K-Means Algorithm with Davies Bouldin Criteria for Clustering Provinces in Indonesia Based on Number of Events and Impacts of Natural Disasters

By Yuli Asriningtias

International Journal of Engineering, Technology and Natural Sciences
E-ISSN: 2685-3191 | P-ISSN: 2775-7706
Vol 4 No 1 (2022)

K-Means Algorithm with Davies Bouldin Criteria for Clustering Provinces in Indonesia Based on Number of Events and Impacts of Natural Disasters

Yuli Asriningtias[*, a,1], **Joko Aryanto** [b,2]
[a,b] Universitas Teknologi Yogyakarta, Yogyakarta, Indonesia
[1] yuli_asriningtias@uty.ac.id,
[2] joko.aryanto@uty.ac.id

Abstract

Indonesia has 34 Provinces with a geographical position at the Eurasian, Indo-Australian, Pacific, and Philippine slab meeting zone. It makes Indonesia vulnerable to the threat of geological disasters. Other disaster threats arise due to climate change and people's behavior towards the environment, which impacts natural and environmental damage. Based on data on natural disasters and their impacts over the last five years, this study discovers Indonesia province clusters that fall into disaster-prone criteria, the number of disaster victims, and the impact on building damage. This research relays on *rapidminer* tools with the K-Means Clustering Algorithm with the Davies–Bouldin Index (DBI). The procedures of this research are collecting datasets, preprocessing data, and modeling and analyzing DBI. This research results show that the clusters of disaster-prone in Indonesia are the provinces of East Java, Central Java, and West Java. Many disaster victims are in the provinces of Lampung and West Nusa Tenggara; meanwhile, the biggest impact of damaged buildings is in the West Nusa Tenggara Province.

Keywords: Natural disasters, K-Means Clustering, Davies-Bouldin Index, Dataset

I. INTRODUCTION

Indonesia's geographical position at the confluence zone of the Eurasian, Indo-Australian, Pacific, and Philippine slabs makes this country vulnerable to the threat of earthquakes, tsunamis, volcanic eruptions, and landslides. In addition to the threat of geological disasters, Indonesia also faces threats due to climate change, such as floods, droughts, and forest and land fires, with increasing intensity and frequency. Population growth and increasing demand for space and land have led to increased environmental damage due to uncontrolled land use changes, illegal logging, and unplanned urbanization. In the end, these things cause excessive pressure on the environment, further encouraging environmental damage [1]. Natural disasters are the disasters caused by factors that occur in nature, including geological, hydrological, meteorological, climatological, biological factors, and factors caused by objects in outer space [2].

Types of disasters are divided into 12, namely: floods, landslides, floods and landslides, abrasion, tornadoes, droughts, forest and land fires, earthquakes, tsunamis, earthquakes and tsunamis, volcanic eruptions, and others. The impact of the disaster is the victim died, were injured, and missing. Another impact is damage to houses, educational facilities, health facilities, houses of worship, public facilities, offices, bridges, factories, and kiosks [3]. The K-Means method is a simple method for dividing a collection or data set in a specific number from a cluster, namely the value [4]. K-means clustering is the most frequently used method in unsupervised learning to partition the analyzed dataset into groups, representing the number of clusters determined before clustering analysis [5]. K-means is also a clustering algorithm that divides objects into several clusters [6].

Based on the explanation above, this study aims to cluster Indonesia's natural disasters, which cover disaster-prone criteria, the number of disaster victims, and the impact on building damage. The research dataset is taken from: https://dibi.bnpb.go.id/kwilayah to seek the number of disaster events, victims, and the impact of the damage caused using the K-Means Clustering algorithm and the Criteria Performance of Davies-Bouldin Index (DBI).

II. METHOD

The method used consists of three stages, namely: collecting datasets, preprocessing data, and modeling and analyzing data.

