K-Means Algorithm's Implementation to Facilitate Grouping of Landslide-Prone Areas

Maria Ulfah[1]*, Andi Sri Irtawaty[2], Zulkifli[3], Subur Mulyanto[4]

[1,2] *Electrical Engineering Departement, Politeknik Negeri Balikpapan, Indonesia*
[3,4] *Mechanical Engineering Departement, Politeknik Negeri Balikpapan, Indonesia*

**Corresponding author. Email: maria.ulfah@poltekba.ac.id*

ABSTRACT

Indonesia is one of the countries that has hydrometeorological disaster vulnerability, one of which is Balikpapan City. Natural disasters have the potential to damage the environment, harm property, and cause casualties. Landslide disaster mitigation is still not optimal and there is no grouping of landslide-prone areas in Balikpapan. Therefore, it is necessary to research to produce a grouping of landslide vulnerability levels in 34 sub-districts. To group vulnerable areas, the clustering method with the K-Means algorithm was used from 34 data on the impact of landslides from each sub-district in Balikpapan with Rapid Miner tool. Clusters or groupings formed: not vulnerable, vulnerable, and very vulnerable. The very vulnerable cluster contains one sub-district, namely Sepinggan, the vulnerable cluster consists of 4 sub-districts which are Sungai Nangka, Baru Tengah, Manggar, and Baru Ilir. The non-vulnerable cluster consists of the rest of the 29 sub-districts with a David Bouldin performance value of 0.009

.

Keywords: *K-Means Algorithm, Landslide-Prone Areas, Data Mining, Clustering Area*

1. INTRODUCTION

Landslides are frequent disasters in Indonesia, especially in hilly and mountainous areas. However, Indonesia is a place with a tropical climate, and it has two seasons: hot and rainy. Because of this, the weather, temperature, and wind direction can change a lot. On the other hand, Indonesia is located in a tropical climate area with two seasons namely hot and rainy which causes changes in weather, temperature, and wind direction to be quite extreme. Such climatic conditions coupled with relatively diverse surface and rock topography, both physically and chemically, result in unstable soil conditions. This can lead to many bad things happening to people like floods, landslides, and droughts. This is mentioned in [1] information from BNPD and [2] research. Furthermore, when people do more things that harm the environment, the damage gets worse and causes more and stronger natural disasters related to water and weather (like floods, landslides, and droughts) to happen in several places in Indonesia consecutively [1], [2].

For two consecutive years, 2020 and 2021 natural disasters, landslides are ranked number three, below floods and tornadoes, which are the most common in the Indonesian region with an incident frequency of 1054 in 2020 and 1321 in 2021 [3]. Natural disasters that occur, including landslides that occur, result in various adverse impacts, both loss of life and damage, which are increasing every year.

The topographical condition of the city of Balikpapan, which is dominated by hills with a large slope value, is supported by unstable soil types which results in frequent landslides in the city of Balikpapan. This was reinforced throughout 2021-2022 Landslides are ranked first in the number one place that occurs in Balikpapan compared to floods and residential fires, namely 59 times in 2021, 54 times landslides occur throughout 2022 . This of course results in loss of both lives and houses and facilities in the City of Balikpapan. One of the right steps to reduce disaster risk in Balikpapan City is do grouping areas according to their level of vulnerability to landslides. The use of the

U. Verawardina and A. Mubai (eds.), *Proceedings of the 10th International Conference on Technical and Vocational Education and Training (ICTVET 2023)*,
Advances in Social Science, Education and Humanities Research 836,
https://doi.org/10.2991/978-2-38476-232-3_11

clustering has made it easier to respond to challenges in disaster risk management and reduction. For example, it took a long time for aid to arrive due to limited equipment and food in the disaster area. Therefore it is necessary to map landslide-prone areas by classifying the level of the number of events or events in that year.

Related research describes research reviews that have been conducted previously by other researchers that are relevant to the research being conducted, including: [4] From this study, grouping of landslide-prone areas for Central Java was carried out using the k-means algorithm, it is concluded that the K-Means method can determine the clustering of landslide-prone areas based on the number of events and can find out which areas are included very group vulnerable, prone and not vulnerable. Cluster 0 = 22 regions, cluster 1 = 1 region, and cluster 2 = 4 regions. The difference between the research carried out by the author and previous research is that the dataset used [4] research using the number of landslide disaster events, while the dataset used by the author is the value of the impact of landslides.

