Parameter: Journal of Statistics
Available at https://bestjournal.untad.ac.id/index.php/parameter
Vol. 4 No. 2 2024, 83 - 92
DOI: *https://doi.org/10.22487/27765660.2024.v4.i2.17131*

Journal of Statistics
eISSN: 2776-5660

SOCIAL VULNERABILITY ANALYSIS IN CENTRAL JAVA WITH K-MEDOIDS ALGORITHM

Alwan Fadlurohman[1*], Nila Ayu Nur Roosyidah[2], Nafida Amalia Annisa[3]

[1,3]Program Studi S1 Sains Data, Universitas Muhammadiyah Semarang
[2]Badan Pusat Statistik (BPS) Kabupaten Poso, Sulawesi Tengah

[1]ORCID ID: 0009-0007-7263-6074

***e-mail**: alwan@unimus.ac.id

ABSTRACT

To address the limitations of the Social Vulnerability Index (SoVI) in only providing a general overview without pinpointing areas of social vulnerability, a correlational approach paired with a clustering method can be applied. This approach helps in identifying dominant factors and pinpointing socially vulnerable districts or cities in Central Java. The study employs the K-Medoids algorithm, which is advantageous when dealing with outliers in the dataset. Three different distance measures are considered: Euclidean, Manhattan, and Minkowski distances, to identify the optimal clustering of social vulnerability. The evaluation of the best cluster is conducted using the Davies-Bouldin Index, a metric for validating clustering models by averaging the similarity of each cluster to its most similar counterpart. Findings indicate that using the K-Medoids algorithm with Manhattan distance yields the most effective clustering, resulting in two distinct clusters. Cluster 1, comprising 25 districts/cities, is identified as the most vulnerable to natural disasters and challenges in education, demography, economy, and health. Meanwhile, Cluster 2, encompassing 10 districts/cities, includes urban areas with the highest social vulnerability, notably in the proportion of rental housing.

Keywords: *Clustering, Distance, DBI, Manhattan, Social Vulnerability*.

Cite: Fadlurohman, A., Roosyidah, N. A. N., & Annisa, N. A. (2024). *Social Vulnerability Analysis in Central Java with K-Medoids Algorithm. Parameter: Journal of Statistics, 4(2), 83-92,* https://doi.org/10.22487/27765660.2024.v4.i2.17131.

*Received: June 2024; Reviewed: June 2024; Published: Dec 2024

INTRODUCTION

Recent data indicates a marked uptick in natural calamities affecting Indonesia. According to the latest report from the country's National Disaster Management Agency (BNPB), the nation witnessed a total of 5,400 disaster events throughout 2023, spanning diverse geographical areas (BNPB, 2024). This substantial figure underscores the escalating frequency of natural hazards impacting the archipelagic state. This figure marks a substantial rise from the previous year, 2022, which recorded only 3,544 incidents. Central Java stands out as one of the provinces with the highest disaster occurrences, totaling 629 events, accounting for 11.6% of all incidents in 2023. The predominant types of disasters in this province include forest and land fires, landslides, extreme weather, and flooding (BNPB, 2024). Among these, the El Nino phenomenon, characterized by rising temperatures in the Pacific Ocean from May to August 2023, has contributed to some of the disasters (BNPB, 2024). Furthermore, Indonesia's tropical climate makes it susceptible to extreme weather, often resulting in heavy rainfall that leads to flash floods and landslides.

In light of the various threats posed by natural disasters, it is crucial for central, provincial, and local governments to ensure effective preparedness. To aid in this, it is necessary to categorize districts and cities based on indicators of social vulnerability, which can help in formulating appropriate mitigation strategies to lessen the adverse impacts of disasters (UNISDR, 2023). Vulnerability refers to the potential losses that can occur due to natural disasters (Cutter, 2024). It can also be understood as the characteristics and capacity of a community to anticipate and respond to the consequences of such disasters (Kuran et al., 2020). Social vulnerability can be conceptualized as the aggregate of societal resources within a given area that facilitate effective management of potential adverse effects resulting from natural calamities.

