JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 7, Number 1, March 2026, Page 26-34
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v7i1.1469

Clustering of Provincial Health Vulnerability Levels in Indonesia Using the K-Means Method

Okma Arnilia[1*], Sahrial Ihsani Ishak[2], Tri Widodo[3], I Gusti Nyoman Agung Bisma Tatwa[4]
[1] Program Studi Informatika, Universitas Islam Negeri Siber Syekh Nurjati Cirebon, Indonesia
[2] Program Studi Teknik Informatika, Universitas Dian Nusantara, Indonesia
[3] Program Studi Teknik Komputer, Universitas Teknokrat Indonesia, Indonesia
[4] Program Studi Ilmu Komputer, Institut Pertanian Bogor, Indonesia

Abstract: Indonesia faces significant regional health disparities due to socio-economic, environmental, and infrastructural variations, complicating equitable resource allocation. To address this, this study aims to classify the health vulnerability of 38 Indonesian provinces using 2024 socio-economic and health indicators. These include the number of hospitals, access to adequate sanitation and safe drinking water, stunting prevalence, number of health facilities, population size, and poverty percentage. Data normalization was performed using the z-score method to standardize variable scales and prevent indicator dominance. By evaluating inertia decreases via the Elbow method, K-Means clustering optimally grouped the provinces into four clusters (K=4). The results were spatially visualized, revealing clear geographical variations: Cluster 1 represents provinces with very good health conditions, Cluster 2 good, Cluster 3 moderate, and Cluster 4 identifies provinces requiring urgent attention. This mapping provides a comprehensive overview of health vulnerability distribution. The findings offer crucial insights for policymakers and stakeholders to prioritize targeted, region-based health interventions, optimize resource allocation, and promote equitable national health development strategies.
Keywords: Health Vulnerability, Clustering, K-Means, Health Indicators, Indonesia

1. INTRODUCING

Public health conditions in Indonesia exhibit distinct regional patterns due to variations in socio-economic status, environmental conditions, and health infrastructure. Data from the Indonesia Health Profile 2023 reveal significant disparities, for example in the Maternal Mortality Ratio (MMR), which ranges from 45 to 250 per 100,000 live births, and the Infant Mortality Rate (IMR), which varies between 10 and 40 per 1,000 live births across provinces [1]. Similar disparities are observed in stunting prevalence, where provinces such as East Nusa Tenggara (NTT) (35.1%) and Papua (34.6%) still record high rates compared to Bali (8.0%) and DKI Jakarta (14.8%), which show relatively low prevalence [2]. Access to improved sanitation also varies widely, with provincial coverage ranging from 60% to 98%, underscoring that equitable health development remains a serious challenge in Indonesia [1].

These differences across indicators reflect disparities in health service capacity and overall quality of life among provinces, indicating the need for more comprehensive data-driven analytical approaches. Health vulnerability refers to conditions in which individuals or communities face a high risk of health disturbances due to various factors such as exposure to hazards, sensitivity, and adaptive capacity, as defined in the vulnerability assessment literature [3], [4]. High levels of health vulnerability can have serious implications for human development, exacerbate social inequality, and increase the economic and social burden of a region [5]. Therefore, mapping health vulnerability across regions is an important

Okma Arnilia: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 7, Number 1, March 2026, Page 26-34
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v7i1.1469

step to systematically assess which areas or populations are most at risk, thereby providing a basis for development prioritization, health system strengthening, and preparedness for potential health crises [4], [6].