A. Collect Dataset

The dataset used in this study is disaster event data, victim data, and damage data from 34 provinces in Indonesia from January 1, 2018 – May 11, 2022. The dataset is presented in Figure 1.

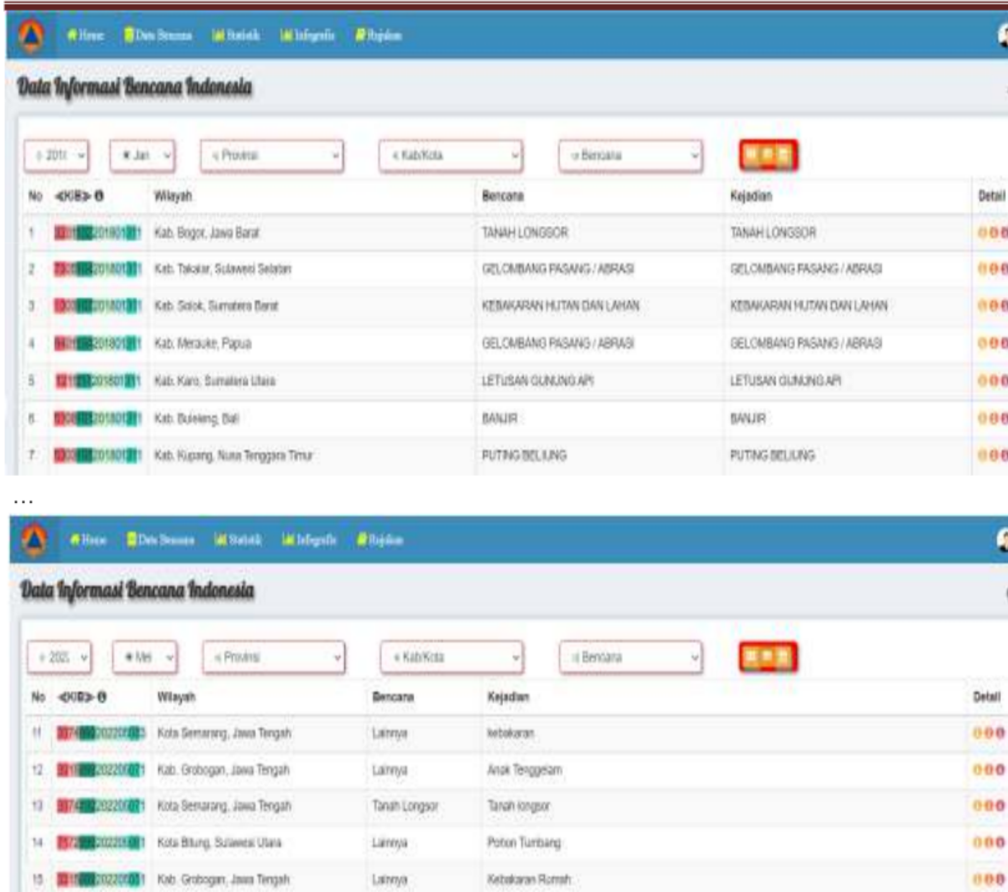

Figure 1. Dataset

B. Preprocessing Data

The next stage is to prepare the data for processing. Double data cleaning is carried out at this stage, setting the data display and changing the Region attribute to ID and other attribute data types to integers. The results of preprocessing are presented in Figure 2, Figure 3, and Figure 4.