This study do grouping of areas prone to natural disasters for West Java was carried out using the k-means algorithm [5]. It was found that the disaster-prone areas of West Java were grouped into 3 low, medium, and high clusters, From the results of K-Means Clustering calculations using the RapidMiner application, it produces low level clusters (C0), including the areas of Cianjur, Cirebon, Indramayu, Subang, Purwakarta, Karawang, Bekasi, Pangandaran, Sukabumi City, Bandung City, Cirebon City, Bekasi City, Kota Depok, Cimahi City, Tasikmalaya City, and Banjar City. Meanwhile, the medium cluster (C1) is in Bogor and Sukabumi. Furthermore, the high/prone cluster (C2) is in the areas of Bandung, Garut, Tasikmalaya, Ciamis, Kuningan, Majalengka, Sumedang, West Bandung and Bogor City. The difference between the research carried out by the author and previous research is that the dataset used is that the research uses the number of events of various natural disasters such as landslides, earthquakes, floods, tornadoes and fires, while the dataset used by the author is the value of the impact of landslides.

In this study, Village Grouping Using K-Means for Implementation of Flood Disaster Management [6]. The K-Means method was chosen to categorize villages based on their traits. It determined the ideal number of groups by comparing two methods called the elbow and silhouette methods. The best groupings obtained from the elbow method determined that there are 7 clusters. The grouping size is the number of villages in each cluster. Cluster 1 has 1 village, Cluster 2 has 2 villages, Cluster 5 has 5 villages, Cluster 6 has 1 village, and Cluster 7 has 1 village. The difference between the research carried out by the author and previous research is that the dataset used research using the number of flood disaster events, while the dataset used by the author is the value of the impact of landslides.

The objective of this study is to reveal areas prone to natural disasters in West Sumatra province using K-Means algorithm and GeoPandas visualization [7]. After doing data processing for the optimal K value, 3 clusters were obtained, namely, cluster 1 is not prone to natural disasters, cluster 2 is prone to natural disasters and cluster 3 is very prone to natural disasters. From this research, it is found that cluster 1 has 7 members of the region, cluster 2 has 11 members of the region, and cluster 3 has 1 member of the region.

Researchers in this study use the K-Means Clustering method to classify natural disaster areas based on their level of vulnerability to natural disasters [8]. Then a search for the optimal K value was carried out using the elbow method and Silhouette Analysis which produced 2 clusters namely cluster 0 and cluster 1. The results of the clustering indicated that cluster 1 was classified as a disaster-prone area in Indonesia including Aceh, North Sumatra, West Java, Central Java, East Java.

The purpose of this study is to Cluster landslide-prone areas, to find out the best centroid points so that they can cluster landslide-prone areas into how many clusters. So that this research can be a reference in efforts to deal with landslides and as a basis for the government to provide appropriate policies to disaster management.

2. RESEARCH METHOD

The research method used is to conduct a literature study and collect data using data from the Balikpapan City BPBD Strategic Plan, and related scientific articles about 6 articles that come from valid resources such us google scholar after doing data preprocessing, namely cleaning data, the calculations will be processed with Rapidminer Tools using the K-Means Algorithm method. This is how to do research:

Figure 1. Research Methods

1. Planning Phase

The first thing researchers need to do in this study is to make a research plan. There are four things to do in planning: figuring out what the researcher wants to learn, finding the issues, deciding what limits the problem has, and studying existing information on the topic.

2. Preprocessing Stage

In the preprocessing step, the researcher typed the obtained data into Microsoft Excel to keep track of all book lending transactions. Once the data is copied, the researcher goes through it to remove any unclear or unreadable information. This cleaning process is done to make sure the calculations are accurate.

3. Data Processing Phase

Using RapidMiner software to analyze data and identify issues in the research. This text is saying that by using the K-Means method, researchers can group data based on the impact of landslides. This method will give us good results when processing the data. Learning outcomes are the goals or objectives that someone hopes to achieve through their learning. Generating new knowledge means creating or discovering new information or ideas. Using Rapidminer software to work with and analyze data.