In this context, data-driven analysis capable of grouping regions with similar vulnerability characteristics is required. Such grouping enables a comprehensive understanding of regional disparities and supports more targeted health program planning. K-Means clustering is an unsupervised learning technique widely used to group data based on attribute similarity. The effectiveness and popularity of K-Means have been extensively discussed in the literature; Jain (2010) notes that K-Means remains one of the most influential clustering algorithms over the past 50 years [7]. Furthermore, Han, Pei, and Kamber (2012) emphasize that K-Means is well suited for data with a large number of variables and complex structures [8]. Research by Kodinariya and Makwana (2013) also shows that K-Means is frequently applied in vulnerability mapping and regional clustering due to its ability to provide objective segmentation [9]. Given that the number of provinces in Indonesia is relatively limited while the complexity of health indicators is high, K-Means is an appropriate method for this study. The clustering results can help identify provinces with low, moderate, and high levels of vulnerability in a more objective manner, thereby providing a foundation for more focused health policy formulation.

Several previous studies have applied K-Means clustering to health and healthcare facility data in Indonesia. For example, a study in Bandung City used K-Means to map the distribution of health facilities and healthcare workers, producing clusters categorized as low, medium, and high [10]. Another study successfully clustered populations based on their health complaints using data mining techniques [11]. Furthermore, K-Means has also been implemented to optimize healthcare budgets specifically for non-communicable diseases (NCDs) [12]. At the facility level, a study in Jakarta compared K-Means and Fuzzy C-Means for clustering hospitals and found that both methods produced meaningful clusters with different distributions [13]. Finally, a study at the provincial level in East Java used K-Means to cluster regions based on disease profiles, demonstrating how clustering can help identify priority areas for health interventions [14].

However, despite the extensive use of K-Means in the healthcare sector, most of these studies have focused on isolated variables—such as specific diseases, localized areas, or facility distribution alone. There remains a significant lack of research that comprehensively clusters regions on a national scale by integrating multi-dimensional factors into a holistic health vulnerability framework. Accordingly, this study aims to perform clustering of health vulnerability levels among Indonesian provinces using the K-Means method in order to obtain groups of regions with similar vulnerability characteristics. The results are expected to serve as a reference for determining health policy priorities, allocating resources, and strengthening targeted health intervention programs across all provinces in Indonesia.

2. RESEARCH METHODOLOGY

This study uses the K-Means Clustering method with the calculation of Euclidean distance. The limitations of this study use data from 2024. Later, the results of the grouping can be used to evaluate policy-making related to public health. The research flow chart can be seen in Figure 1.

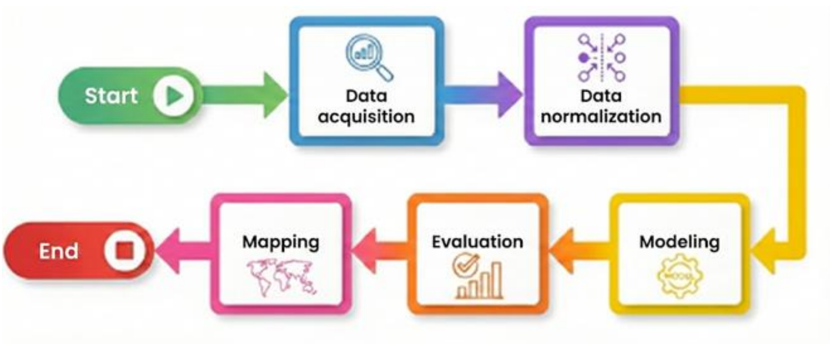

Figure 1. Research Flow

Okma Arnilia: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 7, Number 1, March 2026, Page 26-34
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v7i1.1469

This research began by collecting data from BPS related to the variables that can be seen in Table 1. After that, normalization is carried out using z-score. Then modeling was carried out using K-Means. The modeling results were then evaluated using the elbow method. The last step is Mapping and Interpretation of the modeling results so that conclusions can be drawn more easily and relevant policy recommendations are provided to the authorities.

