Received: 20 Nov 2023 | Revised: 11 Dec 2024 | Accepted: 07 Aug 2025 | Available online: 28 Dec 2025
https://doi.org/10.13057/ijas.v8i1.80562

E-ISSN: 2621-086X
https://jurnal.uns.ac.id/ijas

Indonesian Journal of Applied Statistics Vol. 8, No.1, pp. 38-52, 2025

Comparison of Hard Clustering and Soft Clustering Methods in Grouping Regencies/Cities in West Java Province Based on Regional Vulnerability Indicators to the Impact of Hydrometeorological Disasters in 2021

Hanifah Vida Indrasari[1], Yuliagnis Transver Wijaya[2]*
[1]Badan Pusat Statistik Pulau Morotai, Pulau Morotai, Indonesia
[2]Politeknik Statistika STIS, Jakarta, Indonesia
* Corresponding Author. E-mail: yuliagnis@stis.ac.id

Abstract

Indonesia is an archipelagic country with a high incidence of hydrometeorological disasters and the number is increasing every year. One of the provinces in Indonesia with the highest number of hydrometeorological disasters is West Java Province, where 98.97 percent are hydrometeorological disasters. This is also supported by the characteristics of the area which is dominated by mountains, high rainfall, has 40 watersheds, and has six faults that are suspected to be still active so that it is vulnerable to hydrometeorological disasters. Research on regional vulnerability to hydrometeorological disasters can be carried out by grouping regions based on the same level of vulnerability using the clustering method. The purpose of this study was to group regencies or cities in West Java Province based on indicators of regional vulnerability to the impacts of hydrometeorological disasters in 2021. The clustering method used is hard clustering (single linkage, complete linkage, average linkage, ward's method and k-means) and soft clustering (fuzzy c-means). The most optimal method for grouping regencies or cities in West Java Province is the complete linkage method with a total of 4 clusters. The result is that all the resulting clusters are vulnerable to the characteristics of social vulnerability.

Keywords: cluster analysis; hard clustering; natural disasters; regional vulnerability; soft clustering.

How to Cite:
H. V. Indrasari and Y. T. Wijaya, "Comparison of hard clustering and soft clustering methods in grouping regencies/cities in West Java Province based on regional vulnerability indicators to the impact of hydrometeorological disasters in 2021," *Indonesian Journal of Applied Statistics,* vol. 8, no. 1, pp. 38-52, 2025, doi: 10.13057/ijas.v8i1.80562.

1. INTRODUCTION

Indonesia is an archipelagic country prone to natural disasters. Indonesia is geographically placed between two continents and two oceans. Because of its geographical location, Indonesia is at risk of floods, severe waves, landslides, extreme weather, and drought. Indonesia is also located at the meeting point of the world's three significant plates, which contributes to the country's high catastrophe risk and vulnerability to natural disasters [1]. According to UNISDR [2], disaster is classified into four distinct categories; one of those is caused by hydrometeorological phenomena, also known as hydrometeorological disasters. Hydrometeorological disasters are atmosphere, hydrology, or oceanography natural phenomena that may cause loss of life, injury or other health effects, property damage, loss of livelihoods and services, social and economic upheaval, or environmental damage.

The number of hydrometeorological disasters that occur in Indonesia yearly is increasing, as recorded by The National Agency for Disaster Countermeasure (BNPB). According to disaster data verification by BNPB, the number of natural disasters in Indonesia in 2021 were 6,235 incidents. The total disasters were dominated by hydrometeorological disasters with an occurrence percentage of

98.96% with details: 1,932 floods; 1,817 extreme weather events; 1,727 landslides; 15 droughts; 585 forest and land fires; and 94 tidal waves and abrasion. Meanwhile, the province with the highest number of hydrometeorological disasters is West Java Province, with 2009 incidents.

The West Java Province's BPBD calculated that 98.97% of the disasters in 2021 were hydrometeorological disasters such as floods, landslides, tornadoes, and tidal waves. This is reinforced by the regional characteristics of West Java Province, which is dominated by hills. There are 17 mountains and volcanoes, relatively significant rainfall, 40 watersheds (DAS), and six active faults [3]. West Java Province is at high risk of disaster due to the previous condition described.

Disaster risk assessment is an essential aspect of disaster mitigation, which attempts to reduce the impacts of natural disasters [4],[5]. The government is drafting the 2015-2045 Disaster Management Master Plan (RIPB) with the objective of "Making Indonesia Disaster Resilient for Sustainable Development" which aligns with the vision and purpose of the 2005-2025 RPJPN. Aside from that, Indonesia helped carry out the Sendai Framework for Disaster Risk Reduction (SFDRR) 2015-2030 [6]. Per objective 11 of the SDGs' 5th target, which aims to minimize the number of deaths, the number of people impacted, and the economic losses to global GDP caused by disasters (related to water) by 2030, focusing on safeguarding the poor and vulnerable population.

In 2021, The National Agency for Disaster Countermeasure (BNPB) conducted a disaster risk assessment by computing the Indonesian Disaster Risk Index (IRBI) with three components: hazard, vulnerability, and capacity [7]. IRBI provides an overview of disaster management activities at both the province and district/city levels, and it can also assist the government in developing natural disaster management policies. However, the 2021 IRBI calculation results continue to use hazard and vulnerability data from 2013.

Vulnerability refers to the chance of harm, both to human life and to property. Vulnerability is also described as a measure of prospective losses from hazards and the extent to which society is unable to handle stress caused by disasters [8]. According to [9], vulnerability is classified into two types: social vulnerability and biophysical vulnerability, with the combination of those two resulting in place vulnerability or regional vulnerability. Different loss patterns will come from disparities in social vulnerability in every region [10]. Aside from that, possible losses result from society's interaction with biophysical conditions [9]. If the community's vulnerability is getting higher, natural disasters will have a bigger impact in such instances, and places with high biophysical vulnerability are more likely to suffer losses.

One of the approaches that is widely used to measure social vulnerability is the index approach. One method that is often used is the social vulnerability index (SoVI) developed by Cutter et al. [11], then implemented in Indonesia by Siagian et al., Pangestu et al., and Wijaya et al. [12]-[14]. However, measuring vulnerability using an index still has several weaknesses. One of them is that calculations using indices cannot provide a deeper understanding of social vulnerability indicators, where each region has different social vulnerability factors [15]. Another weakness is that it oversimplifies the complexity and relationship of the various constituent indicators, making it difficult to detect the diversity of vulnerabilities [16]. To overcome these weaknesses, the cluster analysis method has been pioneered by Rufat [16] can be applied.

