16
		Sukaramai	17
		Source of work	18
		Source mulyorejo	19
		Highlands	20
		Weight Loss	21
		Bhakti works	22
		East Rambung	23
		Red soil	24
		The Land of a Thousand	25
		Turi cloves	26
		Peace	27
		Utopian teak	28
		National identity	29
		Pepper garden	30
		Jackfruit	31
		Hero	32
2.	Damage Rate	Minor Damage	1
		Moderately Damaged	2
		Severely Damaged	3
		Food	1
		Family Kit	2
		Blanket	3
		Mattress	4
		Tent	5
		Food, Family Kit	6
		Food, Blankets	7
		Food, Mattress	8
		Food, Tent	9
3.	Assistance Received	Family Kit, Blanket	10
		Family Kit, Mattress	11
		Family Kit, Tent	12
		Blankets, Mattresses	13
		Blankets, Tents	14
		Mattress, Tent	15
		Food, Family Kit, Blankets	16
		Meals, Family Kits, Mattresses	17
		Food, Family Kit, Tent	18
		Food, Blankets, Mattress	19

NO	VARIABLE	TRANSFORMATION	VALUE TRANSFORMATION
		Food, Blankets, Tents	20
		Food, Mattress, Tent	21
		Family Kit, Blanket, Mattress	22
		Family Kit, Blanket, Tent	23
		Family Kit, Mattress, Tent	24
		Blankets, Mattresses, Tents	25
		Food, Family Kit, Blankets, Mattress	26
		Food, Family Kit, Blankets, Tent	27
		Food, Family Kit, Mattress, Tent	28
		Food, Blankets, Mattresses, Tents	29
		Family Kit, Blanket, Mattress, Tent	30
		Food, Family Kit, Blankets, Mattresses, Tents	31

Table 2: Flood Victim Data

NO	Neighborhoods	Damage Rate	Assistance Received	X	Y	Z
1.	Faithful	Moderately Damaged	Food, Family Kit	3	2	6
2.	Faithful	Minor Damage	Family Kit	3	1	2
3.	Faithful	Moderately Damaged	Food, Family Kit, Tent	3	2	18
4.	Faithful	Moderately Damaged	Food, Family Kit	3	2	6
5.	Faithful	Moderately Damaged	Food, Family Kit	3	2	6
6.	Faithful	Moderately Damaged	Food, Family Kit	3	2	6
7.	Faithful	Moderately Damaged	Food, Blankets	3	2	7
8.	Faithful	Moderately Damaged	Food, Family Kit	3	2	6
9.	Faithful	Moderately Damaged	Food, Family Kit	3	2	6
10.	Faithful	Moderately Damaged	Food, Family Kit	3	2	6
11.	.....	.....	.....	....	....	....
591.	Faithful	Moderately Damaged	Food, Family Kit, Blankets, Tent	3	2	27
592.	Faithful	Moderately Damaged	Meals, Family Kits, Mattresses	3	2	17
593.	Faithful	Moderately Damaged	Food, Family Kit, Blankets, Tent	3	2	27
594.	Faithful	Moderately Damaged	Food, Family Kit, Tent	3	2	18
595.	Faithful	Moderately Damaged	Food, Family Kit, Blankets, Tent	3	2	27
596.	Faithful	Moderately Damaged	Food, Family Kit, Blankets, Tent	3	2	27
597.	Faithful	Moderately Damaged	Meals, Family Kits, Mattresses	3	2	17
598.	Faithful	Moderately Damaged	Food, Family Kit, Blankets, Tent	3	2	27
599.	Faithful	Moderately Damaged	Food, Family Kit, Tent	3	2	18
600.	Faithful	Moderately Damaged	Food, Family Kit, Blankets, Tent	3	2	27

### 4.2. Interface Discussion

In the design of this interface, it will be explained about the results of program planning that uses GUIDE Matlab and can be seen as follows:

1. In the initial Home Menu interface display, which contains a clustering process menu for 2 variables and 3 variables, but in the context of my research I used the K-Means method with  $k = 3$  and 3 variables, here is the picture of the display:



Fig 2. Home Menu

2. Clustering Process Menu

The test will be carried out by processing input data using 2, 3, 4 and 5 Clusters. The view below is a view of the results of the data grouping process with 3 Clusters by applying the clustering method that has been processed, a graph of the data from the grouping results that provide information in the form of data grouping coordinate points and data cluster centers.