1) Disaster Event Data

Row No.	WILAYAH	2022	2021	2020	2019	2018
1	ACEH	0	133	381	188	220
2	SUMATERA UTARA	5	225	150	71	84
3	SUMATERA BARAT	2	53	246	103	99
4	RIAU	0	67	55	58	54
5	JAMBI	0	2	104	28	28
6	SUMATERA SELATAN	34	58	91	95	77
7	BENGKULU	0	10	38	25	11
...						
33	SULAWESI BARAT	3	10	11	11	13
31	MALUKU	3	22	27	39	4
32	MALUKU UTARA	3	13	22	19	23
33	PAPUA BARAT	3	1	9	10	7
34	PAPUA	3	10	9	16	14

Figure 2. Disaster Event Data Results

2) Disaster Victim Data

Row No.	Wilayah	Meninggal	Hilang	Terluka
1	ACEH	20	7	27
2	SUMATERA UTARA	103	32	146
3	SUMATERA BARAT	54	8	137
4	RIAU	4	0	12
5	JAMBI	6	7	3
6	SUMATERA SELATAN	22	1	22
7	BENGKULU	41	4	8
…				
30	SULAWESI BARAT	112	3	11138
31	MALUKU	45	0	1633
32	MALUKU UTARA	16	0	136
33	PAPUA BARAT	10	0	7
34	PAPUA	116	88	1130

Figure 3. Disaster Victim Data Results

3) Disaster Damage Data

Row No.	Wilayah	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1	ACEH	2305	52	4	26	3	11	30	0	128
2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
3	SUMATERA BARAT	[illegible]	56	8	57	3	17	26	0	521
4	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5	JAMBI	2394	[illegible]	2	9	3	0	5	0	1
6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
7	BENGKULU	2500	36	8	22	3	5	[illegible]	0	0
…										
30	SULAWESI BARAT	16737	226	40	111	20	[illegible]	15	0	5
31	MALUKU	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
32	MALUKU UTARA	4732	[illegible]	27	46	3	28	31	0	0
33	PAPUA BARAT	175	3	[illegible]	2	0	3	5	0	7
34	PAPUA	[illegible]	[illegible]	[illegible]	34	[illegible]	[illegible]	[illegible]	0	112

Figure 4. Disaster Damage Data Results

C. Modeling and Analyzing

The last phase is modeling using rapidminer tools using K-Means and the Davies Bouldin Index (DBI) performance criteria. In this modeling, we use three groups of clusters, namely k=2, k=3, and k=4, then observe the DBI performance value; the smallest DBI value shows the most optimal results.

1) Disaster Event Data. The modeling for disaster event data is shown in Figure 5.

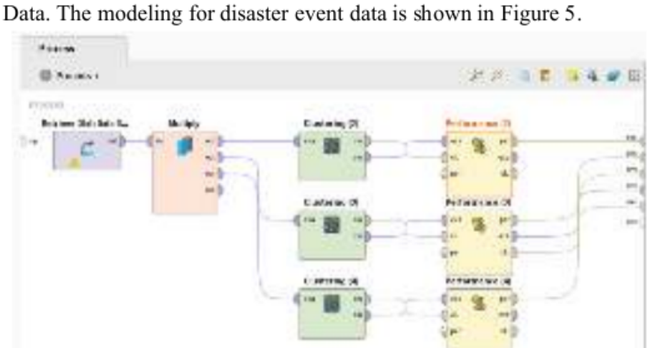

Figure 5. Modeling Disaster Event Data

Based on the modeling, the smallest DBI value is k=3 (shown in table 1).

Table 1. DBI Value

k	DBI Value
2	[illegible]
3	[illegible]
4	[illegible]

2) Disaster Victim Data. The modeling for disaster victim data is presented in Figure 6.