Research on regional vulnerability to hydrometeorological disasters can be conducted by grouping regions based on the same level of vulnerability using the clustering method. The clustering methods that can be used are divided into hard and soft clustering [17],[18]. Pangestu et al. [13] mapped disaster risk areas using the geographic information system (GIS) in the Central Lampung Regency area. Thamrin and Wijayanto [19] grouped the welfare levels of districts/cities in Java using hard clustering and soft clustering analysis.

Based on the previous description, the researcher is interested to know the general picture of indicators of regional vulnerability to the impact of hydrometeorological disasters, comparing the most optimal cluster methods between hard clustering (hierarchical agglomerative and k-means) and soft

https://doi.org/10.13057/ijas.v8i1.80562
E-ISSN: 2621-086X

clustering (FCM) based on indicators of regional vulnerability to the impact of hydrometeorological disasters, and analyzing districts/cities in West Java Province based on indicators of regional vulnerability to the effects of hydrometeorological disasters in 2021. The methods that will be used are hard clustering (single linkage, complete linkage, average linkage, ward's method, and k-means) and soft clustering (FCM). This regional vulnerability grouping is expected can be useful as the plan and to evaluate government program targets and policies so that they can be more focused on the more vulnerable area to natural disasters, especially hydrometeorological disasters.

2. METHODS

2.1. Hydrometeorological Disasters

UNISDR categorizes disasters into four types: disasters caused by dynamic processes inside the earth, disasters caused by dynamic processes on the earth's surface, disasters caused by hydrometeorological events, and disasters caused by biological processes. Hydrometeorological disasters are natural phenomena that may cause loss of life, injury or other health effects, property damage, loss of livelihood, social and economic disruption, and environmental damage [2]. Extreme meteorological and climatic phenomena, such as floods, droughts, storms, tornadoes, or landslides, create hydrometeorological disasters [20].

2.2. Risk of Disaster

Disaster risk is the potential for catastrophic losses in life, health status, livelihoods, assets, and services that can occur in society over a certain period [2],[21]. Meanwhile, according to Law of the Republic of Indonesia Number 24 of 2007, disaster risk is "the potential loss arising from a disaster in an area and a certain period which can be in the form of death, injury, illness, threatened life, loss of a sense of security, displacement, damage or loss. Property, and disruption of community activities". In calculating the Indonesian disaster risk index (IRBI), risk assessment uses the general formula:

$$Risk = Hazard \times \frac{Vulnerability}{Capacity}$$

2.3. Framework, Cutter et al. [11]

This research used the framework by Cutter et al. in 2003 where risk interacts with mitigation will produce potential danger. Risk itself is said to be the possibility of occurrence or probability of danger. Meanwhile, mitigation is said to be an effort to reduce the risks such as the previous plans and experiences. Chances can be weakened by proper mitigation, or they can be strengthened by poor mitigation practices [9].

The interaction between social vulnerability and biophysical vulnerability will produce regional vulnerability or place vulnerability, where place vulnerability provides feedback for risk and mitigation that can reduce or increase risk and mitigation. Potential losses also come from the interaction of society with biophysical conditions [9]. Suppose the vulnerability of the community is higher, in that case, the impact of natural disasters will be more significant, as well as areas that have high biophysical vulnerability are more likely to experience losses as well [22].

Vulnerability is defined as the possibility of damage, both in terms of human life and property. Vulnerability is also defined as a measure of potential losses from hazards and the extent of a community's inability to manage stress due to disasters that occur them [8]. Social vulnerability is a measure of a society's sensitivity to natural hazards and ability to respond to and recover from the impacts of hazards [23]. Meanwhile, biophysical vulnerability can be defined as the exposure of human systems to extreme natural events [24]. Several indicators that can characterize biophysical vulnerability are the location of people's residences in dangerous zones, the level of losses associated with disasters, the frequency of natural disasters, magnitude, duration, availability of natural resources, quality of

https://doi.org/10.13057/ijas.v8i1.80562
E-ISSN: 2621-086X

buildings, as well as land use and land cover [9],[25]. In this study, biophysical vulnerability indicators are approached by the percentage of habitable houses, damage to houses or residences, and other facilities.

2.4. Clustering

Cluster analysis or clustering is a data exploration method to obtain hidden characteristics by forming groups or clusters of data without any information in the form of labels and mechanisms. It is carried out based on similarities or dissimilarities, such as Euclidean distance [17],[18]. Clustering aims to group a set of objects with similar characteristics into one cluster and things with different aspects into another cluster by maximizing the similarity between objects in one cluster and minimizing the similarity between clusters [26]. Groupings of objects that are in the same cluster will be more similar than objects that are outside the cluster. The most extensive Dunn index and silhouette coefficient values can help determine the optimal cluster. The greater the Dunn index value in a cluster, the better the clustering results [27]. Likewise, with the silhouette coefficient, the cluster formed will be better if the value is closer to 1 [19]. The similarity in characteristics between objects can be identified by looking at the close distance between objects. Based on the cluster membership value, clustering methods can be divided into hard and soft clustering.

2.5. Hard Clustering

In the hard clustering method, the membership of an observation unit consists of only two possibilities, namely, whether the observation is included in the cluster or not included in the cluster [17]. Hard clustering is divided into two, namely hierarchical and non-hierarchical. The hierarchical grouping method is a cluster analysis that groups similar objects in adjacent hierarchies and dissimilar things in distant scales. Hierarchical methods can be classified into two, namely agglomerative and divisive. The divisive hierarchical method is also called the top-down approach. Meanwhile, the agglomerative hierarchical method is also called the bottom-up approach. The hierarchical agglomerative methods used in this research are single linkage, complete linkage, average linkage, and ward's method.

Single Linkage uses the smallest or closest distance between one object in a cluster and one thing in another cluster. Complete linkage uses the most significant distance between one object in a cluster and one object in another cluster. Average Linkage uses the average distance between objects from one cluster and objects from others [28]. Meanwhile, Ward's method, where the selection of two clusters to be combined is based on which combination of clusters minimizes the sum of squared error (SSE) value in the cluster across a collection of separate clusters. At each step, the two clusters are combined to give the smallest SSE value [19].

Meanwhile, the non-hierarchical grouping method will group objects into several k clusters, where the number of clusters has been determined in advance or as part of the grouping procedure [28]. Non-hierarchical methods can be applied to much larger data sets than hierarchical methods. One of the most popular methods is the k-means method. K-means clustering will group objects into several k clusters that have the closest centroid (average) [28].

