Clustering Methods in Grouping Rural Destinations 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]

[1,2]Statistics, STIS, Indonesia

Corresponding author: hanifahvida3101@gmail.com

Abstract
Indonesia is an archipelagic country with a high incidence of hydrometeorological disasters, and this incidence is increasing annually. One of the provinces in Indonesia with the highest number of hydrometeorological disasters is West Java Province, where 98.97 per cent of the disasters are hydrometeorological. The area's characteristics also support this: it is dominated by mountains, experiences high rainfall, has 40 watersheds, and contains six faults suspected to remain active, making it vulnerable to hydrometeorological disasters. Research on regional vulnerability to hydrometeorological disasters can be conducted by clustering regions into groups with similar vulnerability levels. The purpose of this study was to group regencies or cities in West Java Province based on indicators of regional vulnerability to hydrometeorological disasters in 2021. The clustering methods used are hard clustering (single linkage, complete linkage, average linkage, Ward's method, and k-means) and soft clustering (Fuzzy C-Means). The optimal method for grouping regencies or cities in West Java Province is complete linkage, yielding 4 clusters. The result is that all clusters are vulnerable to social vulnerability.

Keywords: Clustering methods, rural destination, vulnerability, hydrometeorological disasters

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 intersection of the world's three major tectonic plates, which contributes to the country's high risk of natural disasters and vulnerability (BNPB, 2016). According to UNISDR[1] (2009), disasters are classified into four distinct categories; one of these is caused by hydrometeorological phenomena, also known as hydrometeorological disasters. Hydrometeorological disasters are natural phenomena in the atmosphere, hydrology, or oceanography 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 (UNISDR, 2009).

[1] UNISDR stands for United Nations International Strategy for Disaster Reduction

DOI: xxxxxx

The number of hydrometeorological disasters in Indonesia each year is increasing, as recorded by the National Agency for Disaster Management (BNPB).[2] According to BNPB's disaster data verification, the number of natural disasters in Indonesia in 2021 was 6,235. The total disasters were dominated by hydrometeorological events, accounting for 98.96% of cases, with the following breakdown: 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 estimated that 98.97% of disasters in 2021 were hydrometeorological, including 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 (BPS Jabar, 2022). West Java Province is at high risk of disasters due to the conditions described above.

Disaster risk assessment is an essential aspect of disaster mitigation, which attempts to reduce the impacts of natural disasters (Taghizadeh-Hesary et al., 2021; Siagian et al., 2017). 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. Additionally, Indonesia helped implement the Sendai Framework for Disaster Risk Reduction (SFDRR) 2015-2030 (BNPB & Bappenas, 2018). 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 (BNPB, 2021). 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 risk of harm to human life and property. Vulnerability is also defined as the expected losses from hazards and the extent to which society is unable to cope with disaster-related stress (Jeong & Yoon, 2018). According to Cutter (1996), vulnerability is classified into two types: social vulnerability and biophysical vulnerability, with their combination yielding place or regional vulnerability. Different loss patterns arise from disparities in social vulnerability across regions (Schmidtlein et al., 2011). Additionally, potential losses arise from society's interactions with biophysical conditions (Cutter, 1996). As community vulnerability increases, natural disasters will have a greater impact, and places with high biophysical vulnerability are more likely to suffer losses.

One widely used approach to measure social vulnerability is the index approach. One method often used is the Social Vulnerability Index (SoVI), developed by Cutter et al. (2003) and subsequently implemented in Indonesia by Siagian et al. (2014), Pangestu et al. (2021), and Wijaya et al. (2022). However, measuring vulnerability using an index remains problematic. 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 (Maharani et al., 2020). Another weakness is that it oversimplifies the relationships among the various

[2] The National Disaster Management Agency (BNPB) is a non-ministerial government institution responsible for coordinating disaster management at the national level in Indonesia. BNPB was established to ensure that disaster preparedness, emergency response, and post-disaster recovery are carried out in an integrated, coordinated, and effective manner across the country.

constituent indicators, making it difficult to detect the diversity of vulnerabilities (Rufat, 2013). To overcome these weaknesses, the cluster analysis method, pioneered by Rufat (2013), can be applied.

Research on regional vulnerability to hydrometeorological disasters can be conducted by grouping regions at the same vulnerability level using clustering. Clustering methods can be classified into two categories: hard clustering and soft clustering. Therefore, 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 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 to be helpful in planning and evaluating government program targets and policies, thereby enabling greater focus on the most vulnerable areas to natural disasters, particularly hydrometeorological disasters.