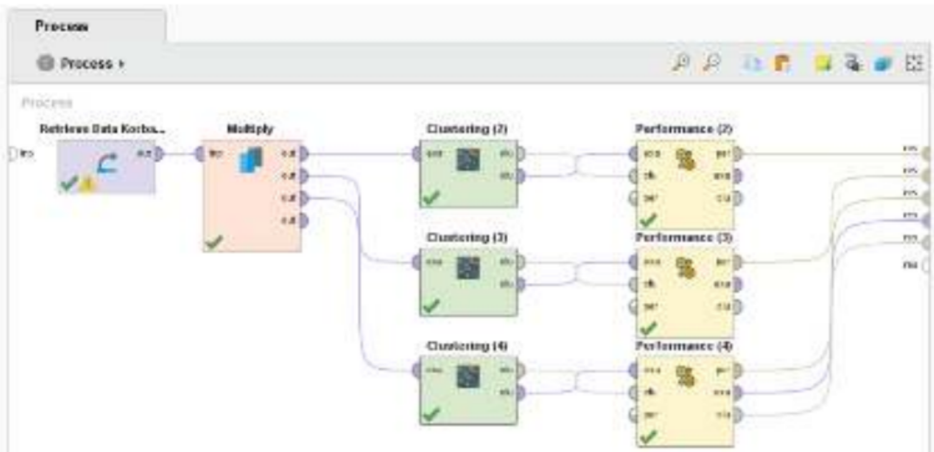

Figure 6. Modeling Disaster Victim Data

For modeling disaster victim data, the smallest DBI value is k=4 (shown in table 2).

Table 2. DBI Value

k	DBI Value
2	0.549
3	0.449
4	0.357

3) Disaster Damage Data. The modeling for disaster damage data is presented in figure 7.

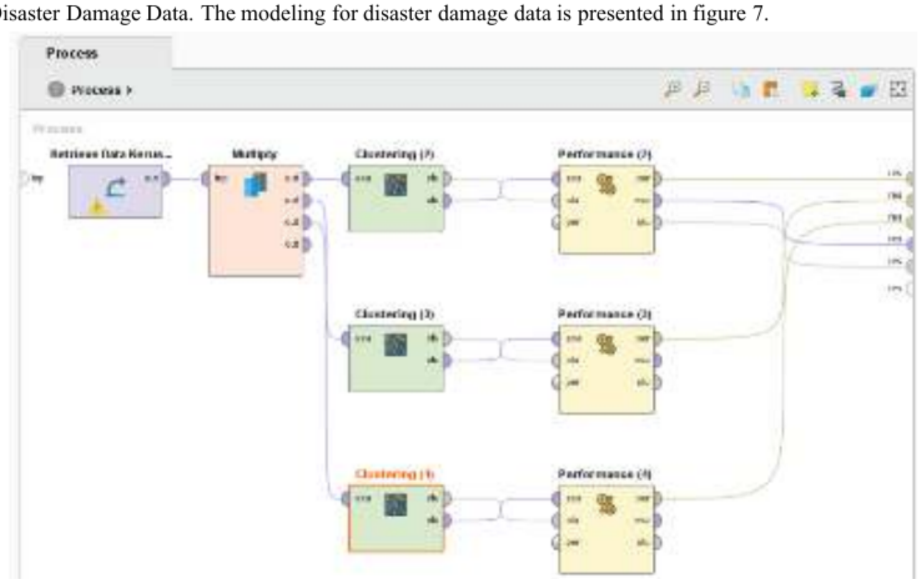

Figure 7. Modeling Disaster Damage Data

The smallest DBI value is k=2 (shown in table 3).

Table 3. DBI Value

k	DBI Value
2	[illegible]
3	[illegible]
4	[illegible]

III. RESULT AND DISCUSSION

Based on the smallest DBI value in the modeling process of the three objects studied, the resulting provincial data grouping is based on disaster events, victims, and the impact of the damage caused. The following are the results of the clustering:

1) Disaster Event. The optimal clustering produced are three clusters, clusters of 0 = 31 Provinces, cluster 1 = 2 Provinces, cluster 2 = 1 Provinces. Provinces based on the clusters formed are presented in table 4.

Table 4. Disaster Event Clustering

Province	Cluster
Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, Dki Jakarta, Di Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua,	cluster 0
Jawa Barat, Jawa Tengah	cluster 1
Jawa Timur	cluster 2

2) Disaster Victim. The resulting clustering consists of four cluster, cluster of 0 = 29 Provinces, cluster 1 = 2 Provinces, cluster 2 = 1 Provinces and cluster 3 = 1 Provinces.
Provinces based on the clusters formed are presented in table 5.

