Journal of Computer and Information Systems Ampera
Vol. 3, No. 3, September 2022 e-ISSN: 2775-2496
https://journal-computing.org/index.php/journal-cisa/index

Implementation of K-Means Method for Clustering Natural Disaster Priority Areas in North Sumatra

Devi Sari Delima Tumanggor[1], Manna Maria Sianipar[2], Sardo Pardingotan Sipayung[3]

[1, 2]Informatics Engineering Departement , Catholic University of Saint Thomas, Medan, Indonesia
[3]Data Science Departement, Catholic University of Saint Thomas, Medan, Indonesia
Email: [1]devisaridelima23@gmail.com, [2]mannaliisecond@gmail.com, [3]pinsarsiphom@gmail.com

Abstract

Indonesia is highly vulnerable to natural disasters due to its geographical and geological conditions, including North Sumatra Province, which experiences diverse disaster risks. This study aims to identify and group natural disaster priority areas in North Sumatra using a data-driven approach. The research employs a quantitative method by applying the K-Means clustering algorithm to group regions based on disaster occurrence and impact characteristics. Secondary data of natural disaster events in 2024 were processed through data preprocessing stages, including data cleaning and normalization. The clustering process was conducted using five clusters representing very low, low, medium, high, and very high disaster priority levels. The quality of clustering results was evaluated using the Davies–Bouldin Index (DBI). The results show that most regions fall into the very low and low priority clusters, while several regions are classified into higher priority levels. Deli Serdang Regency is identified as the very high priority area, indicating the highest disaster impact among regions. The DBI value of 1.102688 indicates that the formed clusters have reasonable compactness and separation. This study concludes that the K-Means method is effective for grouping disaster-prone areas and can support data-driven decision-making in disaster mitigation and planning in North Sumatra Province.

Keywords: K-Means clustering, natural disaster, disaster priority, North Sumatra, data mining

1. INTRODUCTION

Indonesia is a country with a high level of vulnerability to natural disasters due to its geographical and geological characteristics. Located along the

Vol.7, No. 1, Januari 2026 e-ISSN: 2775-2496
https://journal-computing.org/index.php/journal-cisa/index

Pacific *Ring of Fire*, Indonesia frequently experiences earthquakes, volcanic eruptions, floods, and landslides. North Sumatra Province is one of the regions with significant disaster potential, as it is characterized by active fault zones, diverse topography, mountainous areas, and high rainfall intensity [1]. These conditions highlight the importance of identifying and prioritizing disaster-prone areas to support effective disaster mitigation and risk reduction efforts.

Previous studies have applied various approaches to analyze and map disaster-prone areas. Geographic Information Systems (GIS) are widely used to visualize the spatial distribution of disaster risks [2]. In addition, statistical methods have been employed to analyze factors influencing disasters, such as rainfall, slope, land use, and population density. However, these conventional approaches often face limitations in handling complex and large-scale datasets.

With the advancement of information technology, data mining and machine learning techniques have been increasingly applied in disaster analysis. One commonly used unsupervised learning method is K-Means clustering, which groups data based on similarity without requiring predefined labels [3]. Several studies have demonstrated that the K-Means method can be effectively used to cluster disaster-prone areas, including flood-prone, landslide-prone, and earthquake-prone regions based on environmental parameters and historical disaster data [4][5].

Despite its widespread use, most previous studies using the K-Means method focus on a single type of disaster or are limited to small geographic areas. Furthermore, some studies do not integrate multiple disaster indicators comprehensively, resulting in clustering outcomes that are less effective for prioritizing regions in disaster management planning. Research specifically addressing disaster priority clustering at the provincial level, particularly in North Sumatra, remains limited.

Based on these research gaps, this study aims to implement the K-Means method to cluster natural disaster priority areas in North Sumatra. This research integrates multiple disaster-related parameters to classify regions into different priority levels. The novelty of this study lies in its comprehensive clustering approach for determining disaster priority areas at the provincial level, tailored to the specific characteristics of North Sumatra.