Numerous scholars have investigated the assessment of social vulnerability, including studies on environmental susceptibility to catastrophic events (Cutter, 2024). Furthermore, the Social Vulnerability Index (SoVI) has been established by combining socioeconomic and demographic variables through the application of principal component analysis (PCA) methodology (Cutter et al., 2003). In Indonesia, SoVI assessments have been conducted to gauge local tsunami preparedness in Padang (Birkman et al., 2008). However, these studies exhibit some limitations, such as a lack of attention to the geographical dimensions of disaster occurrences (Nasution et al., 2020). Furthermore, previous research primarily focused on assessing area vulnerability without delving deeply into the impacts of natural disasters. Understanding vulnerability impacts necessitates uncovering the underlying reasons why certain regions are susceptible to specific disasters (Rufat, 2014). Consequently, clustering analysis can be employed to evaluate social vulnerability within an area (Rufat, 2014). The analysis methods previously utilized include Hierarchical Clustering (Rufat, 2014) and Fuzzy Geographical Weighted Clustering (FGWC) utilizing an intelligent firefly algorithm (Nasution et al., 2020).

Preliminary assessments of societal vulnerability in Central Java Province revealed anomalies in various categories. This research employs cluster analysis with the K-Medoids method, recognised for its efficacy in managing outlier data (Fadlurohman & Nur, 2023). K-Medoids is a partitioning clustering technique that categorises n items into k clusters, seeking to identify k clusters that most accurately describe the dataset (Wira et al., 2019). This method arranges objects based on their nearness to the cluster centroid to form new clusters (Aurora et al., 2016).

In K-Medoids analysis, objects are grouped based on their similarity, to measure the level of similarity, a distance measure is applied. The greater the distance value obtained, the further the object is located from the cluster centre formed. This study aims to apply distance measurements utilising Euclidean, Manhattan, and Minkowski methods within the K-Medoids framework. Another thing that needs to be considered in cluster analysis is the validation of clustering results. Validation of clustering results is done to obtain the partition that best fits the data. If not validated, it will affect the analysis results. In this study, two validations were used with the internal criteria approach, i.e. silhouette validation and Elbow index. The Davies-Bouldin Index will be employed to determine the most successful clustering approach among these three distance measurements.

Recent research has shown that K-Medoids methods can effectively cluster data using a variety of measurement distances, with each providing its own advantages in different application contexts. Comparative studies have found that Manhattan distance is often more robust to outliers than Euclidean, while Minkowski offers flexibility that can be tailored to the characteristics of the data. The state of the art in this research is an innovation in the K-Medoids algorithm using three distance measures, hopefully improving the performance of K-Medoids on social vulnerability analysis in Central Java.

This study is organised into several primary areas. The introductory part delineates the motivations for the research. The methodology section subsequently delineates the research scope and methodological structure. The third section of the paper delineates the study's aims, principally concentrating on determining the most efficacious clustering technique and doing a thorough assessment of each resultant cluster. This section provides a comprehensive study of the characteristics of the generated clusters and examines their potential policy consequences. The conclusion section integrates the research findings, providing final observations on the study's results and importance.

MATERIALS AND METHODS

This segment is divided into two distinct subsections: the initial part focuses on the materials, while the latter addresses the methodological approach. The materials subsection offers detailed information regarding the utilized dataset and its origins. In the methods subsection, we elucidate the K-Medoids algorithm and delineate the procedural steps for implementing this algorithm using a trio of distinct distance measurement techniques.

Materials

This study employs the K-Medoids algorithm to categorise districts and cities in Central Java Province according to socioeconomic vulnerability indicators, utilising three distinct distance metrics: Euclidean, Manhattan, and Minkowski. The study's variables are derived from "Revisiting Social Vulnerability Analysis in Indonesia Data" (Kurniawan et al., 2022), with modifications to correspond with current data availability. The researchers incorporated the unemployment rate as a macroeconomic variable, alongside socioeconomic and demographic factors, to classify districts and cities in Central Java. Information was obtained from the BPS website (www.bps.go.id), which aggregates data from Indonesia's periodic surveys, including the National Socio-Economic Survey (SUSENAS) and the National Labour Force Survey (SAKERNAS). This study examines eight particular variables using recent data. The study encompasses all districts and municipalities in Central Java, use data from 2023. Table 1 presents a comprehensive review of the variables utilised in this investigation.