2.1 Data Acquisition

This study uses data from BPS for 2024 on public health with variables that can be seen in Table 1. These variables are several factors that affect the public health index in Indonesia. The data is used for 38 provinces in Indonesia. The sample data used can be seen in Table 2

Table 1. Variables Used in Modeling

ID	Variable	Units
1	Number of Hospitals	Unit
2	Percentage of Sanitation Feasible	%
3	Access to Decent Drinking Water	%
4	Prevalence of Stunting	%
5	Number of Health Facilities	Unit
6	Population	Thousand Inhabitants
7	Percentage of Poor Population	%

Table 2. Sample data

Number of Hospitals	Percentage of Sanitation Feasible (%)	Access to Decent Drinking Water (%)	Prevalence of Stunting (%)	Number of Health Facilities	Population (Thousand Inhabitants)	Percentage of Poor Population (%)
75	81,1	90,08	28,60	439	5626,00	12,64
211	85,73	92,94	22,00	1426	15785,80	7,19
73	72,82	86,85	24,90	250	5914,30	5,42
83	86,32	92,24	20,10	551	6811,20	6,36
...	...	...	...	...	...	...
9	12,61	30,64	0,00	17	1484,90	29,66

2.2 Data Normalization

The data normalization carried out in this study was using the Z-score standardization method. Z-score is a normalization technique that converts the original value of a variable into a standard value by subtracting the mean and dividing it by the standard deviation, resulting in a distribution with an average of 0 and a standard deviation of 1 [8], [15]. This method is commonly used in multivariate data processing and clustering algorithms because it is able to balance the scale between variables, especially when the measurement units are different or have a wide range of values [16]. Z-score was chosen in this study because K-Means is sensitive to differences in data scale; variables with a larger range can dominate the process of calculating Euclidean distances. With Z-score normalization, all numerical variables have a balanced proportion of influence on cluster formation [7]. The variables normalized in this study are all numerical variables. The Z-score formula can be seen in Equation (1), while the normalization results are presented in Table 3.

$$z = \frac{x-\mu}{\sigma} \qquad (1)$$

where x is the original value, μ is the mean value of the variable, and σ is the standard deviation of the variable.

Okma Arnilia: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 7, Number 1, March 2026, Page 26-34
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v7i1.1469

2.3 Modeling

The model used in this study is K-Means clustering. K-Means is an unsupervised learning algorithm that aims to group data into clusters based on attribute similarity, where each data will be placed on the cluster closest to the centroid [7], [8]. This algorithm works iteratively by updating the position of the centroid until it reaches a convergent state. K-Means is widely used because it is simple, computationally efficient, and able to provide clear segmentation results on large-scale data [7], [9]. The basic formula for calculating distance in K-Means is shown in Equation (2). In this study, the k-value used is in the range of 2 to 10 to evaluate the various possible most optimal number of clusters.

$$J = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2 \qquad (2)$$

where *k* is the number of clusters, C_i is a dataset in the *i cluster*, *x* is the data of *j*, μ_i is the *centroid of the i cluster*, $\|x - \mu_i\|$ is the Euclidean distance between the data and the centroid.

2.4 Model Evaluation

process using the K-Means algorithm to determine the most optimal number of clusters. In this study, the determination of k values was carried out using the Elbow Method, which is a method that evaluates changes in within-cluster sum of squares (WCSS) values or inertia in various clusters to identify optimal points before the decline in WCSS begins to slow down significantly [16]. The main principle of the Elbow Method is to find the k-value at the "elbow" point, i.e. the position when the addition of clusters no longer provides a substantial decrease in inertia, so that the use of additional clusters is considered inefficient or does not provide a significant improvement in the quality of clustering [17]. This method is widely used in unsupervised learning analysis because it is simple, intuitive, and does not require data labels for the evaluation process [18]. In addition, the Elbow Method has proven to be effective in spatial and public health data analysis, especially when the variables used have different scales and distributions, as inertia is able to capture variations in distance between data globally [19]. The determination of the k-value using the Elbow Method is carried out by calculating the inertia in a certain range of k, then visualizing it in the form of a curve to identify the most stable elbow point as the best k-value [20]. The steps to calculate the Elbow method are as follows:

1. Calculate WCSS/Inertia for various k values =1,2,3,.., k_{max}
2. Plot the WCSS against *k*
3. Look for the "elbow point", which is the point where the decline in WCSS/Inertia starts to slow down drastically.
   a. Before this point: adding clusters significantly reduces WCSS/Inertia.
   b. After this point: the decrease in WCSS/Inertia is relatively small which can be said *k* in this position is the optimal number of clusters.