2.6. Soft Clustering

In the soft clustering method, the membership of an observation unit is expressed through the degree of membership in each cluster [18]. One of the simplest and most frequently used methods in soft clustering is the fuzzy c-means (FCM) method. FCM was first developed by Dunn in 1973 then refined in 1981 by Bezdek [29]. The FCM method is an improvement on the k-means algorithm, where the membership of an observation unit is expressed through the degree of membership in each cluster with a membership value of 0 to 1 [18]. In FCM, there is a value of m defined as a fuzzifier that controls the possibility of intersections between clusters [17]. In most data, a good fuzzifier value is between 1.5 to 3.0 [30]. Then, Wu [31] recommends a value of m ranging from 1.5 to 4 for data that contains noise and outliers.

https://doi.org/10.13057/ijas.v8i1.80562
E-ISSN: 2621-086X

2.7. Data

The data sources used in this research are secondary data obtained from the website and publications of the Central Statistics Agency (BPS) of West Java Province, the West Java Open Data website, and The National Agency for Disaster Countermeasure (BPBD) of West Java Province. This research covers all regencies/cities in West Java Province with analysis units of 27 regencies/cities. The variables used in this research and their data sources are shown in Table 1.

Table 1. List of indicators used and their sources

Indicators	Sources of Data	Sources of Research
The percentage of people with poverty	West Java Province in 2022 (figures)	[32], [11], [33], [34], [14], [12].
Population density	West Java Province in 2022 (figures)	[14], [32], [15], [33], [35], [36], [8].
The number of people with disabilities	The West Java Open Data (website)	[15], [37].
Sex Ratio	West Java Province in 2022 (figures)	[14], [11], [12], [39].
The percentage of the population aged 0 to 4 years The percentage of the population aged >65 years	The West Java Open Data (website) The West Java Open Data (website)	[15], [11], [12], [14], [32].
Population growth rate	West Java Province in 2022 (figures)	[12], [11], [38].
The percentage of female heads of households	The West Java Open Data (website)	[14], [11], [12].
The open unemployement rate Labor-force participation rate	West Java Province in 2023 (figures) West Java Province in 2024 (figures)	[14], [15], [34], [38].
Gross enrollment at high school level	BPS (website)	[15], [35], [36], [8].
The number of health facilities	West Java Province in 2022 (figures)	[11], [39], [40], [15], [37].
The percentage of viable houses	Housing Statistics of West Java Province 2021	[4], [33].
The number of villages/sub-districts that have an early warning system for natural disasters. The number of villages/subdistricts with signage and evacuation routes	Potential Village Statistics of West Java Province 2021 Potential Village Statistics of West Java Province 2021	[34], [41], [13].
GDP at fixed costs per capita	BPS (website)	[4], [36], [35], [11], [42], [43].
Damaged house/residence Damaged facilities	The National Disaster Relief Agency West Java Province The National Disaster Relief Agency West Java Province	[4]

https://doi.org/10.13057/ijas.v8i1.80562
E-ISSN: 2621-086X

3. RESULTS AND DISCUSSION

3.1. General description

West Java Province has the highest number of incidents in 2021. According to the National Disaster Relief Agency (BPBD) of West Java Province, in 2021, there were 2.429 natural disasters in West Java Province, of which 98.97% were hydrometeorological disasters. There are 18 indicators used in this research. Meanwhile, a general overview of regional vulnerability indicators in West Java Province in 2021 is presented in a descriptive table that includes minimum values, maximum values, averages, and standard deviations. An illustrative table is shown in Table 2, where the statistical significance of each indicator varies greatly.

Table 2. Descriptive statistics of regional vulnerability indicators

Indicators	Min	Max	Averages	Standard Deviation
The percentage of people with poverty	2.58	13.13	8.97	2.89
Population density	423	14630	3896	4584.33
The number of people with disabilities	157	8139	1332	1896.98
Sex Ratio	100.30	105.70	102.40	1.52
The percentage of the population aged 0 - 4 years	6.24	8.53	7.39	0.57
The percentage of the population aged >65 years	3.46	9.91	6.24	1.60
Population growth rate	0.48	1.93	1.37	0.38
The percentage of female heads of households	16.57	23.22	20.48	1.50
Open unemployement rate	3.25	13.07	9.40	2.43
Labor-force participation rate	56.86	74.75	65.03	3.58
Gross enrollment at high school level	52.97	114.29	80.63	14.50
The number of health facilities	31	482	224	146.29
The percentage of building resilience	53.93	95.50	79.96	11.26
The number of villages/sub-districts that have an early warning system for natural disasters.	0	146	29.93	29.76
The number of villages/subdistricts with signage and evacuation routes	1	66	20.33	19.56
GDP at fixed costs per capita	13124	81704	29546	19555.95
Damaged house/residence	5	1338	212.80	312.13
Damaged facilities	0	262	50.74	67.58

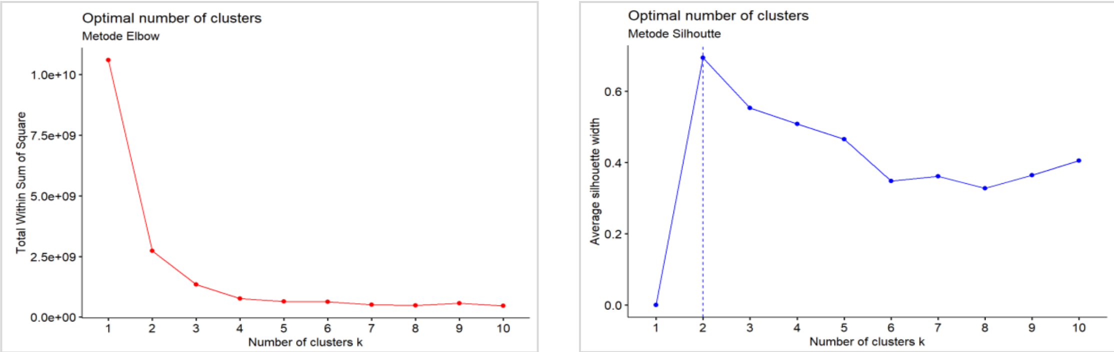
3.2. Comparing the most optimal cluster methods

Table 2 shows that the 18 indicators used have different units, so standardization needs to be done first before carrying out the cluster analysis. The next step is multicollinearity checking. Because there is no multicollinearity between the indicators used in the research, cluster analysis can be done using hard clustering and soft clustering methods using Euclidean distance.