Tabel 1. Social vulnerability variables

Variable	Description	Source
Children	percentage of the population under the age of fifteen	SUSENAS, Statistics Indonesia
Elderly	percentage of the population aged 64 and older	SUSENAS, Statistics Indonesia
Lowedu	The proportion of people aged 15 and older who lack formal education	SUSENAS, Statistics Indonesia
Rented	The proportion of households that rent a home	SUSENAS, Statistics Indonesia
Unemployment	the percentage of the labor force that is without a job but actively seeking employment and willing to work	SAKERNAS, Statistics Indonesia
Water	The proportion of households lacking access to adequate drinking water	SUSENAS, Statistics Indonesia
Poverty	The percentage of households that do not have access to clean drinking water The proportion of people living below the poverty line	SUSENAS, Statistics Indonesia
PopGrowth	lation change as a percentage	Indonesia Population Projection based on 2020 Population Census

Methods

The K-Medoids algorithm represents an adaptation of the K-Means clustering technique. In contrast to K-Means, which relies on cluster averages, K-Medoids employs actual data points, referred to as medoids, for cluster representation (Farissa et al., 2021). This approach helps reduce the partitioning process's susceptibility to dataset outliers (Vercillis, 2009). The primary aim of K-Medoids clustering is to address the shortcomings of K-Means, particularly its vulnerability to extreme values that can skew data distribution (Han & Kamber, 2017). Functioning as a partitioning method, K-Medoids clustering organizes n objects into k clusters, with representative objects in each cluster designated as medoids (Setyawati, 2017). The procedural steps for implementing the K-Medoids algorithm in the clustering process are delineated by Wira et al. (2019) as follows:

1) Perform a descriptive analysis to summarize and characterize the dataset.
2) Create boxplots to identify outliers in each variable.
3) Verify clustering assumptions by conducting KMO and Multicollinearity tests.
4) Establish the number of clusters, k, based on both theoretical and conceptual frameworks. In this study, we examine k values ranging from 2 to 10 and determine the optimal k using the Silhouette and Elbow methods.
5) Compute the proximity of each data point to its closest cluster center, employing and contrasting three distinct distance measurement techniques: the Euclidean metric, Manhattan metric, and Minkowski metric.

a) Euclidean Distance

The Euclidean metric is a technique for measuring spatial distance, assessing the gap between two points in Euclidean space, which can be two-dimensional, three-dimensional, or multi-dimensional. To assess how similar data points are, the Euclidean distance is computed using a particular mathematical formula. (Sunge et al., 202).

$$d_{(x,y)} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (1)$$

where:

$d_{(x,y)}$: the distance between x and y
n : total number of data points
xi : value at the centroid of the i-th cluster
yi : value for each i-th data point

b) Manhattan Distance

The Manhattan distance quantifies the sum of the absolute differences between the corresponding coordinates of two places. This metric utilises a particular mathematical formula, expressed as:

$$d_{(x,y)} = \sum_{i=1}^{n}|x_i - y_i| \qquad (2)$$

where:
$d_{(x,y)}$: the distance between x and y
n : quantity of data elements
xi : value at the i-th cluster's central point
yi : value of each i-th data element

c) Minkowski Distance

The Minkowski metric is employed in normed vector spaces and functions as a generalized form of both Euclidean and Manhattan distances. In the application of Minkowski distance for measuring object separation, the parameter p is commonly assigned a value of 1 or 2. To determine the distance using this approach, a specific mathematical expression is applied.

$$d_{(x,y)} = \left(\sum_{i=1}^{n}|x_i - y_i|^p\right)^{1/p} \qquad (3)$$

where:
$d_{(x,y)}$: represents the distance between x and y
xi : the value at the center of the i-th cluster
yi : the value for each i-th data point
p : the exponent or power

6) Select new cluster centres at random to serve as candidate non-medoids for each object.
7) Calculate the distance from each object in the cluster to the proposed non-medoids.
8) Assess the difference (S) by deducting the previous distance from the current distance. If S is negative, substitute the item with the non-medoids cluster data to form a new collection of k objects referred to as medoids.
9) Proceed by reiterating stages 3 to 5 until the medoids reach stability, culminating in the establishment of clusters and their respective members.