Methods

1. Research Data

The data sources used in this research are secondary data obtained from the Central Statistics Agency (BPS) of West Java Province, the West Java Open Data website, and the National Agency for Disaster Countermeasures (BPBD) of West Java Province. This research covers all regencies/cities in West Java Province, with 27 towns/regencies as the analysis units. The variables used in this research, along with 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	(Habibi, 2013); (Cutter et al., 2003); (Cutter & Finch, 2008); (Rabby et al., 2019); (Dintwa et al., 2019); (Wijaya & Halim, 2022); (Siagian et al., 2014).
Population density	West Java Province in 2022 Figures	(Wijaya & Halim, 2022); (Habibi, 2013); (Maharani et al., 2020); (Rabby et al., 2019); (Songwathana, 2018); (Irmayani et al., 2018); (Jeong & Yoon, 2018).
The number of people with disabilities	The West Java Open Data website	(Maharani et al., 2020); (Nugraha et al., 2022).
Sex Ratio	West Java Province in 2022 Figures	(Wijaya & Halim, 2022); (Cutter et al., 2003); (Cutter & Finch, 2008); (Siagian et al., 2014); (Zhou et al., 2014).
The percentage of the population aged 0 to 4 years	The West Java Open Data website	

The percentage of the population aged >65 years	The West Java Open Data website	(Maharani et al., 2020); (Cutter et al., 2003); (Cutter & Finch, 2008); (Siagian et al., 2014); (Wijaya & Halim, 2022); (Habibi, 2013).
Population growth rate	West Java Province in 2022 Figures	(Siagian et al., 2014); (Cutter et al., 2003); (Cutter & Finch, 2008); (Zhou et al., 2014).
The percentage of female heads of households	The West Java Open Data website	(Wijaya & Halim, 2022); (Cutter et al., 2003); (Cutter & Finch, 2008); (Siagian et al., 2014).
The open unemployement rate	West Java Province in 2023 Figures	(Wijaya & Halim, 2022); (Maharani et al., 2020); (Rabby et al., 2019); (Zhou et al., 2014).
Labor-force participation rate	West Java Province in 2024 Figures	
Gross enrollment at high school level	BPS Website	(Maharani et al., 2020); (Songwathana, 2018); (Irmayani et al., 2018); (Jeong & Yoon, 2018).
The number of health facilities	West Java Province in 2022 Figures	(Cutter et al., 2003); (Watung et al., 2018); (Djuraidah, 2009); (Maharani et al., 2020); (Nugraha et al., 2022).
The percentage of viable houses	Housing Statistics of West Java Province 2021	(Taghizadeh-Hesary et al., 2021); (Rabby et al., 2019) .
The number of villages/sub-districts that have an early warning system for natural disasters.	Potential Village Statistics of West Java Province 2021	(Dintwa et al., 2019); (Peraturan Kepala Badan Nasional Penanggulangan Bencana Nomor 02 Tahun 2012 Tentang Pedoman Umum Pengkajian Risiko Bencana, 2012); (Pangestu et al., 2021).
The number of villages/subdistricts with signage and evacuation routes	Potential Village Statistics of West Java Province 2021	
GDP at fixed costs per capita	BPS Website	(Taghizadeh-Hesary et al., 2021); (Irmayani et al., 2018); (Songwathana, 2018); (Cutter et al., 2003); (Cutter & Finch, 2008); (Schumacher & Strobl, 2011); (Barone & Mocetti, 2014).
Damaged house/residence	The National Disaster Relief Agency West Java Province	(Taghizadeh-Hesary et al., 2021).
Damaged facilities	The National Disaster Relief Agency West Java Province	

2. Hydrometeorological Disasters

UNISDR categorises disasters into four types: those caused by dynamic processes within the Earth, those caused by dynamic processes on the Earth's surface, those caused by hydrometeorological events, and those 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 (UNISDR, 2009). Extreme meteorological and climatic phenomena, such as floods, droughts, storms, tornadoes, or landslides, create hydrometeorological disasters (H. Wu et al., 2016).

3. 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 (Mosby et al., 2021; UNISDR, 2009). 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 in Equation (1).

$$Risk = Hazard \times \frac{Vulnerability}{Capacity} \quad (1)'$$

4. Framework, Cutter et al. (2003)

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 (Cutter, 1996).

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 (Cutter, 1996). 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 (Cutter et al., 2000).