Non-Hierarchical

The initial stage before doing an analysis using the k-means clustering method is determining the number of clusters that will be used first [27]. In this research, the methods used to determine the optimum number of clusters are the elbow and silhouette methods.

https://doi.org/10.13057/ijas.v8i1.80562
E-ISSN: 2621-086X



(a) (b)

Figure 1. Elbow (a) and silhouette (b) in k-means.

Based on Figure 1, researchers will compare the number of clusters 3 and 4. After determining the number of clusters, the next step is to run the k-means clustering algorithm on the dataset. The following are the results of the validity of the k-means method that are shown in Table 3.

Table 3. The result of the validity of k-means

	Number of clusters	
	3	4
Dunn	0.2896	0.4352
Silhouette	0.1633	0.1745

It can be concluded that by using the k-means method, the optimum number of clusters is 4 clusters, where the dunn index and silhouette coefficient values in cluster 4 are bigger than in cluster 3. A plot of cluster analysis using k-means with 4 clusters is shown in Figure 2.

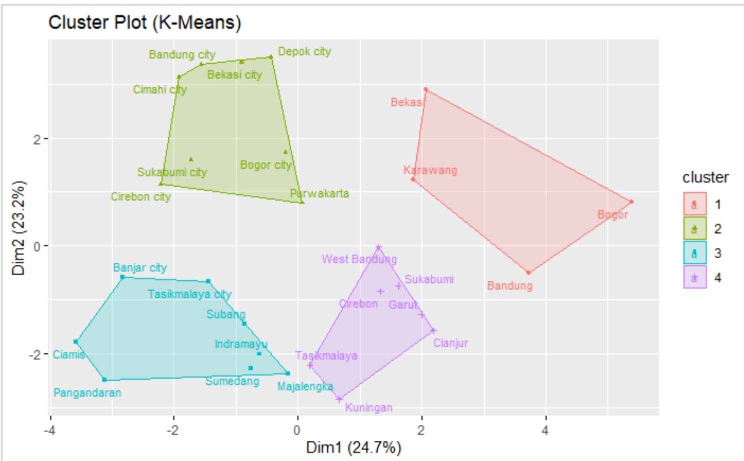


Figure 2. Plot clusters of k-means

Hierarchical

The hierarchical agglomerative method used in this research are single linkage, complete linkage, average linkage, and Ward. In this research, the number of clusters to be compared follows the number of clusters in k-means method which are 3 and 4 clusters. The following are the results of the validity of the hierarchical agglomerative method in Table 4.

https://doi.org/10.13057/ijas.v8i1.80562
E-ISSN: 2621-086X

Table 4. The results of the validity of the hierarchical agglomerative method

Methods	Clusters	*Dunn*	Silhouette
Single	3	0.736	0.1315
	4	0.4889	0.0962
Complete	3	0.3969	0.1724
	4	0.4559	0.1771
Average	3	0.3650	0.1688
	4	0.4559	0.1746
Ward	3	0.3991	0.1648
	4	0.3969	0.1327

In the hierarchical agglomerative method, the most optimal method is complete linkage with a total of 4 clusters. The silhouette coefficient value in cluster 4 is the highest and then followed by Dunn index value is the 3rd highest. The following is a dendrogram from cluster analysis using complete linkage with 4 clusters. Below is a dendrogram of the complete linkage method grouping in Figure 3.

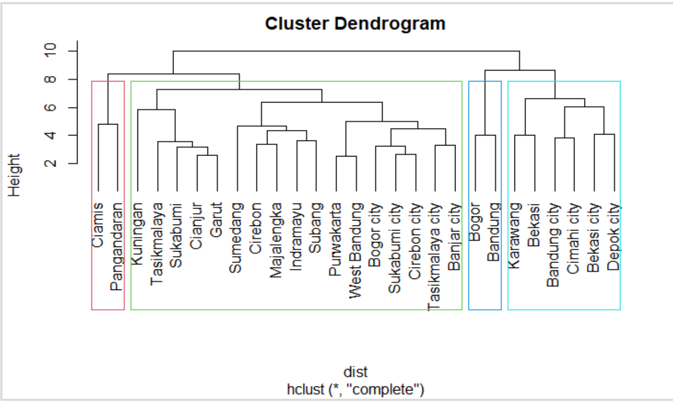
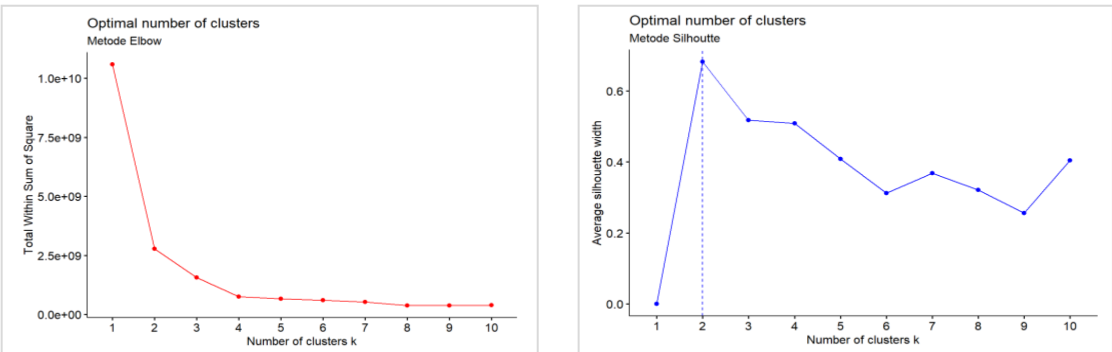


Figure 3. Complete linkage's dendrogram

Fuzzy c-means (FCM)

Like the k-means method, the initial stage before implementing FCM analysis is to determine the number of clusters (k) that will be used. Based on Figure 4, researchers will compare the number of clusters 3 and 4. After determining the number of clusters, the next step is to run the FCM algorithm on the dataset used. Following the suggestions of Bezdek et al. [30], this study used a fuzzifier value between 1.5 to 3.



(a) (b)

Figure 4. Elbow (a) and silhouette (b) in FCM.

https://doi.org/10.13057/ijas.v8i1.80562
E-ISSN: 2621-086X

Using two clusters and predetermined fuzzifier values, the most optimal number of clusters and fuzzifier values is selected by calculating the validity index value. The validity indices used are the partition coefficient (PC) index, modified partition coefficient (MPC) index, classification entropy (CE) index, separation index (S), and Xie and Beni's index (XB). The highest the PC and MPC index values, the more optimal grouping results. Meanwhile, smaller CE, S, and XB index values indicate more optimal grouping results. The results of calculating the five index values are presented in Table 5.