The main research problem addressed in this study is how to effectively group regions in North Sumatra into natural disaster priority clusters using the K-Means method. The results of this study are expected to provide structured and data-driven information to support decision-making in disaster mitigation, preparedness, and resource allocation.

2. METHODS

This study employs a quantitative approach by applying data mining techniques. The methods used consist of clustering and classification techniques to determine natural disaster priority areas. Clustering is used to group regions based on similarity characteristics, while classification is applied to assign priority levels to each region.

The data used in this study are secondary data obtained from the National Disaster Management Agency (BNPB) through the BNPB One Data Portal. The dataset consists of natural disaster occurrence data in North Sumatra Province for the year 2024, including the number of disaster events in each regency/city [6]. Supporting data related to regional characteristics are obtained from Statistics Indonesia (BPS) of North Sumatra Province [7].

Before the analysis process, the data undergo a preprocessing stage to improve data quality. Data preprocessing includes data selection, data cleaning to handle missing or inconsistent values, and data normalization to ensure that all variables have comparable scales and do not dominate the analysis results [8].

2.1. Data Collection

Data collection is the initial stage of the research, which aims to obtain relevant data according to the research objectives. In this study, data collection is carried out by collecting natural disaster occurrence data for the year 2024 from the BNPB One Data Portal. The data include the number of disaster events for each regency/city in North Sumatra Province. In addition, supporting data related to regional characteristics are obtained from Statistics Indonesia (BPS) of North Sumatra Province [6][7].

2.2. Data Preprocessing

Data preprocessing is the stage of preparing data to ensure it is suitable for analysis and capable of producing accurate results. This stage includes selecting relevant attributes, cleaning data to handle missing or inconsistent values, and normalizing data so that each variable has an equal scale and does not disproportionately influence the analysis results [8].

In this study, data normalization is performed using the Min–Max normalization method, which transforms attribute values into a range between 0 and 1 to eliminate scale differences among variables [9]. This normalization ensures that all attributes contribute proportionally to the distance calculation in the clustering and classification processes.

The Min–Max normalization formula is defined as follows:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

where: x' is the normalized value, x is the original data value, $x_{\min}$ is the minimum value of the attribute, $x_{\max}$ is the maximum value of the attribute.

2.3. Determination of the Number of Clusters

Determination of the number of clusters is a stage aimed at defining how many groups will be formed in the clustering process. In this study, the number of clusters (k) is set to five clusters, which represent five levels of natural disaster priority, namely very low priority, low priority, medium priority, high priority, and very high priority. The use of five clusters allows a more detailed and representative classification of disaster-prone areas in North Sumatra Province.

2.4. K-Means Clustering Algorithm

The K-Means clustering algorithm is an unsupervised learning method used to partition data into $k$clusters based on similarity characteristics [9]. The algorithm works by minimizing the distance between data points and the centroid of their assigned cluster.

Vol.7, No. 1, Januari 2026 e-ISSN: **2775-2496**

https://journal-computing.org/index.php/journal-cisa/index

The clustering process begins with the initialization of $k$centroids, followed by distance calculation between each data point and the centroids using the Euclidean distance measure. Each data point is assigned to the nearest cluster, and new centroid values are calculated based on the mean of the data points within each cluster. This process is repeated iteratively until the cluster assignments no longer change or convergence is achieved.

The Euclidean distance formula used in the K-Means algorithm is defined as follows:

$$D(x_i, x_j) = \sqrt{\sum_{k=1}^{n} \left(x_{ik} - x_{Jk}\right)^2} \quad (2)$$

where: x_i represents the test data, x_j represents the training data, n is the number of attributes used.