10) We use the silhouette and elbow approaches to get the best number of clusters. A thing's silhouette score is based on how far away it is on average from other things in its cluster. The elbow technique, on the other hand, requires you to compute the sum of squares inside each cluster. To further evaluate the clustering findings, we use the Davies-Bouldin Index.

RESULTS AND DISCUSSION

In this section, the discussion is organized into several subsections. Subsection 1 focuses on identifying outliers in the dataset. Subsection 2 examines assumptions related to clustering, including the KMO test and VIF-based multicollinearity test. Subsection 3 outlines the process for determining the number of clusters, while subsection 4 presents the results from our clustering comparisons. Finally, subsection 5 interprets the findings from the most effective clustering method.

The proportion for the population under the age of fifteen falls between 17.59% and 24.50%, with a median of 21.26% and a mean of 21.37%, suggesting a fairly even distribution with a little positive skew. With a range of 6.46–16.50%, a median of 10.06%, and an average of 9.96% for the senior population, we can see that the majority of numbers are lower, with a small number of higher numbers somewhat lowering the average. There is a right-skewed distribution with uncommon higher values for the proportion of persons with poor education, which spans from 18.00% to 61.11% and has a median of 43.56% and a mean of 41.18%. A right-skewed distribution with occasional higher values is suggested by the wide range of values for access to water, which varies dramatically from 0.000 to 17.82%. The median value is 4.080 and the mean is 5.34%. The poverty rates range from 4.23% to 16.34%, with a median of 9.81% and an average of 10.40%. This distribution is generally symmetrical with a small right-hand bias. The right-skewed distribution of the population growth rates is shown by a range of 0.19 to 1.43, a median of 1.04, and an average of 0.97. A right-skewed distribution impacted by a few higher values indicates that the fraction of leased homes runs from 0.00% to 12.68%, with a median of 1.08% and a mean of 1.90%. There is a little right skew in the distribution of unemployment rates, which vary from 1.92% to 8.98% and have a median of 4.57% and an average that is somewhat higher at 4.86%.

Our goal is to find the optimal cluster size by combining the silhouette and elbow techniques. While the elbow approach uses the within-cluster sum of squares, the silhouette coefficient finds the average distance of one item to others in its cluster. We will also assess the clustering results using the Davies-Bouldin Index.

Detection of Outliers

In this section, we will examine the dataset for outliers, acknowledging that the K-Medoids method is particularly robust when handling outlier data. We employed boxplots to identify outliers, as illustrated in Figure 1. Data points that fall outside the boxplot are considered outliers. Figure 1 reveals that there are five variables with outlier data: For the Elderly variable, there is one outlier from Wonogiri. The Water variable has three outliers from Purbalingga, Karanganyar, and Jepara. PopGrowth includes three outliers from Wonogiri, Magelang City, and Surakarta City. The Rented variable has seven outliers from Sukoharjo, Magelang City, Surakarta City, Salatiga City, Semarang City, Pekalongan City, and Tegal City. Unemployment has one outlier from Brebes. These data points deviate from the boxplot, making the K-Medoids method appropriate for analyzing these datasets.