4. Analysis Stage

After gathering all the information, the next step is to analyze it. Here's the research flow chart:

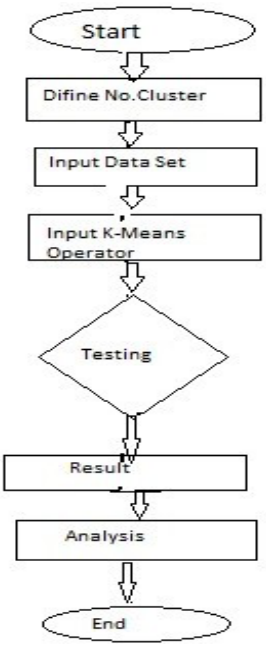

Figure 2. Flowchart research

In this research, author need to difine number of cluster, which takes 3 cluster. Then import dataset into the Rapid miner tools, choose K-means algorithm, run model after that an analysis the result

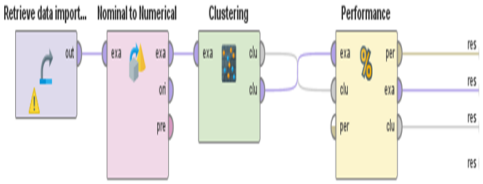

Figure 3. Clustering Modeling with the K-Means operator

3. RESULT AND DISSCUSION

1. Data Sets

This study uses data from the 2021-2026 BPBD Strategic Plan for the city of Balikpapan regarding the impact of the landslide disaster, which is a comparison of the area affected to the number of affected areas. A total of 34 data on the impact of landslides in each village. This data will be processed to find out the results of classifying landslide-prone areas in Balikpapan City. This data grouping can become new information for the BPDB City of Balikpapan and to decide on a policy going forward disaster risk management and reduction. This is in accordance with the function of BNPB, namely formulating and establishing disaster management policies and handling refugees by acting quickly and precisely, effectively and efficiently; and coordinating the implementation of disaster management activities in a planned, integrated and comprehensive manner.[9]

2. Preposing Data

The attributes used are 2 (two), namely the name of the sub-district and the value of the impact of landslide disaster risk describes the impact of the landslide disaster that occurred. The sub-district attribute is needed so that all sub-districts in Balikpapan are included in the data, the value of the impact of landslide disaster risk attribute is needed to determine the impact value on each sub-district. The following is the data on the impact.

Table 1. Data Set

No	District	Impact
1	Damai Bahagia	3
2	Damai Baru	2
3	Gunung Bahagia	0
4	Sepinggan	132
5	Sepinggan Baru	1
6	Sepinggan Raya	1
7	Sungai Nangka	33
8	Lamaru	2
9	Manggar	18
10	Manggar Baru	0
11	Teritip	0
12	Gunung Sari Ilir	10
13	Gunung Sari Ulu	4
14	Mekar Sari	5
15	Karang Rejo	10
16	Sumber Rejo	5
17	Karang Jati	2
18	Prapatan	8
19	Telaga Sari	4
20	Klandasan Ulu	0
21	Klandasan Ilir	8
22	Damai	5
23	Baru Ilir	17
24	Margo Mulyo	4
25	Marga Sari	0
26	Baru Tengah	20
27	Baru Ulu	12
28	Kariangau	0
29	Gunung Samarinda	5
30	Muara Rapak	2
31	Batu Ampar	2
32	Karang Joang	0
33	Gunung Samarinda Baru	2
34	Graha Indah	1

3. Processing Data

After the data has been preprocessed, a clusterization process is carried out using the K-means algorithm in the Rapid Miner application. The following are the steps in processing the data:

- New Prosess

  The New Process view is about making a worksheet on RapidMiner. Bring the data that needs to be tested into a . xls or xlsx file format with a right click and then choose the option "Insert Operator". After that, select "Data Access" and then "File". Finally, click on "Read Excel" like it is shown in figure 4

Figure 4. Import Data Set

- Now, include the K-Means operator. To perform K-Means segmentation, follow these steps: 1. Right-click on the object. 2. Select the "Insert Operator" option. 3. Choose "Modeling" from the menu. 4. Click on "Segmentation". 5. Finally, select "K-Means".

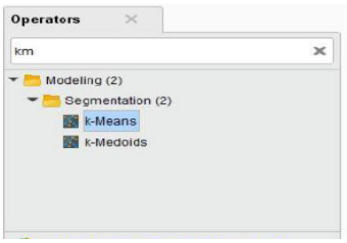

Figure 5. Addition of K-Means Operator

4. Result

In Figure 6, the data results are divided into 3 groups. These are called cluster 0, cluster 1, and cluster 2. Each group contains a different number of sub-districts. Cluster 0 has 29 sub-districts, cluster 1 has 1 sub-district, and cluster 2 has 4 sub-districts.

Cluster 0 is the result of a grouping of sub-districts in Balikpapan that are not prone or non-vulnerable to landslides.

Cluster 1 is the result of a grouping of sub-districts in Balikpapan that very-vulnerable to landslides.