2.5. Davies–Bouldin Index (DBI)

The Davies–Bouldin Index (DBI) is an internal evaluation method used to measure the quality of clustering results by considering both *intra-cluster compactness* and *inter-cluster separation*. A lower DBI value indicates better clustering quality, as it reflects clusters that are more compact and better separated from each other [11].

In this study, the DBI is used to evaluate the clustering results obtained from the K-Means method for grouping natural disaster priority areas. The evaluation is performed after the clustering process reaches convergence. The DBI value is calculated based on the average ratio between the within-cluster distances and the distances between cluster centroids [12].

Mathematically, the Davies–Bouldin Index is defined as follows:

$$DBI = \frac{1}{k} \sum_{i=1}^{k} \max_{j \neq i} \left(\frac{S_i + S_j}{M_{ij}} \right)$$

(3)

where: k is the number of clusters, S_i is the average distance between data points in cluster i and the centroid of cluster i, S_j is the average distance between data points in cluster j and the centroid of cluster j, M_{ij} is the distance between the centroids of cluster i and cluster j.

The resulting DBI value is used as an indicator to assess whether the number of clusters and the clustering results for natural disaster-prone areas are optimal. Therefore, DBI provides an objective basis for evaluating the quality of clustering produced by the K-Means method [11][12].

3. RESULTS AND DISCUSSION

This section presents the research results and discussion based on the stages of the applied research methodology.

3.1 Data Collection

The data collection stage produced natural disaster occurrence data for the year 2024 in North Sumatra Province, obtained from the National Disaster Management Agency (BNPB). The data include the number of disaster event count, fatality count, injury/sick victim count, affected population count, evacuee count, and damage count for each regency/city.

Tabel 1. Shows the initial disaster data

Region Name	Event	Fatality	Injury /Sick	Affected	Evacuee	Damage
Asahan	8	0	0	0	488	0
Batu Bara	3	0	0	60	0	13
Dairi	13	0	0	2288	16	59
Deli Serdang	34	18	37	66923	323	350
Humbang Hasundutan	2	0	0	5	0	6
Karo	33	13	0	482	173	18
Labuhanbatu	22	2	0	158	0	0
Labuhanbatu Selatan	13	0	0	5979	1492	0
Labuhanbatu Utara	24	0	0	23275	2	59
Langkat	24	1	11	28306	12	0
Mandailing Natal	10	0	1	740	220	24

Vol.7, No. 1, Januari 2026 e-ISSN: 2775-2496

https://journal-computing.org/index.php/journal-cisa/index

Nias	0	0	0	0	0	0
Nias Barat	2	0	0	0	0	0
Nias Selatan	0	0	0	0	0	0
Nias Utara	2	0	0	0	0	0
Pakpak Bharat	2	0	0	1095	0	0
Padang Lawas	19	4	1	7653	331	108
Padang Lawas Utara	19	0	0	408	0	0
Samosir	15	2	5	8	0	8
Serdang Bedagai	10	0	0	174	408	20
Simalungun	15	2	1	1373	0	162
Tapanuli Selatan	13	3	87	0	883	412
Tapanuli Tengah	4	0	0	0	0	14
Tapanuli Utara	7	5	24	8	0	38
Toba	9	0	0	20	0	6
Kota Medan	13	0	0	42544	90	48
Kota Binjai	6	0	0	28593	0	6
Kota Gunungsitoli	4	0	0	2455	0	0
Kota Pematangsiantar	4	0	0	180	0	23
Kota Padangsidimpuan	0	0	0	0	0	0
Kota Sibolga	1	0	0	1565	0	0
Kota Tanjungbalai	4	0	0	62322	0	0
Kota Tebing Tinggi	2	0	0	11037	0	0

3.2 Data Preprocessing

After data collection, preprocessing was conducted to prepare the dataset for clustering analysis. This stage included attribute selection, data cleaning, and data normalization using the Min–Max normalization method.