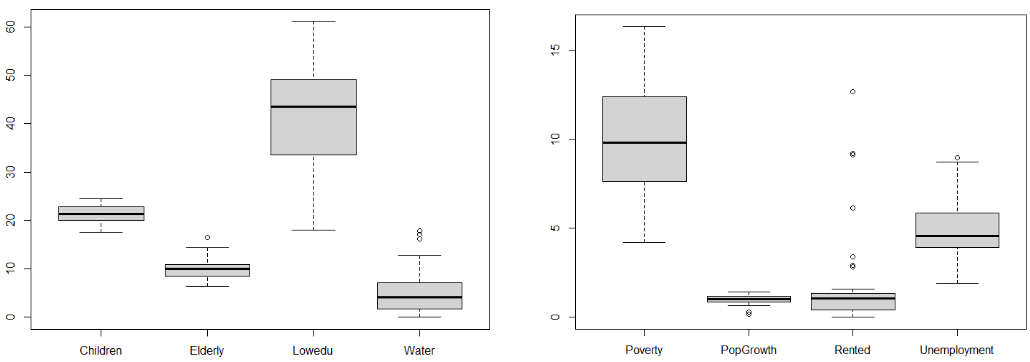


Figure 1. Outlier detection using boxplot

Assumptions of Cluster Analysis

To conduct cluster analysis, it is essential to meet certain assumptions, such as ensuring a representative sample through the KMO test and checking for multicollinearity using the VIF test. A representative sample is one that accurately reflects the overall population. The KMO (Kaiser-Meyer-Olkin) test can be used to determine if a sample is representative by assessing sample adequacy. A KMO value between 0.5 and 1 indicates a representative sample (Ningrat et al., 2016). Additionally, if the VIF value is less than 10, it suggests that multicollinearity is not present. The outcomes of the KMO and VIF tests are detailed in Table 2.

Tabel 2. VIF and KMO test

Variable	VIF	KMO	
Children	7,11	0,69	Overall = 0,73
Elderly	2,65	0,66	
Lowedu	4,78	0,76	
Water	1,53	0,86	
Poverty	2,68	0,80	
PopGrowth	3,99	0,72	
Rented	4,07	0,72	
Unemployment	1,91	0,68	

Table 2 shows that there is no multicollinearity among the variables as their VIF values are less than 10 (VIF < 10). It may be inferred from this that there is little correlation between the variables. In addition, the data or sample used seems to be representative, as the KMO value is 0.73. So, it's safe to say that clustering analysis works well with the social vulnerability data.

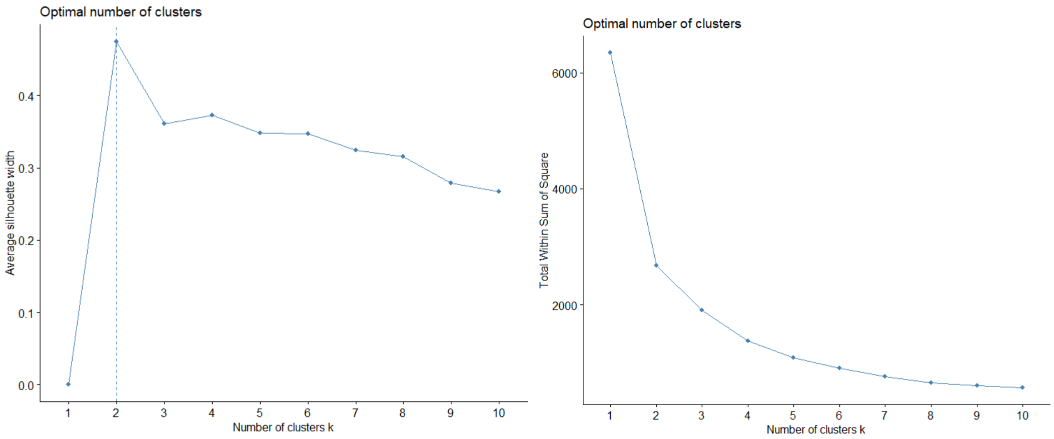
Determination the Optimum Number of Cluster

Utilising the silhouette and elbow techniques, we determine the optimal number of clusters. The silhouette technique evaluates the quality of the clusters created by counting the number of clusters and then calculating the average distance between items in the same cluster and their closest neighbouring cluster (Anggara et al., 2016). The quality of a cluster is improved with an increase in the average value. On the other hand, the elbow technique looks at the total of squares inside each cluster, which becomes smaller as the number of clusters gets more. When the within-cluster sum of squares drops significantly, that's when you know you've found the sweet spot for cluster counts. Figure 2 shows the results visually, while Table 3 presents the results from the silhouette and elbow approaches.

Tabel 3. Determination of the number of clusters with silhouette and elbow methods

Cluster	Silhouette	Elbow
1	0,00	6340,26
2	**0,47**	**2673,08**
3	0,36	1899,51
4	0,37	1371,01
5	0,35	1086,92
6	0,35	901.85
7	0,32	764.35
8	0,31	652,69
9	0,28	602,72
10	0,27	563,58

According to Table 3, the silhouette approach reveals that the optimal number of clusters (k) is 2, yielding an average score of 0.47. As the within-cluster sum of squares decreases the most from k = 1 to k = 2, the elbow technique also suggests that k = 2 is the ideal number of clusters. For this reason, it is clear that the social vulnerability data should be input into the K-Medoids algorithm using two clusters.