Table 5. The validity index based on the number of clusters and fuzzifier value

Cluster	Fuzzifier	PC	MPC	CE	XB	S
3	1.5	0.582	0.372	0.739	0.918	0.652
	2	0.333	0.000	1.099	11.451.612	11.451.612
	2.5	0.333	0.000	1.099	8.295815E+17	1.436877E+18
	3	0.333	0.000	1.099	3.952384E+15	1.185715E+22
4	1.5	0.542	0.390	0.881	0.903	0.606
	2	0.250	0.000	1.386	636754.264	636754.264
	2.5	0.250	0.000	1.386	2.293597E+16	4.587195E+16
	3	0.250	0.000	1.386	1.534616E+18	6.138465E+18

Based on the values in Table 5, the optimal number of clusters and fuzzifier value for the FCM method is obtained which are 4 clusters with a fuzzifier value of 1.5. The number of clusters and fuzzifier values chosen are based on the highest MPC index value and the smallest CE, XB, and S index values. Using the FCM method, the number of clusters formed to group regencies/cities based on regional vulnerability to the impact of hydrometeorological disasters in West Java Province in 2021 is 4 clusters. Determining the cluster for each regency/city in West Java Province is based on the highest membership degree value. A plot of cluster analysis using FCM with 4 clusters is shown in Figure 5.

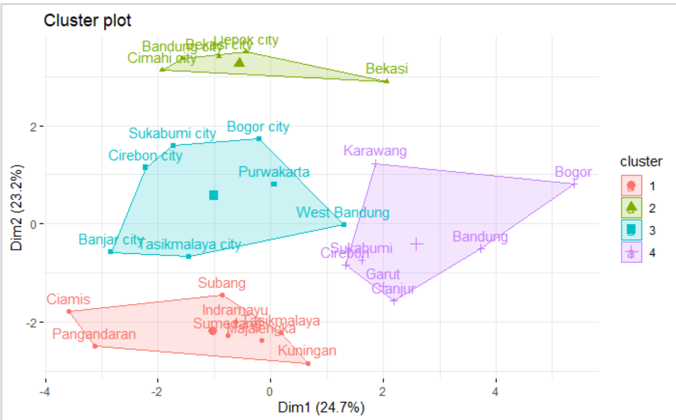


Figure 5. Cluster plot of FCM

Determining the best cluster

The best or the most optimal cluster method in this research is seen from the highest values of Dunn index and Silhouette coefficient. The highest the Dunn index value in a cluster, the better the clustering results [26]. Likewise, with the silhouette coefficient, if the value is closer to 1, the cluster formed will be better [27]. To find out the best cluster method between hard clustering and soft clustering that has been done before, validation was carrried out using the Dunn index and silhouette coefficient on the k-means, complete linkage, and FCM methods (with a fuzzifier value of 1.5) with 4 clusters. The validity results of the three methods are presented in Table 6.

Table 6. The validity results of three methods

	Complete	K-means	FCM
Dunn	0.4559	0.4352	0.4103
Silhouette	0.1771	0.1745	0.1376

It can be seen from Table 6 that the most optimal method for grouping regencies/cities in West Java Province based on data of regional vulnerability to the impact of hydrometeorological disasters is the complete linkage method with the optimal number of clusters are 4 clusters. The Dunn index and silhouette coefficient values for the complete linkage method are the highest compared to other methods that are 0.4559 and 0.1771. So, it can be concluded that in this research, hard clustering (complete linkage) is better than soft clustering, as seen from the Dunn index value and silhouette coefficient values. This aligns with research conducted by [44] that the complete linkage method is the best method for grouping regencies/cities in West Java Province.

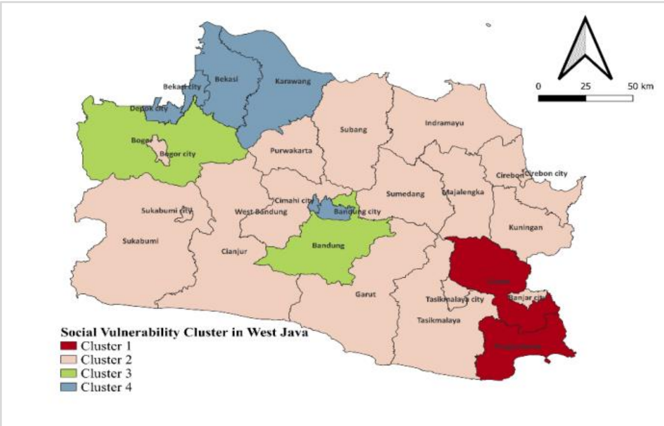


Figure 6. Thematic maps cluster with complete linkage method

Figure 6 presents a thematic map of the results of grouping regencies/cities based on indicators of regional vulnerability to hydrometeorological disasters in West Java Province in 2021 using the hard clustering (complete linkage) method with 4 clusters.

The number of members for each cluster are: cluster 1 consists of 2 regencies/cities, cluster 2 consists of 17 regencies/cities, cluster 3 consists of 2 regencies/cities, and cluster 4 consists of 6 regencies/cities. Then, the average of each indicator in each cluster was calculated to see the characteristics of the cluster obtained, presented in Table 7.

Table 7. The average of each indicator in each cluster

Indicators	Cluster 1	Cluster 2	Cluster 3	Cluster 4
The percentage of people with poverty	8.81	10.48*	7.64	5.20
Population density	649.00	2570.71	2049.50	9347.83*
The number of people with disabilities	5799.50*	559.59	1425.50	1998.33
Sex ratio	100.30*	102.54	104.85	102.07
The percentage of the population aged 0 to 4 years	6.26	7.58	7.71*	7.11
The percentage of the population aged >65 years	9.62*	6.48	4.61	4.98
Population growth rate	1.09	1.41	1.55*	1.31
The percentage of female heads of households	20.73	21.16*	18.85	19.00
Open unemployement rate	4.16	9.29	10.27	11.18*
Labor-force participation rate	72.29*	64.66	63.84	64.09
Gross enrollment at high school level	99.54	75.15	64.49*	95.26
The number of health facilities	68.50*	180.71	423.50	330.67
The percentage of building resilience	78.23	82.18	80.38	74.13*

https://doi.org/10.13057/ijas.v8i1.80562
E-ISSN: 2621-086X

Indicators	Cluster 1	Cluster 2	Cluster 3	Cluster 4
The number of villages/sub-districts that have an early warning system for natural disaster	38.50	33.53	37.00	14.50*
The number of villages/sub-districts with signane and evacuation routes	18.50	23.00	28.50	10.67*
GDP at fixed cost per capita	18581.19*	22804.85	25985.03	53488.85
Damaged house/residence	257.00	133.47	1067.00*	138.00
Damaged facilities	15.50	49.53	196.50*	17.33