Tabel 2. The normalized dataset

Region Name	**Event**	**Fatality**	**Injury /Sick**	**Affected**	**Evacuee**	**Damage**
Asahan	0.235	0	0	0	0.327	0
Batu Bara	0.088	0	0	0.001	0	0.032
Dairi	0.382	0	0	0.034	0.011	0.143
Deli Serdang	1	1	0.425	1	0.216	0.850
Humbang Hasundutan	0.059	0	0	0	0	0.015
Karo	0.971	0.722	0	0.007	0.116	0.044
Labuhanbatu	0.647	0.111	0	0.002	0	0
Labuhanbatu Selatan	0.382	0	0	0.089	1	0
Labuhanbatu Utara	0.706	0	0	0.348	0.001	0.143

Langkat	0.706	0.056	0.126	0.423	0.008	0
Mandailing Natal	0.294	0	0.011	0.011	0.147	0.058
Nias	0	0	0	0	0	0
Nias Barat	0.059	0	0	0	0	0
Nias Selatan	0	0	0	0	0	0
Nias Utara	0.059	0	0	0	0	0
Pakpak Bharat	0.059	0	0	0.016	0	0
Padang Lawas	0.559	0.222	0.011	0.114	0.222	0.262
Padang Lawas Utara	0.559	0	0	0.006	0	0
Samosir	0.441	0.111	0.057	0	0	0.019
Serdang Bedagai	0.294	0	0	0.003	0.273	0.049
Simalungun	0.441	0.111	0.011	0.021	0	0.393
Tapanuli Selatan	0.382	0.167	1	0	0.592	1
Tapanuli Tengah	0.118	0	0	0	0	0.034
Tapanuli Utara	0.206	0.278	0.276	0	0	0.092
Toba	0.265	0	0	0	0	0.015
Kota Medan	0.382	0	0	0.636	0.060	0.117
Kota Binjai	0.176	0	0	0.427	0	0.015
Kota Gunungsitoli	0.118	0	0	0.037	0	0
Kota Pematangsiantar	0.118	0	0	0.003	0	0.056
Kota Padangsidimpuan	0	0	0	0	0	0
Kota Sibolga	0.029	0	0	0.023	0	0
Kota Tanjungbalai	0.118	0	0	0.931	0	0
Kota Tebing Tinggi	0.059	0	0	0.165	0	0

3.3 Cluster Initialization

The clustering process was performed by defining five clusters (k = 5), which represent the levels of natural disaster priority areas, namely very low priority, low priority, medium priority, high priority, and very high priority.

Tabel 3. Presents the initial centroid values

Centroid	Priority Levels	Region Name	Event	Fatality	Injury /Sick	Affected	Evacuee	Damage
C1	very low priority	Nias	0	0	0	0	0	0
C2	low priority	Asahan	0.235	0	0	0	0.327	0
C3	medium priority	Padang Lawas	0.559	0.222	0.011	0.114	0.222	0.262

C4	high priority	Langkat	0.706	0.056	0.126	0.423	0.008	0
C5	very high priority	Deli Serdang	1	1	0.425	1	0.216	0.850

3.4 K-Means Clustering Algorithm

The K-Means clustering algorithm was applied to group regions in North Sumatra Province based on similarities in disaster characteristics.

1) First Iteration

In the first iteration, Euclidean distances between each regional data point and the initial centroids were calculated. Each region was assigned to the cluster with the minimum distance.