(a) (b)

Figure 2. Graph of determining the number of clusters: (a) silhouette method and (b) elbow method.

Comparison of Cluster Results

Using the K-Medoids technique and the three distinct distance metrics—Euclidean, Manhattan, and Minkowski—we ran a cluster analysis. Using these distance metrics, we applied the Davies-Bouldin Index (DBI) to the K-Medoids algorithm's clustering output. In most cases, better clustering is indicated by a lower DBI score. Table 4 displays the DBI values for every clustering result.

Tabel 4. Evaluation of cluster result with Davies Bouldin Index (DBI)

Distance	Davies Bouldin Index (DBI)
Euclidean	0,795
Manhattan	**0,750**
Minkowski (p=1)	**0,750**
Minkowski (p=2)	0,795
Minkowski (p=3)	0,795

The Davies-Bouldin Index, as shown in Table 4, is a product of the clustering outcomes obtained by the K-Medoids algorithm via the use of five distinct distance measuring techniques. At $p = 2$, the DBI value is clearly the same for both the Minkowski and the Euclidean distances. This happens because, as can be shown in equation 3, the Minkowski distance formula with $p = 2$ is identical to the formula for the distance in geometry. In a similar vein, the Manhattan distance formula and the Minkowski distance both provide the same DBI value when $p = 1$. Thus, the Manhattan distance or the Minkowski distance with $p = 1$ provide the ideal clustering result for social vulnerability data in Central Java using the K-Medoids method, according to the DBI value, as they both generate the lowest DBI value.

Interpretation of the Best Cluster Results

The optimal clustering results for social vulnerability data in Central Java were obtained using the K-Medoids algorithm with the Manhattan distance metric and two clusters. Figure 3 illustrates the distribution of districts and cities in Central Java based on this clustering approach. The distinct characteristics of each cluster are detailed in Table 5.

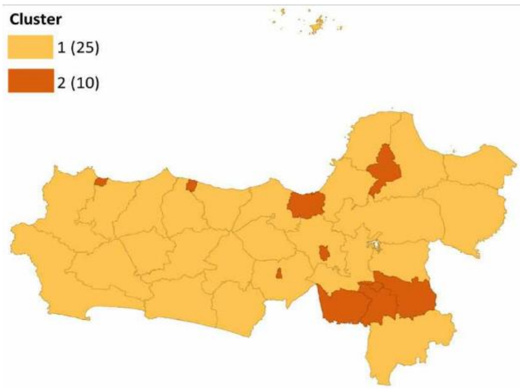


Figure 3. Visualisation of clustering results using the K-Medoids algorithm with Manhattan distance.

Tabel 5. Characteristics of each cluster (mean)

Variable	Cluster 1	Cluster 2
Children	21,8	20,2
Elderly	10,4	8,93
Lowedu	47,2	26,1
Water	7,08	1,00
Poverty	11,6	7,48
PopGrowth	1,02	0,87
Rented	0,64	5,04
Unemployment	4,94	4,67

Roughly 71% of Central Java's administrative entities are located in Cluster 1, which consists of 25 districts and municipalities. Natural catastrophes pose the greatest threat to this cluster in terms of their potential influence on health, education, demography, and the economy. Within the field of education, the cluster is characterized by a high average percentage of residents with low educational attainment. Demographically, Cluster 1 has a high dependency ratio, as indicated by the substantial percentages of elderly individuals (65 years and older) and children under 15. Furthermore, this area has a relatively high average population growth rate in comparison to other clusters. This rapid increase in population suggests a rising number of children, which may heighten security concerns.

As noted by Cutter et al. (2003) and Rufat (2014), a substantial proportion of children and elderly residents can intensify the responsibilities placed on the working-age population during disaster response and recovery efforts. From a macroeconomic perspective, regions in cluster 1 display the highest average poverty rate. Interpreting poverty as limited access (Siagian et al., 2014), this elevated poverty rate suggests that areas in cluster 1 face heightened vulnerability during disasters. Impoverished populations often lack the means to acquire disaster preparedness equipment or establish emergency funds, potentially prolonging their recovery period following a disaster.