2.5 Mapping and Interpretation Results

The mapping was carried out after modeling results were obtained using the K-Means algorithm because the resulting cluster labels are still numerical and need to be interpreted spatially so that the patterns of health vulnerability between regions can be understood contextually and meaningfully [21]. The interpretation stage is important to convert statistical outputs into informative geographic information, so that the relationships between health indicators, environmental conditions, and risk distribution can be analyzed more comprehensively and support evidence-based decision-making [22]. The mapping process was carried out by integrating the results of clustering into provincial administrative spatial data using regional codes as a link, then visualized to display the distribution of health vulnerability clusters between provinces [23]. This visualization helps to show the spatial patterns of clusters, such as regions with similar vulnerability categories and areas that appear as outliers, so that geographical characteristics and regional contexts can be analyzed in a more targeted manner [24]. Furthermore, interpretation is carried out by reviewing the average value of health indicators in each cluster to determine the level of vulnerability of each regional group, so that the results of clustering can be used as a basis for determining intervention priorities and formulating more targeted health policies [25].

Okma Arnilia: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 7, Number 1, March 2026, Page 26-34
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v7i1.1469

3. RESULT AND DISCUSSIONS

3.1. Data Normalization

The results of data normalization using the z-score method in Table 3 show the change in the value of each variable to a standardized scale that has a mean of zero and a standard deviation of one. This transformation makes all variables such as the number of hospitals, decent sanitation, access to decent drinking water, stunting prevalence, number of health facilities, population population, and percentage of poor people can be compared equally in the next analysis. A positive z-score value, such as 1.260 on the number of hospitals or 0.509 on access to decent drinking water, indicates that the original value of the variable is above the population average. On the other hand, negative values such as –0.559 in proper sanitation or –0.834 in the population indicate below-average conditions. This pattern of value variation provides an initial indication of inequality between regions, for example, there are areas with health facilities that are much better than other regions or there are areas that are still lagging behind in access to basic sanitation.

Table 3. Data Normalization Results Using Z-Score

Number of Hospitals	Percentage of Sanitation Feasible (%)	Access to Decent Drinking Water (%)	Prevalence of Stunting (%)	Number of Health Facilities	Population (Thousand Inhabitants)	Percentage of Poor Population (%)
-0,068	-0,003	0,264	0,990	-0,025	0,314	-0,165
1,2607029	0,308329	0,509337	0,120425	1,22698	-0,553	0,736
-0,0876	-0,559	-0,012	0,502	-0,265	-0,834	-0,139
0,010	0,348	0,449	-0,130	0,116	-0,685	-0.0598

3.2. Modeling and Evaluation

Modeling was carried out using the K-Means algorithm to classify the level of health vulnerability in 38 provinces in Indonesia based on health and socio-economic variables. The modeling stage begins with normalization of data using z-score to equalize the scale of variables, considering that indicators such as the number of hospitals, stunting prevalence, and population have different ranges of values. This normalization is important to avoid bias due to large-scale variables, as recommended by previous research related to clustering in public health data [7], [8].

After the data is normalized, the process continues by determining the optimal number of clusters using the Elbow Method. This method evaluates the change in the inertia value (within-cluster sum of squares/WCSS) for various k-values, and identifies the elbow point i.e. the position when the addition of clusters no longer results in a significant decrease in inertia [17]. Based on the results of the inertial calculation in the range k = 2 to k = 10 visualized in Figure 2, the elbow point is most clearly visible at k = 4, so this study determines four clusters as the most optimal configuration to describe the level of vulnerability of the province's health.