Tabel 4. First Iteration

Region Name	C1	C2	C3	C4	C5	First Iteration
Asahan	0.403	0	0.497	0.722	1.871	C2
Batu Bara	0.094	0.360	0.621	0.761	1.930	C1
Dairi	0.410	0.379	0.382	0.544	1.743	C2
Deli Serdang	1.987	1.871	1.449	1.472	0	C5
Humbang Hasundutan	0.061	0.372	0.650	0.785	1.952	C1
Karo	1.216	1.053	0.700	0.847	1.380	C3
Labuhanbatu	0.657	0.537	0.388	0.447	1.691	C3
Labuhanbatu Selatan	1.074	0.695	0.869	1.104	1.931	C2
Labuhanbatu Utara	0.800	0.685	0.434	0.213	1.496	C4
Langkat	0.834	0.722	0.522	0	1.472	C4
Mandailing Natal	0.335	0.198	0.421	0.615	1.811	C2
Nias	0	0.403	0.702	0.834	1.987	C1
Nias Barat	0	0.372	0.656	0.785	1.958	C1
Nias Selatan	0	0.403	0.702	0.834	1.987	C1
Nias Utara	0.059	0.372	0.656	0.785	1.958	C1
Pakpak Bharat	0.061	0.372	0.653	0.777	1.950	C1
Padang Lawas	0.702	0.497	0.000	0.522	1.449	C3
Padang Lawas Utara	0.559	0.460	0.423	0.463	1.770	C3

Samosir	0.459	0.407	0.387	0.507	1.724	C3
Serdang Bedagai	0.405	0.093	0.425	0.662	1.822	C2
Simalungun	0.602	0.563	0.318	0.635	1.577	C3
Tapanuli Selatan	1.589	1.456	1.306	1.549	1.603	C3
Tapanuli Tengah	0.122	0.349	0.599	0.738	1.916	C1
Tapanuli Utara	0.452	0.519	0.537	0.713	1.672	C1
Toba	0.265	0.329	0.509	0.627	1.861	C1
Kota Medan	0.753	0.714	0.632	0.430	1.502	C4
Kota Binjai	0.462	0.541	0.636	0.547	1.712	C1
Kota Gunungsitoli	0.123	0.350	0.607	0.717	1.912	C1
Kota Pematangsiantar	0.130	0.352	0.590	0.738	1.905	C1
Kota Padangsidimpuan	0.000	0.403	0.702	0.834	1.987	C1
Kota Sibolga	0.038	0.387	0.675	0.798	1.961	C1
Kota Tanjungbalai	0.939	0.994	1.015	0.790	1.653	C4
Kota Tebing Tinggi	0.175	0.407	0.648	0.710	1.880	C1

After the cluster assignment process in the first iteration, new centroid values were calculated by computing the average of all data points within each cluster. These updated centroids represent the new cluster centers.

Tabel 5. New centroid

Centroid	Event	Fatality	Injury/Sick	Affected	Evacuee	Damage
C1	0.088	0.017	0.017	0.042	0	0.016
C2	0.318	0	0.002	0.027	0.352	0.050
C3	0.571	0.206	0.154	0.022	0.133	0.245
C4	0.478	0.014	0.032	0.584	0.017	0.065
C5	1	1	0.425	1	0.216	0.850

2) Second Iteration

The second iteration was performed using the updated centroid values obtained from the first iteration. Euclidean distances between each data point and the new centroids were recalculated, and cluster assignments were updated accordingly.