Cluster 1 has the lowest average proportion of families having access to appropriate drinking water sources, which is a major concern for health indicators. Having access to safe drinking water is essential for good health. It may be especially important to pay close attention to low-lying regions in the case of a natural catastrophe so that people there don't have even more trouble getting clean water if the water delivery system is damaged. Governments at all levels should pay more attention to the regions included in cluster 1. Because most of these areas are located on the southern shore, which is vulnerable to natural calamities like earthquakes and tsunamis due to tectonic plate subduction, this is of the utmost importance. Building strong evacuation infrastructure, installing tsunami early warning systems, and educating people thoroughly on how to be prepared for disasters are all important parts of a comprehensive mitigation strategy.

Cluster 2 comprises 10 districts/municipalities, accounting for 29% of Central Java's administrative units, primarily urban areas characterized by a high proportion of rented housing. The dramatic rise in property values, coupled with increased housing demand due to urbanization in Indonesia's major cities, has led low-income individuals to favor rental accommodations (Berawi et al., 2019). The elevated open unemployment rates in these areas indicate intense job competition. Urban employment opportunities, predominantly in industrial and service sectors, typically demand advanced skills and expertise. This, combined with high population density, results in labor market inefficiencies and inadequate job absorption (Mankiw, 2010).

To mitigate vulnerabilities in Cluster 2, suggested solutions involve pre-employment training programs and enhanced access to business capital. As shown in Table 5, individuals in Cluster 2 typically have higher education levels, indicating that these strategies could provide more sustainable solutions to their vulnerabilities. To tackle housing ownership challenges, the government could expand its current initiatives, such as affordable apartment rental programs in low-income areas.

CONCLUSION

The K-Medoids algorithm with Manhattan distance emerged as the optimal clustering method for grouping districts and cities in Central Java, as determined by the Davies-Bouldin Index. The Silhouette and Elbow criteria indicated that two clusters were ideal. Cluster 1 encompasses 25 districts/municipalities, representing approximately 71% of Central Java's administrative units. Cluster 2 comprises urban areas characterized by high social vulnerability, particularly in terms of rental housing prevalence. This cluster includes 10 districts/municipalities, accounting for 29% of the region's administrative divisions.

Each district in Central Java faces unique challenges, as evidenced by the distinct characteristics of the formed clusters. Policy interventions should be tailored to address these specific vulnerabilities. The most prevalent issues across the regions are poverty and low educational attainment. Consequently, the government should prioritize and enhance job training initiatives and financial inclusion programs to mitigate community vulnerability. Additionally, implementing comprehensive disaster education is crucial to equip residents with effective strategies for disaster preparedness and response.

REFERENCES

Anggara, M., Sujiani, H., & Nasution, H. (2016). Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokkan Member di Alvaro Fitness. *Jurnal Sistem dan Teknologi Informasi (JUSTIN), 1*(1), 1-6.

Aurora, P., Deepali, & Varshney, S. (2016). Analysis of K-Means and K-Medoids Algorithm for Big Data. *Procedia Computer Sains*, 78: 7-12.

Berawi, M.A., Aulidtyo, A., Miraj, P., & Gunawan. Developing affordable housing for low-income households in developing countries using participative payment. I*OP Conference Series: Materials Science and Engineering*. DOI: 10.1088/1757-899X/650/1/012013.

Birkmann, J., Setiadi, N.J., & Gebert, N. (2008). *Socio-economic Vulnerability at The Local Level in Context of Tsunamy Early War Evacuation Planning in The City of Padang, West Sumatera*. Proceedings of the International Conference on Tsunami Warning (ICTW). Bali, Indonesia, 1-8.

BNPB, PDSI Pusdatinkom. (2024). *Buku Data Bencana Tahun 2023*. Jakarta, Indonesia: Pusdatinkom BNBP.