a. Testing for 2 Clusters

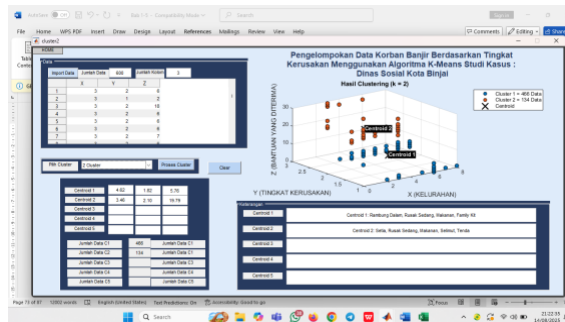


Fig 3: Testing 1 for 2 Clusters

Testing 2 clusters is done once because the number of clusters is predetermined, resulting in Cluster 1 (466 data) with low damage and little to medium assistance, and Cluster 2 (134 data) with moderate damage and more assistance. The main difference lies in the level of damage and the amount of help received, where cluster 2 gets more help than cluster 1.

b. Testing for 3 Clusters

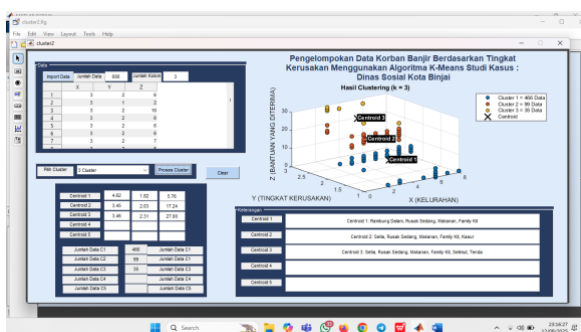


Fig 4: 1 to 3 Cluster Testing

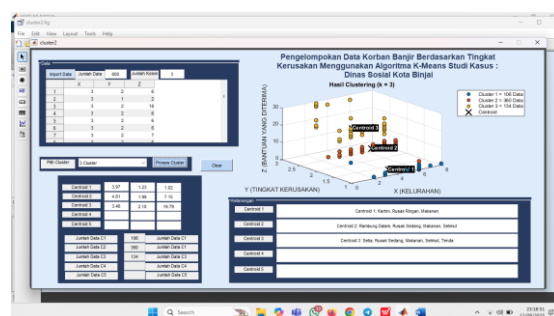


Fig 5: 2 to 3 Cluster Testing

Of the two tests with 3 clusters, in the 2nd test showed that the first cluster was the most optimal, this showed that the data in the cluster had a high level of similarity. This cluster consists of flood victims, the majority of whom are located in Kartini Village, with a minor level of damage. The assistance received in this group is only in the form of food, because it is adjusted to the level of need and urgency, where food assistance is considered sufficient and on target for the condition of those who are mildly affected.

c. Testing for 4 Clusters

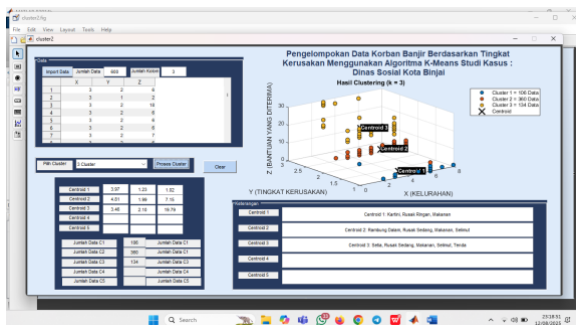


Fig 6: Testing 1 for 4 Clusters

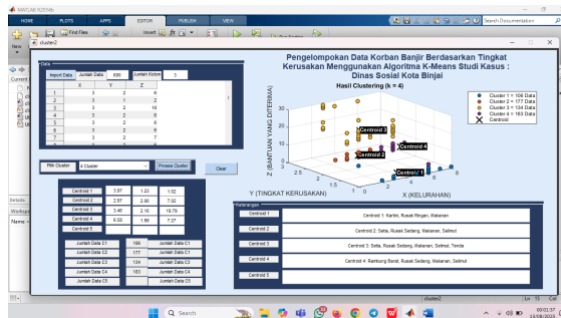


Fig 7: 2 to 4 Cluster Testing

When 2 tests were carried out for grouping into 4 clusters, in the 2nd test the data often appeared Cluster 2 as the group with the highest cohesiveness. This cluster is dominated by flood victims in Setia Village, with a moderate level of damage and received assistance in the form of food and blankets. This shows that victims in this cluster can still survive at home or in the surrounding environment without the need for evacuation tents, but still need food assistance and protection from the weather during the recovery period.