Cluster 2 is the result of a grouping of sub-districts in Balikpapan that vulnerable to landslides.

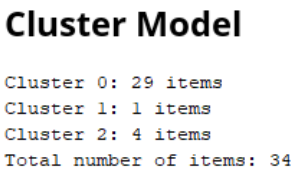

Figure 6. Result

Data processing using the K-Means algorithm with the Rapid Miner tool. Clusters/groupings formed: not vulnerable, vulnerable, and very vulnerable. The very vulnerable cluster contains one sub-district, namely Sepinggan, the vulnerable cluster consists of 4 sub-districts (Sungai Nangka, Manggar, Baru Ilir, Baru Ulu), the non-vulnerable cluster consists of 29 sub-district

The Davies Bouldin Index helps to evaluate how well a cluster is performing. It looks at both the amount and closeness of the data in the cluster to determine whether the results are good or bad.

Figure 7. Performance Result

From the results of the performance, the Davies Bouldin index value was obtained at 0.009 as figured 7. The Davies Bouldin Index is measure to evaluate clustering performance[10]. The value 0.009 is the performance of regional grouping based on the level of vulnerability to regional landslides.

4. CONCLUSIONS

From analyzing the data using the Rapid Miner application with 3 clusters, we obtained the following results., not vulnerable, vulnerable, and very vulnerable area in sub-district Balikpapan form landslide. The very vulnerable cluster contains one sub-district, namely Sepinggan, the vulnerable cluster consists of 4 sub-districts which are Sungai Nangka, Baru Tengah, Manggar, and Baru Ilir. The non-vulnerable cluster consists of the rest of the 29 sub-districts with a David Bouldin performance value of 0.009. The results of this research can be used by the government or related agencies BPBD Balikpapan city in disaster management and disaster prevention

REFERENCES

[1] I. PRIYADI, F. HADI, Y. S. PRANATA, and M. R. RAZALI, “Rancangan dan Implementasi Sistem Deteksi Longsor Berbasis SMS dan Progressive Web Apps,” *ELKOMIKA J. Tek. Energi Elektr. Tek. Telekomun. Tek. Elektron.*, vol. 10, no. 1, p. 243, 2022, doi: 10.26760/elkomika.v10i1.243.

[2] BNPB, “Potensi Ancaman Bencana,” *Badan nasional penanggulangan bencana*, 2020. .

[3] BNPB, “Bencana Indonesia 2021,” *Badan nasional penanggulangan bencana*, 2021. .

[4] N. Fadilah, “Penerapan Metode Algoritma K-Means Untuk Clustering Daerah Rawan Tanah Longsor Di Provinsi Jawa Tengah,” *J. BATIRSI*, vol. 6, no. 1, pp. 1–5, 2022.

[5] M. Firman, A. Halik, and L. Septiana, “Analisa Data Untuk Prediksi Daerah Rawan Bencana Alam Di Jawa Barat Menggunakan Algoritma K-Means Clustering,” *J. Inf. Syst. Applied, Manag. Account. Res.*, vol. 6, no. 4, pp. 856–870, 2022, doi: 10.52362/jisamar.v6i4.939.

[6] S. F. Susilo, A. Jamaludin, and I. Purnamasari, “Pengelompokan Desa Menggunakan K-Means Untuk Penyelenggaraan Penanggulangan Bencana Banjir,” *JOINS (Journal Inf. Syst.*, vol. 5, no. 2, pp. 156–167, 2020, doi: 10.33633/joins.v5i2.3709.

[7] D. A. Wicaksono and Y. A. Susetyo, “Clustering Zonasi Daerah Rawan Bencana Alam Di Provinsi Sumatera Barat Menggunakan Algoritma K-Means Dan Library Geopandas,” *J. Indones. Manaj. Inform. dan Komun.*, vol. 4, no. 2, pp. 426–438, 2023, doi: 10.35870/jimik.v4i2.225.

[8] A. Baldah, V. Duarisah, and R. A. Maulana, “Clustering Daerah Rawan Bencana Alam Di Indonesia Berdasarkan Provinsi Dengan Metode K-Means,” *J. Ilm. Inform. Glob.*, vol. 14, no. 2, pp. 31–36, 2023.

[9] BNPB, “Tugas Dan Fungsi BNPB.” .

[10] Y. Sopyan, A. D. Lesmana, and C. Juliane, “Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1464–1470, 2022, doi: 10.47065/bits.v4i3.2697.