Tabel 6. Second Iteration

Region Name	C1	C2	C3	C4	C5	Second Iteration
Asahan	0.362	0	0.527	0.708	1.871	C2
Batu Bara	0.050	0.421	0.603	0.704	1.930	C1
Dairi	0.322	0.359	0.357	0.565	1.743	C1
Deli Serdang	1.903	1.802	1.489	1.493	0	C5
Humbang Hasundutan	0.057	0.439	0.633	0.722	1.952	C1
Karo	1.136	1.002	0.700	1.043	1.380	C3
Labuhanbatu	0.569	0.498	0.342	0.618	1.691	C3
Labuhanbatu Selatan	1.044	0.656	0.959	1.107	1.931	C2
Labuhanbatu Utara	0.701	0.620	0.468	0.340	1.496	C4
Langkat	0.735	0.668	0.528	0.305	1.472	C4
Mandailing Natal	0.259	0.207	0.419	0.617	1.811	C2
Nias	0	0.477	0.687	0.759	1.987	C1
Nias Barat	0	0.440	0.638	0.723	1.958	C1
Nias Selatan	0.102	0.477	0.687	0.759	1.987	C1
Nias Utara	0.059	0.440	0.638	0.723	1.958	C1
Pakpak Bharat	0.049	0.440	0.638	0.710	1.950	C1
Padang Lawas	0.615	0.421	0.194	0.593	1.449	C3
Padang Lawas Utara	0.473	0.430	0.380	0.589	1.770	C3
Samosir	0.370	0.395	0.323	0.596	1.724	C3
Serdang Bedagai	0.347	0.085	0.450	0.663	1.822	C2
Simalungun	0.525	0.519	0.293	0.661	1.577	C3
Tapanuli Selatan	1.547	1.410	1.238	1.586	1.603	C3
Tapanuli Tengah	0.060	0.406	0.579	0.688	1.916	C1
Tapanuli Utara	0.395	0.539	0.442	0.739	1.672	C1
Toba	0.183	0.359	0.482	0.625	1.861	C1
Kota Medan	0.673	0.681	0.708	0.132	1.502	C4
Kota Binjai	0.396	0.552	0.677	0.346	1.712	C4
Kota Gunungsitoli	0.042	0.408	0.592	0.660	1.912	C1
Kota Pematangsiantar	0.068	0.405	0.571	0.685	1.905	C1
Kota Padangsidimpuan	0.102	0.477	0.687	0.759	1.987	C1
Kota Sibolga	0.068	0.458	0.662	0.722	1.961	C1
Kota Tanjungbalai	0.890	0.992	1.085	0.506	1.653	C4
Kota Tebing Tinggi	0.130	0.461	0.654	0.598	1.880	C1

Vol.7, No. 1, Januari 2026 e-ISSN: **2775-2496**
https://journal-computing.org/index.php/journal-cisa/index

Based on the clustering results in the second iteration, centroid values were recalculated once again. The updated centroids reflect more stable cluster centers compared to the previous iteration.

Tabel 7. New centroid

Centroid	Event	Fatality	Injury/Sick	Affected	Evacuee	Damage
C1	0.101	0.017	0.017	0.017	0	0.024
C2	0.318	0	0.002	0.027	0.352	0.050
C3	0.571	0.206	0.154	0.022	0.133	0.245
C4	0.418	0.011	0.025	0.553	0.014	0.055
C5	1	1	0.425	1	0.216	0.850

3) Third Iteration

The third iteration was conducted using the centroid values obtained from the second iteration. Distance calculations and cluster assignments were repeated for all regions.

The results of the third iteration show that no changes occurred in cluster membership compared to the second iteration. This indicates that the K-Means clustering process has reached convergence.

Therefore, the cluster assignments obtained in the third iteration were used as the final clustering results.

Tabel 8. Third Iteration

Region Name	C1	C2	C3	C4	C5	Third Iteration
Asahan	0.355	0	0.527	0.664	1.871	C2
Batu Bara	0.033	0.421	0.603	0.644	1.930	C1
Dairi	0.307	0.359	0.357	0.528	1.743	C1
Deli Serdang	1.906	1.802	1.489	1.533	0	C5
Humbang Hasundutan	0.053	0.439	0.633	0.661	1.952	C1
Karo	1.126	1.002	0.700	1.059	1.380	C3
Labuhanbatu	0.555	0.498	0.342	0.608	1.691	C3
Labuhanbatu Selatan	1.041	0.656	0.959	1.092	1.931	C2