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 (Jeong & Yoon, 2018). 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 (Cutter & Finch, 2008). Meanwhile, biophysical vulnerability can be defined as the exposure of human systems to extreme natural events (Clare & Weninger, 2011). 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 buildings, as well as land use and land cover (Cutter, 1996; Clare & Weninger, 2011).

5. 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 (Johnson & Wichern, 2007; Pramana et al., 2018). 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 (Ramadhan et al., 2021). Groupings of objects that are in the same cluster will be more similar than objects that are outside the cluster. 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.

6. Hard Clustering

In the hard clustering method, the membership of an observation unit is binary: whether it is included in the cluster or excluded (Pramana et al., 2018). Hard clustering is divided into two, namely hierarchical and non-hierarchical. The hierarchical grouping method is a clustering method that groups similar objects at adjacent levels and dissimilar objects at distant levels. 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 distance between an object in one cluster and an object in another cluster. Complete Linkage uses the maximum distance between any pair of objects, one from each cluster. Average Linkage uses the average distance between objects in one cluster and those in other clusters (Johnson & Wichern, 2007). 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 merged to yield the smallest SSE (Hair et al., 2009).

Meanwhile, the non-hierarchical grouping method partitions objects into k clusters, where the number of clusters is specified in advance or determined as part of the grouping procedure (Johnson & Wichern, 2007). 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) (Johnson & Wichern, 2007).

7. Soft Clustering

In the soft clustering method, the membership of an observation unit is expressed through the degree of membership in each cluster (Balasko et al., 2005). 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 (Hanniva et al., 2022). The FCM method is an improvement on the k-means algorithm, in which the membership of an observation unit is expressed as a degree of membership in each cluster, with values ranging from 0 to 1 (Balasko et al., 2005). In FCM, m is a fuzzification parameter that controls the degree of intersection between clusters (Pramana et al., 2018). In most data, a good fuzzifier value is between 1.5 to 3.0 (Bezdek et al., 1984). Wu (2012) recommends a value of m ranging from 1.5 to 4 for data containing noise and outliers.

Findings and Discussion

General description of indicators of regional vulnerability to the impact of hydrometeorological disasters in West Java Province

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 (Table 2) shows that the statistical significance of each indicator varies substantially.

Table 2. Descriptive Statistics of Regional Vulnerability Indicators

Indicators	Minimum	Maximum	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 unemployment 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

Comparing the most optimal cluster methods between hard clustering (hierarchical agglomerative and k-means) and soft clustering (FCM) based on indicators of regional vulnerability to the impact of hydrometeorological disasters in West Java Province in 2021
Table 2 shows that the 18 indicators used have different units, so standardizstion needs to be done first before carrying out the cluster analysis. The next step is multicollinearity checking. Because there is no multicollinearity among the indicators used in the study, cluster analysis can be conducted using hard and soft clustering methods based on Euclidean distance.

1. Hard Clustering

a. Non-Hierarchical

The first step in k-means clustering is to determine the number of clusters to be used (Johnson & Wichern, 2007). In this research, the methods used to determine the optimum number of clusters are the elbow and silhouette methods.

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. The following is a plot of cluster analysis using k-means with 4 clusters.

b. 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.

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.

2. Soft Clustering

a. 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. (1984), this study used a fuzzifier value between 1.5 to 3. 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
	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
3	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
	1.5	0.542	**0.390**	0.881	**0.903**	**0.606**
	2	0.250	0.000	1.386	636754.264	636754.264
4	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

Explanation:

The highest index value
The lowest index value

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. The following is a plot of cluster analysis using FCM with 4 clusters.

3. 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 (Han et al., 2022). Likewise, for the silhouette coefficient, a higher value indicates a better-formed cluster (Thamrin & Wijayanto, 2021). 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 Syafiyah et al. (2022) that the complete linkage method is the best method for grouping regencies/cities in West Java Province. *Analyzing regencies/cities in West Java Province based on indicators of regional vulnerability to the impact of hydrometeorological disasters in 2021*

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

Explanation:

Characteristics that are deemed susceptible

The results of clustering districts/cities in West Java using the complete linkage method produced 4 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. Besides, this cluster also has a high average 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 (BPBD Jawa Barat, 2021). 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. (2021), 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. In addition, this cluster has a relatively high average number of hydrometeorological disasters, with around 74 disasters in 2021, and the most frequent disaster is landslides (BPBD Jawa Barat, 2021), 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. Besides, this cluster has the highest average number of hydrometeorological disasters, with around 412 disasters in 2021. The most frequent disasters in the areas in this cluster are landslides.
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 with natural-disaster early-warning systems, and number of villages/subdistricts with signages and evacuation routes. Besides, this cluster has the lowest average number of hydrometeorological disasters, with around 23 disasters in 2021, with the most frequent disaster is tornadoes.