Okma Arnilia: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 7, Number 1, March 2026, Page 26-34
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v7i1.1469

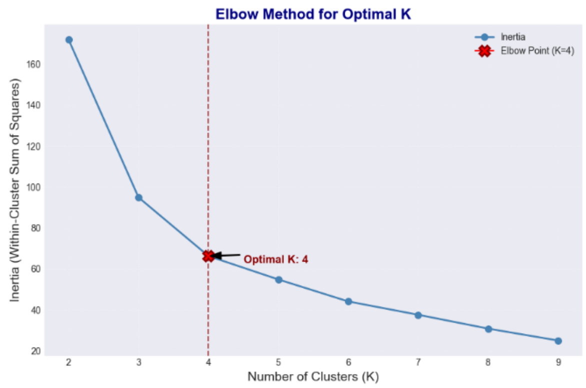

Figure 2. Evaluation of Modeling Results Using the Elbow Method

These results are in line with several previous studies that show that regional health data in Indonesia generally form a cluster structure with a total of 3–4 groups. For example, research on mapping health facilities in the city of Bandung through E-Jurnal Swadharma shows three to four clusters that are stable in the analysis of health spaces [10]. Research in JournalShub related to provincial-level clustering of health complaints also indicates a similar pattern, where the optimal number of clusters is in that range [11]. This alignment shows that the diversity of health and socio-economic characteristics between provinces tends to be divided into several distinct groups, such as low, medium, high, and very high risk groups.

The model evaluation showed that the decrease in inertia from k = 2 to k = 3 and from k = 3 to k = 4 was still quite significant. However, after k = 4, the decrease in inertia became much smaller in the next k (k ≥ 5), so the addition of clusters was considered not to provide a significant improvement in the quality of clustering. These findings support the findings of Kodinariya & Makwana [17], who stated that the Elbow Method is effective for determining the optimal number of clusters before the model structure is overfitting. In addition, other literature that discusses spatial analysis and public health also confirms that the Elbow Method is a stable and reliable approach to multidimensional data [18], [19].

3.3. Mapping and Interpretation Results

Before the mapping process is carried out, the first step that must be prepared is a shapefile (SHP) file which contains the administrative boundaries of all provinces in Indonesia, as many as 38 provinces. A visualization of those administrative boundaries is shown in Figure 3 as the basis for the next data overlay. Once the baseline map is available, the analysis continues by displaying Table 4 which contains the average value of each variable in each cluster, so that the main characteristics of each cluster can be compared more clearly.

Figure 3. Visualization of provincial administrative boundaries in Indonesia

Okma Arnilia: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 7, Number 1, March 2026, Page 26-34
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v7i1.1469

Table 4 presents the average value for each variable in each cluster. Cluster C1 shows characteristics with a relatively high number of hospitals and health facilities, supported by good access to sanitation and decent drinking water, as well as lower poverty rates. Cluster C2 has moderate average access to sanitation and drinking water, but the prevalence of stunting and the percentage of poor people tend to be higher than other clusters. The C3 cluster shows quite good conditions in terms of access to basic services and the number of health facilities, with a low poverty rate. Meanwhile, the C4 cluster displays the most disadvantaged conditions, characterized by low access to sanitation and drinking water, a very limited number of health facilities, and a much higher poverty rate than other clusters. This paragraph provides a clear overview of the differences in characteristics between clusters in the context of health and demographic indicators.