Vol.7, No. 1, Januari 2026 e-ISSN: **2775-2496**

https://journal-computing.org/index.php/journal-cisa/index

Labuhanbatu Utara	0.700	0.620	0.468	0.366	1.496	C4
Langkat	0.738	0.668	0.528	0.339	1.472	C4
Mandailing Natal	0.246	0.207	0.419	0.572	1.811	C2
Nias	0	0.477	0.687	0.696	1.987	C1
Nias Barat	0	0.440	0.638	0.662	1.958	C1
Nias Selatan	0.108	0.477	0.687	0.696	1.987	C1
Nias Utara	0.057	0.440	0.638	0.662	1.958	C1
Pakpak Bharat	0.054	0.440	0.638	0.649	1.950	C1
Padang Lawas	0.605	0.421	0.194	0.586	1.449	C3
Padang Lawas Utara	0.459	0.430	0.380	0.568	1.770	C3
Samosir	0.356	0.395	0.323	0.565	1.724	C3
Serdang Bedagai	0.336	0.085	0.450	0.622	1.822	C2
Simalungun	0.511	0.519	0.293	0.639	1.577	C3
Tapanuli Selatan	1.539	1.410	1.238	1.584	1.603	C3
Tapanuli Tengah	0.036	0.406	0.579	0.630	1.916	C1
Tapanuli Utara	0.388	0.539	0.442	0.697	1.672	C1
Toba	0.167	0.359	0.482	0.576	1.861	C1
Kota Medan	0.688	0.681	0.708	0.122	1.502	C4
Kota Binjai	0.417	0.552	0.677	0.277	1.712	C4
Kota Gunungsitoli	0.043	0.408	0.592	0.600	1.912	C1
Kota Pematangsiantar	0.046	0.405	0.571	0.628	1.905	C1
Kota Padangsidimpuan	0.108	0.477	0.687	0.696	1.987	C1
Kota Sibolga	0.080	0.458	0.662	0.660	1.961	C1
Kota Tanjungbalai	0.915	0.992	1.085	0.487	1.653	C4
Kota Tebing Tinggi	0.157	0.461	0.654	0.532	1.880	C1

Results of Regional Clustering Using the K-Means Method

1. Cluster C1 (Very Low Priority)
   Regions included in cluster C1 consist of Asahan, Batu Bara, Dairi, Humbang Hasundutan, Mandailing Natal, Nias, West Nias, South Nias, North Nias, Pakpak Bharat, North Padang Lawas, Central Tapanuli, North Tapanuli, Toba, Gunungsitoli City, Pematangsiantar City, Padangsidimpuan City, Sibolga City, and Tebing Tinggi City. This cluster is dominated by regions with relatively low levels of disaster occurrence and impact.
2. Cluster C2 (Low Priority)

Vol.7, No. 1, Januari 2026 e-ISSN: 2775-2496

https://journal-computing.org/index.php/journal-cisa/index

Regions included in cluster C2 are South Labuhanbatu, Langkat, Padang Lawas, Serdang Bedagai, and Samosir. These regions exhibit higher disaster occurrence and impact compared to cluster C1, but they are still categorized as low priority.

3. Cluster C3 (Medium Priority)
   Cluster C3 consists of Karo, Labuhanbatu, North Padang Lawas, Simalungun, and South Tapanuli. This cluster represents regions with moderate levels of disaster occurrence and impact.
4. Cluster C4 (High Priority)
   Regions classified in cluster C4 include North Labuhanbatu, Medan City, Binjai City, and Tanjungbalai City. These regions experience relatively high disaster impacts and therefore require greater attention in disaster mitigation planning.
5. Cluster C5 (Very High Priority)
   Cluster C5 consists solely of Deli Serdang Regency. This cluster represents the region with the highest disaster occurrence and impact among all regions, making it the top priority for disaster mitigation and management efforts in North Sumatra Province.

3.5 Davies–Bouldin Index (DBI)

The quality of the clustering results was evaluated using the Davies–Bouldin Index (DBI). The DBI calculation process began by determining the cluster width, defined as the average distance of data points within a cluster to its centroid.

Tabel 9. Presents the calculated width values for each cluster

Cluster	C1	C2	C3	C4	C5
Distance (S)	0.133	0.337	0.496	0.318	0

Next, the distances between centroids of all cluster pairs were calculated to measure the degree of separation between clusters.