Table 5. Disaster Victim Clustering

Province	Cluster
Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Kepulauan Bangka Belitung, Kepulauan Riau, Dki Jakarta, Jawa Barat, Jawa Tengah, Di Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara, Papua Barat, Papua	cluster_0
Banten, Sulawesi Barat	cluster_1
Sulawesi Tengah	cluster_2
Lampung, Nusa Tenggara Barat	cluster_3

3) Damage. The resulting clustering consists of two clusters, cluster of 0 = 33 Provinces and cluster 1 = 1 Province.
Provinces based on the clusters formed are presented in table 6.

Table 6. Disaster Damage Clustering

Province	Cluster
Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, Dki Jakarta, Jawa Barat, Jawa Tengah, Di Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua	cluster 0
Nusa Tenggara Barat	cluster_1

IV. CONCLUSION

Based on the results of modeling and analysis of the DBI value on the optimal number of Clusters on disaster, victim, and damage event data, it can be concluded that:

1. Provinces included in Cluster two, namely East Java, have a high potential for natural disasters. The next things to watch out for are West Java and Central Java. Other provinces are included in the low – moderate category of natural disasters.
2. Provinces that have the biggest impact on disaster victims are Lampung and West Nusa Tenggara. Next, Central Sulawesi, Banten, West Sulawesi. In addition, they are included in the low-medium group.
3. Natural disasters impacting damage such as houses, education facilities, health facilities, houses of worship, public facilities, offices, bridges, factories, and kiosks are mostly Cluster one, namely West Nusa Tenggara Province.

Finally, the results of this study can be used as a reference for the government to conduct disaster response mapping so that the community's preparedness for disaster events is better to reduce the number of victims and the impact of damage caused by natural disasters.

V. REFERENCES

[1]. Indonesia,“ Rencana Induk Penanggulangan Bencana 2015 – 2045”, 2018.
[2]. Chaudhary, M. T., & Piracha, A, “Natural Disasters — Origins Impacts," Management. 1101–1131, 2021.
[3]. BNPB, B. N. P. B. , “Data Informasi Bencana Alam. Pusat Data Informasi dan Komunikasi Kebencanaan blob:file:///f45ce3e4-dab9-4307-b537-048beacbe5be(Pusdatinkom)”. https://dibi.bnpb.go.id/. 2021.
[4]. Larose, D. T., & Larose, C. D., "Discovering Knowledge in Data: An Introduction to Data Mining," Second Edition. (Vol. 9780470908747), 2014.
[5]. Kassambara, A., "Practical Guide To Cluster Analysis in R," http://www.sthda.com/sthda/ebooks/clustering_english_edition1_preview.pdf, 2017.
[6]. Javed Mehedi Shamrat, F. M., Tasnim, Z., Mahmud, I., Jahan, N., & Nobel, N. I, "Application of k-means clustering algorithm to determine the density of demand of different kinds of jobs," International Journal of Scientific and Technology Research, 9(2), 2550–2557. 2020.

K-Means Algorithm with Davies Bouldin Criteria for Clustering Provinces in Indonesia Based on Number of Events and Impacts of Natural Disasters

ORIGINALITY REPORT

14%
SIMILARITY INDEX

PRIMARY SOURCES

1	journal.uty.ac.id Internet	158 words — 12%
2	www.mdpi.com Internet	11 words — 1%
3	"Pervasive Computing and Social Networking", Springer Science and Business Media LLC, 2022 Crossref	8 words — 1%
4	Muhammad Zakariyah, Umar Zaky. "Analysis of Machine Learning Algorithm for Sleep Apnea Detection Based on Heart Rate Variability", JUITA : Jurnal Informatika, 2022 Crossref	8 words — 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	OFF