Table 4. Average Grouping Results

Cluster	Number of Hospitals	Percentage of Sanitation Eligible (%)	Access tc Decent Drinking Water (%)	Prevalence of Stunting (%)	Number of Health Facilities	Population (Thousand Inhabitants)	Percentage of Poor Population (%)
C1	396	82	95,75	15,86	2687	43.694, 06	8,74
C2	31,53	80,06	86,26	27,66	95,07	2524,30	14,89
C3	75,70	87,06	89,25	19,70	404,60	5878,16	6,40
C4	11,50	27,02	56,08	0.00	28	1488,60	28,63

Figure 4. Cluster Result Mapping Using K-Means

Figure 4 shows the spatial distribution of K-Means clustering with K=4 on provincial health conditions in Indonesia. The pattern of cluster distribution clearly shows that most provinces are in Cluster 3 (yellow), which describes health conditions in the "adequate" category with moderate access to basic services. Cluster 2 (green color) is spread across various regions, especially in the eastern region and several provinces of Sumatra, indicating relatively better health conditions. Cluster 1 (blue), which is the "very good" category, appears in only a few provinces, especially in the southern Java region, showing superior levels of health services and facilities. In contrast, Cluster 4 (in red) is concentrated in the Papua region, reflecting areas that need special attention due to a combination of low health facilities, access to sanitation, and high social vulnerability. This visualization provides a clear spatial picture of health disparities between provinces, while reinforcing the significance of the clustering results in identifying areas that require more intensive health interventions.

3.4. Discussion

The clustering results clearly illustrate the multidimensional nature of health vulnerability across Indonesian provinces, validating the significant disparity in health infrastructure and socio-economic conditions between regions. Cluster 1 (C1), predominantly in Java, demonstrates high health resilience characterized by highly centralized health resources, abundant hospitals, and low poverty rates (8.74%). Conversely, Cluster 4 (C4) highlights a critical level of health vulnerability, heavily concentrated in the

Okma Arnilia: *Corresponding Author

JURNAL INFORMATIKA DAN REKAYASA PERANGKAT LUNAK (JATIKA)
Volume 7, Number 1, March 2026, Page 26-34
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v7i1.1469

Papua region, where the highest poverty rate (28.63%) directly correlates with alarmingly low access to proper sanitation (27.02%) and a severe lack of hospitals (average 11.50). Meanwhile, Clusters 2 and 3 represent transitional vulnerabilities; although Cluster 3 shows relatively adequate basic services and the lowest poverty rate (6.40%), Cluster 2 reveals a specific vulnerability regarding child health, marked by the highest stunting prevalence (27.66%). These findings strongly imply that national health policy formulation and budget allocation must be distributed equitably based on cluster needs—prioritizing massive infrastructure and sanitation development for highly vulnerable regions (C4), targeted nutritional and poverty alleviation programs for C2, and service quality maintenance for more developed regions (C1 and C3).

4. CONCLUSION

This study successfully applied the K-Means algorithm to group provinces in Indonesia based on seven health and socio-economic indicators. Through data normalization and Elbow Method analysis, the optimal number of clusters is set to four clusters. In general, Cluster 1 shows the best health performance with adequate facilities and low stunting prevalence, while Cluster 4 describes the most vulnerable health conditions, especially seen in the Papua region which has a lack of health facilities, low access to sanitation, and high poverty. Spatial visualization of the clustering results shows that there is a fairly prominent health inequality between regions, where provinces with good and very good status are spread unevenly, while provinces in the vulnerable category are concentrated in certain regions. Thus, the results of this research can be an important basis for the government in formulating health intervention programs that are more targeted, effective, and based on local needs.