Tabel 10. Shows the distances between cluster centroids

D	1	2	3	4	5
1	0	0.551	0.560	0.606	1.894
2	0.551	0	1.894	1.894	1.852
3	0.560	1.894	0	0.643	1.492

Vol.7, No. 1, Januari 2026 e-ISSN: 2775-2496
https://journal-computing.org/index.php/journal-cisa/index

4	0.606	1.894	0.643	0	1.540
5	1.894	1.852	1.492	1.540	0

After that, the ratio between the sum of cluster widths and the distance between centroids was computed for each pair of clusters. For each cluster, the maximum ratio value was selected.

Tabel 11. Presents the ratio values for all cluster pairs

R	1	2	3	4	5	Max
1	0	0.854	1.123	0.744	0.070	R13
2	0.854	0	1.396	0.876	0.182	R23
3	1.123	1.396	0	1.266	0.332	R32
4	0.744	0.876	1.266	0	0.207	R43
5	0.070	0.182	0.332	0.207	0	R53

The Davies–Bouldin Index value was then obtained by calculating the average of the maximum ratio values across all clusters. Based on the calculation, the resulting DBI value is 1.102

The obtained DBI value indicates that the clusters formed by the K-Means algorithm have a reasonable level of compactness and separation. This result suggests that the clustering outcome using the K-Means method with five clusters (k = 5) is acceptable and relevant for determining natural disaster priority areas in North Sumatra Province.

4. CONCLUSION

This study applied the K-Means clustering method to identify natural disaster priority areas in North Sumatra Province based on disaster occurrence and impact characteristics. Using five clusters (k = 5), regions were grouped into very low, low, medium, high, and very high disaster priority levels. The results indicate that most regions fall into the very low and low priority clusters, while Deli Serdang Regency is classified as a very high priority area, requiring special attention in disaster mitigation planning.

The clustering evaluation using the Davies–Bouldin Index (DBI) produced a value of 1.102688, indicating that the formed clusters have reasonable

Vol.7, No. 1, Januari 2026 e-ISSN: **2775-2496**
https://journal-computing.org/index.php/journal-cisa/index

compactness and separation. Overall, the findings demonstrate that the K-Means method is effective for supporting data-driven decision-making in disaster risk mitigation and regional planning in North Sumatra Province.

REFERENCES

[1] National Disaster Management Agency (BNPB), *Indonesia Disaster Risk Assessment*, Jakarta: BNPB, 2021.

[2] A. R. Nugroho and S. Hadi, "Disaster-Prone Area Mapping Using Geographic Information Systems," *Journal of Geography*, vol. 12, no. 2, pp. 85–94, 2019.

[3] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. San Francisco: Morgan Kaufmann, 2012.

[4] R. Kurniawan, Y. Prasetyo, and D. Suryani, "Application of K-Means Method for Flood-Prone Area Clustering," *Journal of Information Technology*, vol. 10, no. 1, pp. 45–53, 2020.

[5] S. Wibowo and M. Lestari, "Clustering of Landslide-Prone Areas Using K-Means Method," *Journal of Informatics*, vol. 8, no. 3, pp. 210–218, 2021.

[6] National Disaster Management Agency (BNPB), *Natural Disaster Occurrence Data of North Sumatra Province 2024*, BNPB One Data Portal, 2024.

[7] Statistics Indonesia (BPS) of North Sumatra Province, *North Sumatra in Figures*, Medan: BPS, 2024.

[8] C. C. Aggarwal, *Data Mining: The Textbook*, Cham: Springer, 2015.

[9] J. Han, M. Kamber, and J. Pei, "Data Preprocessing and Normalization Techniques," in *Data Mining: Concepts and Techniques*, 3rd ed. San Francisco: Morgan Kaufmann, 2012.

[10] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, 1979.

[11] P. J. Rousseeuw, "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.