d. Testing for 5 Clusters

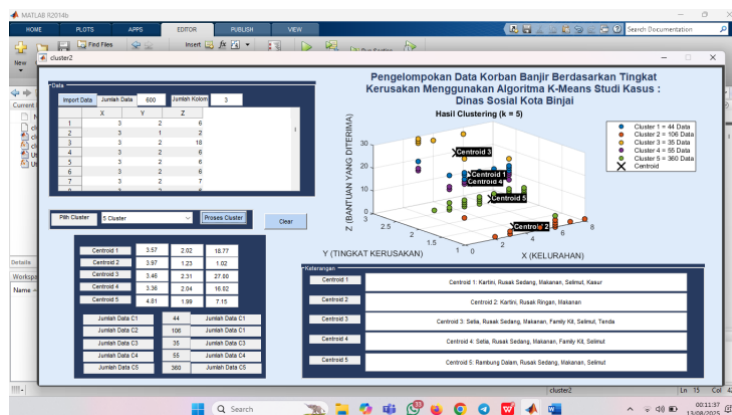


Fig 8: Testing 1 for 5 Clusters

Of the 600 data, 5 Clusters were obtained, of which Cluster 1 amounted to 44 data, Cluster 2 amounted to 35 data, Cluster 3 amounted to 105 data, Cluster 4 amounted to 56 data, and Cluster 5 amounted to 360 data. Here is a description of the cluster center on the graph:

1. (3.57, 2.02, 18.77)

It can be known that Cluster 1 is the Kartini area with a moderate level of damage and received sufficient assistance, with the dominant type of assistance being food, blankets.

2. (3.97, 1.23, 1.02)

It can be known that Cluster 2 is the Kartini area with a level of Light Damage and receives very little assistance, with only food.

3. (3.36, 2.31, 17.02)

It can be known that Cluster 3 is a loyal village area with a moderate level of damage and receives considerable assistance, with the dominant type of assistance being Food, Family Kit, Blankets, Tents.

4. (4.81, 1.99, 7.15)

It can be known that Cluster 4 is a loyal area with a moderate level of damage and receives moderate assistance, with the type of food, family kit, and mattress assistance.

5. (3.45, 2.03, 17.24)

It can be known that Cluster 5 is the Rambung Dalam area with a moderate level of damage and received considerable assistance, with the dominant type of assistance in the form of food, blankets.

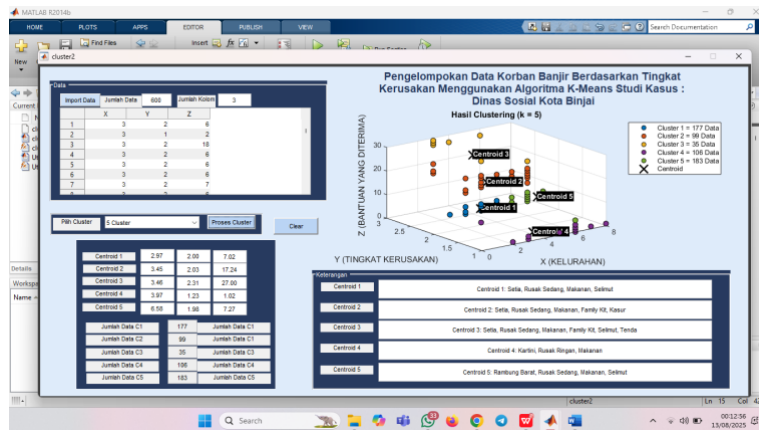


Fig 9: Testing 2 for 5 Clusters

Of the 600 data, 5 Clusters were obtained, of which Cluster 1 amounted to 177 data, Cluster 2 amounted to 99 data, Cluster 3 amounted to 35 data, Cluster 4 amounted to 106 data, and Cluster 5 amounted to 183 data. Here is a description of the cluster center on the graph:

1. (2.97, 2.00, 7.02)

Cluster 1 consisting of 177 victims, with the majority located in Setia Village, with a moderate level of damage and dominant assistance in the form of (food, blankets), this shows that victims in this cluster indicate that victims are generally still able to live at home or in the surrounding environment, but still need food assistance and protection from the weather during the recovery period.

2. (3.45, 2.03, 17.24)

Cluster 2 has 99 flood victim data, the majority of which are located in Setia Village, with a moderate level of damage. Obtaining assistance received in this group is in the form of food, family kits, mattresses, because the victim suffered quite serious damage that affected basic needs and comfort of life. Family kits and mattresses are provided to help victims meet their personal needs and get a decent rest during recovery.