Conclusion

It can be concluded that hydrometeorological disasters predominate among natural disasters in West Java Province, and the distribution of each vulnerability indicator across regions is highly heterogeneous. 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. Grouping using the complete linkage method yielded 4 clusters.

1. Cluster 1 comprises 2 regencies/cities with the six most vulnerable indicators: 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. Cluster 1 is vulnerable to social vulnerability.
2. Cluster 2 comprises 17 regencies/cities with the two most vulnerable indicators: the percentage of people living in poverty and the percentage of female heads of households. Cluster 2 is vulnerable to social vulnerability.
3. Cluster 3 comprises 2 regencies/cities with the five most vulnerable indicators: the percentage of the population aged 0-4 years, population growth rate, Gross Enrollment Rate (APK) at the high school education level, damaged houses/residences, and damaged other facilities. Cluster 3 is also more vulnerable to social vulnerability characteristics than to biophysical vulnerability characteristics.
4. Cluster 4 comprises 6 regencies/cities with the five most vulnerable indicators: population density, Open Unemployment Rate (TPT), number of villages/subdistricts with natural-disaster early-warning systems, number of villages/subdistricts with signane and evacuation routes, and percentage of viable houses. Cluster 4 is also more vulnerable to social vulnerability characteristics than biophysical vulnerability characteristics.

Acknowledgments

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

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

Barone, G., & Mocetti, S. (2014). Natural disasters, growth and institutions: A tale of two earthquakes. *Journal of Urban Economics*, *84*, 52–66. https://doi.org/10.1016/j.jue.2014.09.002

Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The Fuzzy C-Means Clustering Algorithm. *Computers & Geosciences*, *10*(2–3), 191–203. https://doi.org/10.1109/igarss.1988.569600

BNPB. (2016). *Risiko bencana indonesia*.

BNPB. (2021). *IRBI Indeks Risiko Bencana Indonesia Tahun 2021*.

BNPB, & Bappenas. (2018). *RENCANA INDUK PENANGGULANGAN BENCANA 2015-2045*.

BPBD Jawa Barat. (2021). *Rekapitulasi Kejadian Bencana di Wilayah Provinsi Jawa Barat*.

BPS Jabar. (2022). *STATISTIK DAERAH PROVINSI JAWA BARAT 2022*.

Clare, L., & Weninger, B. P. (2011). Social and biophysical vulnerability of prehistoric societies to Rapid Climate Change. *Documenta Praehistorica*, *37*(1), 283–292. https://doi.org/10.4312/dp.37.24

Cutter, S. L. (1996). Vulnerability to hazards. *Progress in Human Geography*, *20*(4), 529–539.

Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). *Social Vulnerability to Environmental Hazards*. *84*(2).

Cutter, S. L., & Finch, C. (2008). Temporal and spatial changes in social vulnerability to natural hazards. *Planning for Climate Change: A Reader in Green Infrastructure and Sustainable Design for Resilient Cities*, *105*(7), 2301–2306. https://doi.org/10.4324/9781351201117-16

Cutter, S. L., Mitchell, J. T., & Scott, M. S. (2000). *Revealing the Vulnerability of People and Places : A Case Study of Georgetown County, South Carolina*. *90*(4), 713–737.

Dintwa, K. F., Letamo, G., & Navaneetham, K. (2019). Measuring social vulnerability to natural hazards at the district level in Botswana. *Jamba: Journal of Disaster Risk Studies*, *11*(1), 1–11. https://doi.org/10.4102/JAMBA.V11I1.447

Djuraidah, A. (2009). Indeks Kerentanan Sosial Ekonomi untuk Bencana Alam di Indonesia. *Prosiding Seminar Nasional Matematika Dan Pendidikan Matematika*.

Habibi, M. (2013). *MODEL SPASIAL KERENTANAN SOSIAL EKONOMI DAN KELEMBAGAAN TERHADAP BENCANA GUNUNG MERAPI (" Spatial Model of Social Economic and Institutional Vulnerability Of Merapi Disaster ") Pendahuluan Indonesia mempunyai karakteristik bencana yang kompleks , karena t*. *2*(1), 1–10.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate Data Analysis (7th ed.)*.