*Characteristics that are deemed susceptible

The results of clustering districts/cities in West Java using the hard clustering (complete linkage method) produced four groups, those are:

1. Cluster 1 is the cluster with the most regional vulnerability indicators compared to other clusters, with the six most vulnerable indicators. These indicators are the number of people with disabilities, sex ratio, percentage of population aged > 65 years, TPAK, number of health facilities, and GDP at fixed cost per capita. People with disabilities and the elderly are categorized as vulnerable populations and have a higher potential for experiencing loss, damage or losses due to natural disasters, including hydrometeorological disasters [15]. The characteristics of a low sex ratio indicate that a female population dominates the area. In recovering from natural disasters, women also experience a more difficult time than men [11]. Characteristics of the labor force participation level (TPAK): When a disaster occurs, the working population will risk losing their jobs, resulting in slow post-disaster recovery [11]. The following characteristics are the number of health facilities and GRDP per capita at constant prices. According to Irmayani et al. [36], high GDP per capita at constant prices in a region can reduce the amount of losses from natural disasters. In addition, the lack of nearby medical services will prolong first aid and long-term disaster recovery, so adequate health facilities are essential [15]. On average, this cluster has a high number of hydrometeorological disasters, with around 89 disasters in 2021, which caused 3,928 fatalities. In this cluster, the most frequent disasters are landslides followed by tornadoes [45]. The frequency of natural disasters is an aspect of the natural environment that can influence the level of damage. According to Taghizadeh-Hesary et al. [4] the intensity or frequency of disasters is the factor that has the highest impact on the total damage caused by disasters.

2. Cluster 2 is the cluster with the lowest regional vulnerability indicators compared to other clusters, where there are two most vulnerable indicators. These indicators are the percentage of people with poverty and the percentage of female heads of households. Poverty level is one factor that causes communities to be affected by disasters differently [33],[34]. According to Cutter et al. [11], wealth allows people to recover from losses more quickly and vice versa. Furthermore, the high percentage of female heads of households will increase vulnerability to the impacts of hydrometeorological disasters. According to Cutter et al. [11], households headed by women are more vulnerable to natural disasters, including hydrometeorological disasters, because the majority of women have lower wages and have to be responsible for family care. On average, this cluster has a relatively high number of hydrometeorological disasters, around 74 disasters in 2021 and the most frequent disasters are landslides [45]. This cluster has a relatively high average number of hydrometeorological disasters, with around 74 disasters in 2021, and the most frequent disaster is landslides [45], with an average death toll of around 25,559 people.

3. Cluster 3 is a cluster with five regional vulnerability indicators that are most vulnerable compared to other clusters. These indicators are the percentage of the population aged 0-4 years, population growth rate, gross enrollment rate (APK) at high school education level, damage to houses/residences, and damage to other facilities. According to Maharani et al. [15], children (aged 0-14 years) and the elderly have a higher potential to experience loss, damage or loss due to natural

https://doi.org/10.13057/ijas.v8i1.80562
E-ISSN: 2621-086X

disasters. Characteristics of the gross enrollment rate (APK) at the high school education level: According to Irmayani et al. [36], disaster preparedness increases with a person's education level. The next indicator is the amount of damage due to hydrometeorological disasters. The average damage to houses/residences and other facilities in this cluster is the highest. This aligns with the average number of hydrometeorological disasters, around 412 incidents, with the fatalities average of 22,409 people.

4. Cluster 4 has five vulnerability indicators for the most vulnerable areas. These indicators are population density, open unemployment rate (TPT), percentage of viable houses, number of villages/subdistricts that have natural disaster early warning systems, and number of villages/subdistricts that have signane and evacuation routes. An area with a high population density will increase socio-economic activity and hurt the environment if it does not align with the population's quality [38]. Thus, the higher the population density of an area, the higher the vulnerability to the impacts of hydrometeorological disasters. According to Maharani et al. [15], unemployment will slow the recovery from the impact of hydrometeorological disasters. Someone unemployed will become vulnerable because they do not have sufficient income and inadequate resources to meet post-disaster needs [34]. Livable houses are crucial because they have a direct disaster potential [7]. Taghizadeh-Hesary et al. [4] explain that the quality of the infrastructure will reduce long-term damage due to natural disasters. The final characteristic is the number of villages/subdistricts with a natural disaster early warning system and the number of villages/subdistricts with signs and evacuation routes. In dealing with those two vulnerability indicators, the government has outlined it in Law Number 24 of 2007 concerning Disaster Management in Article 44. The government has also prioritized providing a disaster early warning system and mitigation infrastructure, as well as preparedness (such as temporary evacuation sites, evacuation routes, and evacuation signs) to reduce vulnerability to natural disasters [7]. Besides, this cluster has the lowest average number of hydrometeorological disasters, with around 23 disasters in 2021, with the most frequent disaster is tornadoes.

4. CONCLUSIONS

It can be concluded that hydrometeorological disasters dominate natural disasters in West Java Province, and the distribution of each vulnerability indicator in each region is very diverse and different. Then, the most optimal cluster method between hard clustering and soft clustering in grouping regencies/cities based on indicators of regional vulnerability to the impact of hydrometeorological disasters in West Java Province in 2021 is the hard clustering (complete linkage) method. The results of grouping using the complete linkage method produced 4 clusters.

1. Cluster 1 consists of 2 regencies/cities with the six most vulnerable indicators, those are: the number of people with disabilities, sex ratio, percentage of the population aged >65 years, TPAK, number of health facilities, and GDP at a fixed cost per capita. It can be said that cluster 1 is vulnerable to social vulnerability.
2. Cluster 2 consists of 17 regencies/cities with the two most vulnerable indicators, those are: the percentage of people with poverty and the percentage of female heads of households. It can be said that cluster 2 is vulnerable to social vulnerability.
3. Cluster 3 consists of 2 regencies/cities with the five most vulnerable indicators, those are: the percentage of population aged 0-4 years, population growth rate, gross enrollment rate (APK) at high school education level, damaged houses/residences, and damaged other facilities. It can be said that cluster 3 is also more vulnerable to social vulnerability characteristics than biophysical vulnerability characteristics.
5. Cluster 4 consists of 6 regencies/cities with the five most vulnerable indicators, those are: population density, open unemployment rate (TPT), number of villages/subdistricts that have

https://doi.org/10.13057/ijas.v8i1.80562
E-ISSN: 2621-086X

natural disaster early warning systems, number of villages/subdistricts that have signane and routes evacuation, and percentage of viable houses. It can be said that cluster 4 is also more vulnerable to social vulnerability characteristics than biophysical vulnerability characteristics.