3. (3.97, 1.23, 1.02)

Cluster 3 consists of 35 flood victims, the majority of locations in Setia Village, with moderate levels of damage. Obtaining assistance received in the form of food, family kits, blankets, tents) because victims need support for basic needs, personal protection, and temporary housing. This combination of assistance shows that the impact of flooding is quite disruptive to daily activities and the comfort of residence.

4. (6.56, 1.98, 7.27)

Cluster 4 consists of 106 flood victims, the majority of whom are located in Kartini Village, with a minor level of damage. The assistance received in this group is only in the form of food, because it is adjusted to the level of need and urgency, where food assistance is considered sufficient and on target for the condition of those who are mildly affected.

5. (3.50, 2.23, 12.23)

Cluster 5 has 183 data on flood victims, the majority of which are located in West Rambung Village, with moderate levels of damage. Obtaining assistance received from this group in the form of food and blankets indicates that victims are generally still able to live at home or in the surrounding environment, but still need food assistance and protection from the weather during the recovery period.

In clustering into 5 clusters, the data that frequently appears in the 2nd test Cluster 4 returns to being the most compact, which is the lowest value of all the clustering configurations tested. This cluster consists of victims located in West Rambung Village with a moderate damage category. The type of assistance received, namely food and blankets, indicates that the victims did not experience total homelessness, but were affected significantly enough that they needed basic support to maintain health and comfort, especially in the face of uncertain post-flood weather.

## 5. Conclusion

Based on the results of the research related to the grouping of business types based on the assistance received in Binjai City using the K-Means algorithm, the following conclusions were obtained:

1. The application of the K-means algorithm has proven to be successfully implemented in a program based on Matlab R2014b for the grouping of flood victim data based on the level of damage using the K-Means algorithm.

2. Based on the analysis that has been carried out, namely the grouping of flood victim data based on the level of damage, the results of clustering with 2, 3, 4, and 5 clusters were obtained. Each cluster shows a different pattern related to the village, the level of damage, and the type of assistance received.
3. Based on the results of testing 2, 3, 4, and 5 clusters, a consistent grouping pattern was obtained that differentiated flood victims based on the level of damage and the type of assistance received. In the 2 clusters, a clear difference was seen between the group with low damage and little help and the group with moderate damage and more assistance. In the 3 clusters, the second test produced the most optimal cluster, namely Kartini Village with minor damage and limited assistance in the form of food as needed. In 4 clusters, the compact cluster in the second test was found in Setia Village with moderate damage and food assistance and blankets, while in 5 clusters, the most compact cluster in the second test was West Rambung Village with moderate damage and food assistance and blankets. Overall, second tests on each cluster configuration tend to provide more specific and targeted results.

## References

- [1] W. O. Bakara, Sugianto, and N. Ahmadi, "BPBD Kota Binjai Kota Binjai," vol. 6, no. 1, pp. 71–82, 2023.
- [2] R. F. Purba, A. A. Panjaitan, L. T. Butar-butur, J. F. E. D. L. Sitorus, R. Rumapea, and I. M. S. S, "Implementation of k-means clustering to determine the level of flood-prone disasters in the North Sumatra region," vol. 4, no. 2, pp. 210–215, 2024.
- [3] M. M. Effendi and A. Siswandi, "Analysis of Flood Prone Prediction with K-Means Algorithm," *Journal of Information System Research (JOSH)*, vol. 5, no. 2, pp. 697–703, 2024, doi: 10.47065/josh.v5i2.4770.
- [4] A. Maulana, R. Danar Dana, and N. Dienwati Nuris, "Implementation of K-Means Clustering Algorithm in the Grouping of Data on House Damage Due to Natural Disasters in Cirebon Regency," *JATI (Student Journal of Informatics Engineering)*, vol. 8, no. 2, pp. 1417–1424, 2024, doi: 10.36040/jati.v8i2.9024.
- [5] M. F. Al Halik and L. Septiana, "Data Analysis for Prediction of Natural Disaster-Prone Areas in West Java Using K-Means Clustering Algorithm," *Journal of Information System, Applied, Management, Accounting and Research*, vol. 6, no. 4, pp. 856–870, 2022, doi: 10.52362/jisamar.v6i4.939.
- [6] A. M. Nugraha and N. Nurullaeli, "Matlab Graphical User Interface (GUI) for Solving First-Order Ordinary Differential Equations," *National Seminar on Research and Technology (National Seminar on Research and Technological Innovation)*, vol. 7, no. 1, pp. 182–185, 2023, doi: 10.30998/semnasristek.v7i1.6269.
- [7] M. K. Juniar Hutagalung, S.Kom., *Combination of K-Means Clustering and MOORA Method*. Medan: CV BUDI UTAMA, 2021.