5. REFERENCES

[1] Kementerian Kesehatan Republik Indonesia, *Profil Kesehatan Indonesia 2023*. Jakarta: Kemenkes RI, 2023.
[2] Kementerian Kesehatan Republik Indonesia, *e-PPGBM (Elektronik Pencatatan dan Pelaporan Gizi Berbasis Masyarakat)*, Jakarta: Kemenkes RI, 2023.
[3] Direktorat Kesehatan Lingkungan, *Pedoman Penilaian Kerentanan dan Adaptasi*. Jakarta: Ditjen Penanggulangan Penyakit, Kementerian Kesehatan RI, 2025.
[4] World Health Organization (WHO), *Climate change and health vulnerability and adaptation assessment*. Geneva: WHO, 2021.
[5] Y. Ma, Q. Xiang, C. Yan, H. Liao, and J. Wang, "Poverty vulnerability and health risk action path of families of rural elderly with chronic diseases: Empirical analysis of 1,852 families in central and western China," *Frontiers in Public Health*, vol. 10, p. 776901, 2022, doi: 10.3389/fpubh.2022.776901.
[6] Pusat Krisis Kesehatan, *Pedoman Nasional Penanggulangan Krisis Kesehatan*. Jakarta: Kementerian Kesehatan Republik Indonesia, 2023.
[7] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.
[8] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Waltham: Morgan Kaufmann, 2012.
[9] T. R. Kodinariya and P. R. Makwana, "Review on determining number of clusters in K-means clustering," *International Journal*, 2013.
[10] F. Hamami and I. A. Dahlan, "Penerapan algoritma K-Means untuk memetakan persebaran fasilitas dan tenaga kesehatan di Kota Bandung," *JRIS: Jurnal Rekayasa Informasi Swadharma*, vol. 4, no. 2, 2024, doi: 10.56486/jris.vol4no2.561.
[11] N. Rofiqo, A. P. Windarto, and D. Hartama, "Penerapan clustering pada penduduk yang mempunyai keluhan kesehatan dengan datamining K-Means," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 2, no. 1, 2018, doi: 10.30865/komik.v2i1.929.
[12] G. M. M. Sujak, H. N. Rofiq, and F. I. Tawakal, "Implementasi K-Means Clustering untuk Optimalisasi Anggaran Penyakit Tidak Menular: Implementation of K-Means Clustering for Optimizing Non-Communicable Disease Budgets", MALCOM, vol. 5, no. 1, pp. 67-74, Nov. 2024.
[13] K. E. Setiawan, A. Kurniawan, A. Chowanda, and D. Suhartono, "Clustering models for hospitals in Jakarta using fuzzy c-means and k-means," *Procedia Computer Science*, vol. 216, pp. 356–363, 2023, doi: 10.1016/j.procs.2022.12.146.

Okma Arnilia: *Corresponding Author

Volume 7, Number 1, March 2026, Page 26-34
E-ISSN 2797-2011 P-ISSN 2797-3492
https://publikasi.teknokrat.ac.id/index.php/jatika/index
DOI: https://doi.org/10.33365/jatika.v7i1.1469

[14] Karmilasari, "Klasterisasi penyakit berdasarkan wilayah dengan metode K-Means (studi kasus: Propinsi Jawa Timur)," *JURNAL TEKNOLOGI INFORMASI: Teori, Konsep dan Implementasi*, vol. 15, no. 1, 2024, doi: 10.36382/jti-tki.v15i1.517.
[15] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, Springer, 2013.
[16] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. Morgan Kaufmann, 2011.
[17] T. R. Kodinariya and P. R. Makwana, "Review on determining number of clusters in K-means clustering," International Journal of Advance Research in Computer Science and Management Studies, vol. 1, no. 6, pp. 90–95, 2013.
[18] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.
[19] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. London: Academic Press, 2009.
[20] L. Kaufman and P. J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis. New York: Wiley, 2005.
[21] M. F. Goodchild, "GIS and spatial analysis," *International Journal of Geographical Information Systems*, vol. 6, no. 4, pp. 385–395, 1992.
[22] P. Longley, M. F. Goodchild, D. J. Maguire, and D. W. Rhind, *Geographic Information Systems and Science*, 3rd ed. Hoboken, NJ: Wiley, 2011.
[23] T. A. Slocum, R. B. McMaster, F. C. Kessler, and H. H. Howard, *Thematic Cartography and Geovisualization*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2009.
[24] M. De Smith, M. Goodchild, and P. Longley, Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools, 6th ed. Winchelsea Press, 2018.
[25] N. Andrienko and G. Andrienko, Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach, Berlin: Springer, 2006.

Okma Arnilia: *Corresponding Author