ACKNOWLEDGEMENTS

The author would like to thank the National Disaster Management Agency, the Regional Disaster Management Agency of West Java Province, and Mr. Ebet Nugraha, S.Kom., M.A as the resource person; As well as everyone who assisted and supported this research.

REFERENCES

[1] Badan Nasional Penanggulangan Bencana, *Risiko Bencana Indonesia*. Jakarta, Indonesia: BNPB, 2016.

[2] United Nations International Strategy for Disaster Reduction (UNISDR), *United Nations Report*, 2009.

[3] Badan Pusat Statistik Provinsi Jawa Barat, *Statistik Daerah Provinsi Jawa Barat 2022*. Bandung, Indonesia: BPS Jabar, 2022.

[4] F. Taghizadeh-Hesary, T. Sarker, N. Yoshino, A. Mortha, and X. V. Vo, "Quality infrastructure and natural disaster resiliency: A panel analysis of Asia and the Pacific," *Econ. Anal. Policy*, vol. 69, pp. 394–406, 2021, doi: 10.1016/j.eap.2020.12.021.

[5] T. H. Siagian and N. B. Parwanto, *Mengukur Risiko dan Kerentanan Bencana pada Skala Lokal di Indonesia melalui Downscaling World Risk Index*. 2017.

[6] BNPB and Bappenas, *Rencana Induk Penanggulangan Bencana 2015-2045*, 2018.

[7] BNPB, *IRBI Indeks Risiko Bencana Indonesia Tahun 2021*. Jakarta, Indonesia: BNPB, 2021.

[8] S. Jeong and D. K. Yoon, "Examining vulnerability factors to natural disasters with a spatial autoregressive model: The case of south Korea," *Sustain.*, vol. 10, no. 5, pp. 1–13, 2018, doi: 10.3390/su10051651.

[9] S. L. Cutter, "Vulnerability to hazards," *Prog. Hum. Geogr.*, vol. 20, no. 4, pp. 529–539, 1996.

[10] M. C. Schmidtlein, J. M. Shafer, M. Berry, and S. L. Cutter, "Modeled earthquake losses and social vulnerability in Charleston , South Carolina," *Appl. Geogr.*, vol. 31, no. 1, pp. 269–281, 2011, doi: 10.1016/j.apgeog.2010.06.001.

[11] S. L. Cutter, B. J. Boruff, and W. L. Shirley, "Social vulnerability to environmental hazards," *Social Science Quarterly*, vol. 84, no. 2, pp. 242-261, 2003, https://doi.org/10.1111/1540-6237.8402002.

[12] T. H. Siagian, P. Purhadi, S. Suhartono, and H. Ritonga, "Social vulnerability to natural hazards in Indonesia: Driving factors and policy implications," *Nat. Hazards*, vol. 70, no. 2, pp. 1603–1617, 2014, doi: 10.1007/s11069-013-0888-3.

[13] H. D. Pangestu, A. D. Putra, and A. Syah, "Analisis indeks risiko dan potensi kebencanaan (studi untuk wilayah Kabupaten Lampung Tengah)," *J. Rekayasa Sipil …*, vol. 9, no. 3, pp. 481–490, 2021, [Online]. Available: http://repository.lppm.unila.ac.id/id/eprint/36857

[14] Y. T. Wijaya and I. T. Halim, "Measuring and profiling social vulnerability to natural disaster in Indonesia in 2019," *J. Mat. Stat. dan Komputasi*, vol. 19, no. 1, pp. 183–194, 2022, doi: 10.20956/j.v19i1.21686.

https://doi.org/10.13057/ijas.v8i1.80562
E-ISSN: 2621-086X

[15] Y. N. Maharani, A. R. B. Nugroho, D. F. Adiba, and I. Sulistiyowati, "Pengaruh kerentanan sosial terhadap ketangguhan masyarakat dalam menghadapi bencana erupsi Gunung Merapi di Kabupaten Sleman," *J. Dialog Penanggulangan Bencana*, vol. 11, no. 1, pp. 1–12, 2020, [Online]. Available: https://www.bnpb.go.id/jurnal/jurnal-dialog-penanggulangan-bencana-vol-11-no-1-tahun-2020

[16] S. Rufat, "Spectroscopy of urban vulnerability," *Ann. Assoc. Am. Geogr.*, vol. 103, no. 3, pp. 505–525, 2013, doi: 10.1080/00045608.2012.702485.

[17] S. Pramana, B. Yuniarto, S. Mariyah, I. Santoso, and R. Nooraeni, *Data mining dengan R Konsep Serta Implementasi*. Jakarta: IN MEDIA, 2018.

[18] B. Balasko, J. Abonyi, and B. Feil, *Fuzzy Clustering and Data Analysis Toolbox For Use with Matlab*. Veszprem: Department of Process Engineering University of Veszprem, 2005.

[19] N. Thamrin and A. W. Wijayanto, "Comparison of soft and hard clustering: a case study on welfare level in cities on Java Island," *Indones. J. Stat. Its Appl.*, vol. 5, no. 1, pp. 141–160, 2021, doi: 10.29244/ijsa.v5i1p141-160.

[20] H. Wu, M. Huang, Q. Tang, D. B. Kirschbaum, and P. Ward, "Hydrometeorological hazards: monitoring, forecasting, risk assessment, and socioeconomic responses," *Adv. Meteorol.*, vol. 2016, pp. 11–14, 2016, doi: 10.1155/2016/2367939.

[21] J. Han, J. Pei, and H. Tong, *Data Mining: Concepts and Techniques, 4th Ed*. Cambridge, MA, US: Morgan Kaufmann publications, 2022.

[22] S. L. Cutter, J. T. Mitchell, and M. S. Scott, "Revealing the vulnerability of people and places : a case study of Georgetown County, South Carolina," *Annals of the Association of American Geographers*, vol. 90, no. 4, pp. 713–737, 2004, https://doi.org/10.1111/0004-5608.00219.