Han, J., Pei, J., & Tong, H. (2022). *Data Mining: Concepts and Techniques* (Fourth). Elsevier Science. https://books.google.co.id/books?hl=id&lr=&id=NR1oEAAAQBAJ&oi=fnd&pg=PP1&dq=Han,+J.,+%26+Kamber,+M.+(2006).+Data+Mining:+Concepts+and+Techniques,+Second+Edition.+San+Francisco:+Elsevier,+Inc&ots=_M9GNJsct-&sig=ksWO0XJnwK4adMvMPf8ImJnBiy4&redir_esc=y#v=one

Hanniva, Kurnia, A., Rahardiantoro, S., & Mattjik, A. A. (2022). Penggerombolan Kabupaten/Kota di Indonesia Berdasarkan Indikator Indeks Pembangunan Manusia Menggunakan Metode K-Means dan Fuzzy C-Means. *Xplore: Journal of Statistics*, *11*(1), 36–47. https://doi.org/10.29244/xplore.v11i1.855

Irmayani, S., Azhar, Z., & Adry, M. R. (2018). PENGARUH FAKTOR EKONOMI, SOSIAL EKONOMI DAN IKLIM TERHADAP BENCANA ALAM DI INDONESIA. *Photosynthetica*, *2*(1), 1–13. http://link.springer.com/10.1007/978-3-319-76887-8%0Ahttp://link.springer.com/10.1007/978-3-319-93594-2%0Ahttp://dx.doi.org/10.1016/B978-0-12-409517-5.00007-3%0Ahttp://dx.doi.org/10.1016/j.jff.2015.06.018%0Ahttp://dx.doi.org/10.1038/s41559-019-0877-3%0Aht

Jeong, S., & Yoon, D. K. (2018). Examining vulnerability factors to natural disasters with a spatial autoregressive model: The case of south Korea. *Sustainability (Switzerland)*, *10*(5), 1–13. https://doi.org/10.3390/su10051651

Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis (6th ed.)*.

Maharani, Y. N., Nugroho, A. R. B., Adiba, D. F., & Sulistiyowati, I. (2020). Pengaruh Kerentanan Sosial Terhadap Ketangguhan Masyarakat dalam Menghadapi Bencana Erupsi Gunung Merapi di Kabupaten Sleman. *Jurnal Dialog Penanggulangan Bencana*, *11*(1), 1–12. https://www.bnpb.go.id/jurnal/jurnal-dialog-penanggulangan-bencana-vol-11-no-1-tahun-2020

Mosby, K., Birch, T., Moles, A., & Cherry, K. E. (2021). Disasters. *Handbook of Rural Aging*, 111–115. https://doi.org/10.7591/9781501701498-008

Nugraha, A. L., Awaluddin, M., Sukmono, A., & Wakhidatus, N. (2022). Pemetaan Dan Penilaian Kerentanan Bencana Alam Di Kabupaten Jepara Berbasis Sistem Informasi Geografis. *Geoid*, *17*(2), 185. https://doi.org/10.12962/j24423998.v17i2.9370

Pangestu, H. D., Putra, A. D., & Syah, A. (2021). Analisis Indeks Risiko dan Potensi Kebencanaan (Studi untuk Wilayah Kabupaten Lampung Tengah). *Jurnal Rekayasa Sipil ...*, *9*(3), 481–490. http://repository.lppm.unila.ac.id/id/eprint/36857

Peraturan Kepala Badan Nasional Penanggulangan Bencana Nomor 02 Tahun 2012 Tentang Pedoman Umum Pengkajian Risiko Bencana, Peraturan Kepala Badan Nasional Penanggulangan Bencana 1 (2012). https://www.bnpb.go.id/uploads/24/peraturan-kepala/2012/perka-2-tahun-2012-tentang-pedoman-umum-pengkajian-resiko-bencana.pdf

Pramana, S., Yuniarto, B., Mariyah, S., Santoso, I., & Nooraeni, R. (2018). *Data mining dengan R Konsep Serta Implementasi*. IN MEDIA.