Cutter, S.L. (2024). The origin and diffusion of the social vulnerability index (SoVI). *International Journal of Disaster Risk Reduction*, *109*, 104576. DOI: https://doi.org/10.1016/j.ijdrr.2024.104576

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

Farissa, R.A., Mayasari, R., & Umaidah, Y. (2021). Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Data Obat dengan Sillhouette Coefficient. *Journal of Applied Informatics and Computing (JAIC), 5*(2), 109-116.

Han, J. & Kamber, M. (2017). *Data Mining, Concept and Techniques*. Waltham: Morgan Kauffman Publisher.

Kuran, C. H. A., Morsut, C., Kruke, B. I., Krüger, M., Segnestam, L., Orru, K., Naevestad, T. O., Airola, M., Keränen, J., Gabel, F., Hansson, S., & Torpan, S. (2020). Vulnerability and vulnerable groups from an intersectionality perspective. *International Journal of Disaster Risk Reduction*, *50*. DOI: https://doi.org/10.1016/j.ijdrr.2020.101826

Kurniawan, R., Nasution, B.I., Agustina, N., & Yuniarto, B. (2022). Revisiting Social Vulnerability Analysis in Indonesia Data. *Data in* Brief; *40*, 1-7. DOI: https://doi.or/10.1016/j.dib.2021.107743

Mankiw, G.N. (2010). *Macroeconomics*. Seventh Edition. New York: Worth Publishers.

Nasution, B.I., Kurniawan, R., Siagian, T.H., & Fudholi, A. (2020). Revisiting Social Vulnerability

Analysis in Indonesia: An Optimized Spatial Fuzzy Clustering Approach. *International Journal of Disaster Risk Reduction*, 51: 1-11. DOI: https://doi.org/10.1016/j.ijdrr.2020.101801

Ningrat, F.R., Maruddani, D.A.I., & Wuryandari, T. (2016). Analisis Cluster dengan Algoritma K-Means dan Fuzzy C-Means Clustering untuk Pengelompokkan Data Obligasi Korporas. *Jurnal Gaussian, 5*(4), 641-650.

Perumahan dan Kesehatan Lingkungan Menggunakan Metode KMedoids. In Prosiding Seminar Nasional Unimus (Vol. 6).

Rufat, S. (2014). Spectroscopy of Urban Vulnerability, *Annals of the Association of American Geographers*, *103*(3): 505-525. DOI: https://doi.or/10.1080/00045608.2012.702485

Setyawati, A. W. (2017). Impelmentasi Algoritma Partitioning Around Medoid (PAM) untuk Pengelompokkan Sekolah Menengah Atas di DIY Berdasarkan Nilai Daya Serap ujian Nasional. Skripsi. Yogyakarta: Fakultas Sains dan Teknologi, Universitas Sanata Darma.

Siagian, T.H., Purhadi, P., Suhartono, S., & Ritonga, H. (2014). Social vulnerability to natural hazards in Indonesia: driving factors and policy implications, Nat. Hazards. *Journal of the International Society for the Prevention and Mitigation of Natural Hazards*, *70*(2): 1603-1617. DOI: https://doi.org/10.1007/s11069-013-0888-3

Social Science Quarterly. 2003; 84(2): 242–261. DOI : https://doi.or/10.1111/1540-6237.8402002 Fadlurohman, A., & Nur, I. M. (2023). Pengelompokkan Provinsi di Indonesia Berdasarkan Indikator

Sunge, A.S., Heryadi, Y., Religia, Y., & Lukas. (2020). Comparison of Distance Function to Performance of K-Medoids Algorithm for Clustering. *International Conference on Smart Technology and Applications (ICoSTA), Surabaya, Indonesia*, 1-6. DOI: 10.1109/ICoSTA48221.2020.1570615793.

UNISDR. (2023). Implementation of the Hyogo Framework for Action: Summary of Reports 2007-2013. United Nations Secretariat of The International Strategy of Disaster Reduction (UNISDR). Geneva: Switzerland.

Vercillis, C. (2009) *Business Intelligence, Data Mining and Optimization for Decision making*. Milan: Wiley.

Wira B, Budianto AE, Wiguna AS. Impelmentasi Metode K-Medoids Clustering Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 di Universitas Kanjuruhan Malang. *Jurnal Terapan Sains & Teknologi*. 2019; 1(3): 54-69.