[23] S. L. Cutter and C. Finch, "Temporal and spatial changes in social vulnerability to natural hazards," *Plan. Clim. Chang. A Read. Green Infrastruct. Sustain. Des. Resilient Cities*, vol. 105, no. 7, pp. 2301–2306, 2008, doi: 10.4324/9781351201117-16.

[24] L. Clare and B. P. Weninger, "Social and biophysical vulnerability of prehistoric societies to rapid climate change," *Doc. Praehist.*, vol. 37, no. 1, pp. 283–292, 2011, doi: 10.4312/dp.37.24.

[25] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis, 7th ed*. Upper Saddle River, NJ, USA: Pearson Prentice Hall, 2010.

[26] A. Ramadhan, K. Prawita, M. A. Izzudin, and G. Amandha, "Analisis strategi dan klasterisasi ketahanan pangan nasional dalam menghadapi pandemi covid-19," *Teknol. Pangan Media Inf. dan Komun. Ilm. Teknol. Pertan.*, vol. 12, no. 1, pp. 110–122, 2021, doi: 10.35891/tp.v12i1.2179.

[27] J. Han, J. Pei, and H. Tong, *Data Mining: Concepts and Techniques, 4th Ed*. Cambridge, MA, US: Morgan Kaufmann publications, 2022.

[28] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis (6th ed.)*. 2007.

[29] H. Hanniva, A. Kurnia, S. Rahardiantoro, and A. A. Mattjik, "Penggerombolan kabupaten/kota di Indonesia berdasarkan indikator indeks pembangunan manusia menggunakan metode K-Means dan Fuzzy C-Means," *Xplore J. Stat.*, vol. 11, no. 1, pp. 36–47, 2022, doi: 10.29244/xplore.v11i1.855.

[30] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: the fuzzy c-means clustering algorithm," *Comput. Geosci.*, vol. 10, no. 2–3, pp. 191–203, 1984, doi: 10.1109/igarss.1988.569600.

[31] K.-L. Wu, "Analysis of parameter selections for fuzzy c-means," *Pattern Recognit.*, vol. 45, no. 1, pp. 407–415, 2012, doi: 10.1016/j.patcog.2011.07.012.

https://doi.org/10.13057/ijas.v8i1.80562
E-ISSN: 2621-086X

[32] M. Habibi and I. Buchori, "Model spasial kerentanan sosial ekonomi dan kelembagaan terhadap bencana Gunung Merapi," *Teknik PWK (Perencanaan Wilayah Kota)*, vol. 2, no. 1, pp. 1–10, 2013, https://doi.org/10.14710/tpwk.2013.1402.

[33] Y. W. Rabby, B. Hossain, and M. U. Hasan, "Social vulnerability in the coastal region of Bangladesh: An investigation of social vulnerability index and scalar change effects," *Int. J. Disaster Risk Reduct.*, p. 101329, 2019, doi: 10.1016/j.ijdrr.2019.101329.

[34] K. F. Dintwa, G. Letamo, and K. Navaneetham, "Measuring social vulnerability to natural hazards at the district level in Botswana," *Jamba J. Disaster Risk Stud.*, vol. 11, no. 1, pp. 1–11, 2019, doi: 10.4102/JAMBA.V11I1.447.

[35] K. Songwathana, "The relationship between natural disaster and economic development: a panel data analysis," *Procedia Eng.*, vol. 212, no. 2017, pp. 1068–1074, 2018, doi: 10.1016/j.proeng.2018.01.138.

[36] S. Irmayani, Z. Azhar, and M. R. Adry, "Pengaruh faktor ekonomi, sosial ekonomi dan iklim terhadap bencana alam di Indonesia," *Jurnal Ecogen*, vol. 1, no. 3, pp. 1–13, 2018, doi: http://dx.doi.org/10.24036/jmpe.v1i3.5023

[37] A. L. Nugraha, M. Awaluddin, A. Sukmono, and N. Wakhidatus, "Pemetaan dan penilaian kerentanan bencana alam di Kabupaten Jepara berbasis sistem informasi geografis," *Geoid*, vol. 17, no. 2, p. 185, 2022, doi: 10.12962/j24423998.v17i2.9370.

[38] Y. Zhou, N. Li, W. Wu, J. Wu, and P. Shi, "Local Spatial and Temporal Factors Influencing Population and Societal Vulnerability to Natural Disasters," *Risk Anal*, vol. 34, no. 4, pp. 614–639, 2014, doi: 10.1111/risa.12193.

[39] C. H. T. Watung, R. L. E. Sela, and L. Tondobala, "Tingkat ketangguhan dan ketahanan Kota Manado terhadap bencana," *J. Perenc. Wil. dan Kota*, vol. 5, 2018, doi: 10.5614/jpwk.2014.25.1.1.

[40] A. Djuraidah, "Indeks kerentanan sosial ekonomi untuk bencana alam di Indonesia," *Pros. Semin. Nas. Mat. dan Pendidik. Mat.*, 2009.

[41] perka BNPB, *Peraturan Kepala Badan Nasional Penanggulangan Bencana Nomor 02 Tahun 2012 Tentang Pedoman Umum Pengkajian Risiko Bencana*. 2012, pp. 1–67. [Online]. Available: https://www.bnpb.go.id/uploads/24/peraturan-kepala/2012/perka-2-tahun-2012-tentang-pedoman-umum-pengkajian-resiko-bencana.pdf

[42] I. Schumacher and E. Strobl, "Economic development and losses due to natural disasters: The role of hazard exposure," *Ecol. Econ.*, vol. 72, pp. 97–105, 2011, doi: 10.1016/j.ecolecon.2011.09.002.

[43] G. Barone and S. Mocetti, "Natural disasters, growth and institutions: A tale of two earthquakes," J. Urban Econ., vol. 84, pp. 52–66, 2014, doi: 10.1016/j.jue.2014.09.002.

[44] U. Syafiyah, D. P. Puspitasari, I. Asrafi, B. Wicaksono, and F. M. Sirait, "Analisis perbandingan hierarchical dan non-hierarchical clustering pada data indikator ketenagakerjaan di Jawa Barat tahun 2020," *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, pp. 803–812, 2022, doi: 10.34123/semnasoffstat.v2022i1.1221.

[45] Badan Penanggulangan Bencana Daerah Provinsi Jawa Barat, *Rekapitulasi Kejadian Bencana di Wilayah Provinsi Jawa Barat*. Bandung, Indonesia: BPBD Jabar, 2021.

https://doi.org/10.13057/ijas.v8i1.80562
E-ISSN: 2621-086X