Rabby, Y. W., Hossain, B., & Hasan, M. U. (2019). Social vulnerability in the coastal region of Bangladesh: An investigation of social vulnerability index and scalar change effects. *International Journal of Disaster Risk Reduction*, 101329. https://doi.org/10.1016/j.ijdrr.2019.101329

Ramadhan, A., Prawita, K., Izzudin, M. A., & Amandha, G. (2021). Analisis strategi dan klasterisasi ketahanan pangan nasional dalam menghadapi pandemi covid-19. *Teknologi Pangan : Media Informasi Dan Komunikasi Ilmiah Teknologi Pertanian*, *12*(1), 110–122. https://doi.org/10.35891/tp.v12i1.2179

Rufat, S. (2013). Spectroscopy of Urban Vulnerability. *Annals of the Association of American Geographers*, *103*(3), 505–525. https://doi.org/10.1080/00045608.2012.702485

Schmidtlein, M. C., Shafer, J. M., Berry, M., & Cutter, S. L. (2011). Modeled earthquake losses and social vulnerability in Charleston , South Carolina. *Applied Geography*, *31*(1), 269–281. https://doi.org/10.1016/j.apgeog.2010.06.001

Schumacher, I., & Strobl, E. (2011). Economic development and losses due to natural

disasters: The role of hazard exposure. *Ecological Economics*, *72*, 97–105. https://doi.org/10.1016/j.ecolecon.2011.09.002
Siagian, T. H., & Parwanto, N. B. (2017). *Mengukur Risiko dan Kerentanan Bencana pada Skala Lokal di Indonesia melalui Downscaling World Risk Index*.
Siagian, T. H., Purhadi, P., Suhartono, S., & Ritonga, H. (2014). Social vulnerability to natural hazards in Indonesia: Driving factors and policy implications. *Natural Hazards*, *70*(2), 1603–1617. https://doi.org/10.1007/s11069-013-0888-3
Songwathana, K. (2018). The Relationship between Natural Disaster and Economic Development: A Panel Data Analysis. *Procedia Engineering*, *212*(2017), 1068–1074. https://doi.org/10.1016/j.proeng.2018.01.138
Syafiyah, U., Puspitasari, D. P., Asrafi, I., Wicaksono, B., & Sirait, F. M. (2022). Analisis Perbandingan Hierarchical dan Non-Hierarchical Clustering Pada Data Indikator Ketenagakerjaan di Jawa Barat Tahun 2020. *Seminar Nasional Official Statistics*, *2022*(1), 803–812. https://doi.org/10.34123/semnasoffstat.v2022i1.1221
Taghizadeh-Hesary, F., Sarker, T., Yoshino, N., Mortha, A., & Vo, X. V. (2021). Quality infrastructure and natural disaster resiliency: A panel analysis of Asia and the Pacific. *Economic Analysis and Policy*, *69*, 394–406. https://doi.org/10.1016/j.eap.2020.12.021
Thamrin, N., & Wijayanto, A. W. (2021). Comparison of Soft and Hard Clustering: A Case Study on Welfare Level in Cities on Java Island. *Indonesian Journal of Statistics and Its Applications*, *5*(1), 141–160. https://doi.org/10.29244/ijsa.v5i1p141-160
UNISDR. (2009). *United Nations International Strategy for Disaster Reduction (UNISDR)*.
Watung, C. H. T., Sela, R. L. E., & Tondobala, L. (2018). TINGKAT KETANGGUHAN DAN KETAHANAN KOTA MANADO TERHADAP BENCANA. *Jurnal Perencanaan Wilayah Dan Kota*, *5*. https://doi.org/10.5614/jpwk.2014.25.1.1
Wijaya, Y. T., & Halim, I. T. (2022). Measuring and Profiling Social Vulnerability to Natural Disaster in Indonesia in 2019. *Jurnal Matematika, Statistika Dan Komputasi*, *19*(1), 183–194. https://doi.org/10.20956/j.v19i1.21686
Wu, H., Huang, M., Tang, Q., Kirschbaum, D. B., & Ward, P. (2016). Hydrometeorological Hazards: Monitoring, Forecasting, Risk Assessment, and Socioeconomic Responses. *Advances in Meteorology*, *2016*, 11–14. https://doi.org/10.1155/2016/2367939
Wu, K.-L. (2012). Analysis of parameter selections for fuzzy c-means. *Pattern Recognition*, *45*(1), 407–415. https://doi.org/10.1016/j.patcog.2011.07.012
Zhou, Y., Li, N., Wu, W., Wu, J., & Shi, P. (2014). *Local Spatial and Temporal Factors Influencing Population and Societal Vulnerability to Natural Disasters*. *34*(4), 614–639. https://doi.org/10.1111/risa.12193