

Comparison of K-Means and DBSCAN Algorithms in Regional Segmentation in Indonesia Based on Mortality and Fertility Indicators

Bunga Nafandra¹, Dwi Sulistiowati², Nonong Amalita^{3*}

^{1,2,3} *Statistics Study Program, Padang State University, Padang, Indonesia*

*Corresponding author: nongmat@fmipa.unp.ac.id

*Submission date: 11 October 2025, Revision: 13 November 2025, Accepted: 07 February 2026

ABSTRACT

Mortality and fertility indicators can be interpreted through maternal mortality rates, infant mortality rates, under-five mortality rates, child mortality rates, and crude birth rates. Disparities in mortality and fertility between regions indicate differences in health quality. In Indonesia, the maternal mortality rate has declined over the past decade, from 346 to 189 per 100000 live births, but this figure has not yet reached the global target set by the SDGs of 70 per 100000 live births. To reduce these mortality indicators, it is necessary to identify regions with low health quality so that interventions can be more targeted and focused. Therefore, this study aims to form regional segmentation based on mortality and fertility indicators by comparing two clustering algorithms, namely K-Means and Density-Based Spatial Clustering of Applications with Noise (DBSCAN). The data used is secondary data from the Indonesian Central Statistics Agency, taken from the 2020 Population Census. The results of the study found that the K-Means algorithm is better than DBSCAN based on the Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index evaluation matrices, with values of 0.485, 0.804, and 36.34, respectively. This algorithm formed two clusters, with western Indonesia tending to have better health quality than eastern Indonesia. Based on the results of this study, it is hoped that the government can provide focused interventions in eastern Indonesia to improve maternal and child health quality and reduce mortality rates.

KEYWORDS

Clustering, DBSCAN, Fertility, K-Means, Mortality.

1. INTRODUCTION

Maternal and child health remains an important issue in health development in Indonesia. Mortality and fertility indicators such as the Maternal Mortality Rate (MMR), Infant Mortality Rate (IMR), and Crude Birth Rate (CBR) are important measures for assessing the social and economic conditions and the effectiveness of the health care system in a region. These three indicators are interrelated and provide an overview of the population's welfare, environmental conditions, and the quality of health care received by the community.

The MMR, IMR, and CBR indicators are often used in research in the context of maternal and child health [1]. The MMR describes the number of maternal deaths during pregnancy, childbirth, and the postpartum period per 100000 live births [2]. This indicator is important because it reflects the quality of maternal health services, nutritional status, and environmental conditions. Meanwhile, the IMR measures the number of infant deaths before the age of 1 per 100000 live births, which reflects the vulnerability of infants to disease, nutritional status, and access to basic health services [2]. In addition, the Crude Birth Rate (CBR) reflects the number of live births per 100000 population per year, which can be used as an illustration of fertility dynamics in society [2]. High maternal, infant, and child mortality rates are generally influenced by factors such as poor

nutrition, inadequate sanitation, and high prevalence of infectious diseases [3]. Sharp differences in mortality and fertility rates between regions in Indonesia indicate disparities in quality of life and access to health facilities.

Based on a literature study by Syairaji et al. [4], the maternal mortality rate (MMR) in Indonesia shows a downward trend, but it is still high compared to other ASEAN countries and far above the Sustainable Development Goals (SDGs) target. This is supported by data published by Statistics Indonesia in the Indonesian Demographic Statistics publication, which shows that the maternal mortality rate has decreased from 346 to 189 per 100,000 live births based on population census data from 2010 and 2020. However, this result has not yet reached the global target set by the SDGs, which aims for a maternal mortality rate of less than 70 deaths per 100,000 live births [2]. This condition emphasizes the need for regional mapping based on mortality and fertility characteristics so that health policies can be formulated in a more targeted manner according to regional conditions.

In an effort to group regions based on mortality and fertility characteristics, the clustering method is one effective approach to identify patterns and similarities between regions. However, previous studies have mostly used only one type of clustering algorithm, namely K-Means, which has limitations in detecting data patterns with irregular distributions or noise. For example, in a study by Eka Sabna et al. [3], regions in Indonesia were grouped based on child mortality rates using the K-Means algorithm. This study had limitations in terms of variables, as it only used one indicator and did not use a comparative clustering algorithm. Therefore, a comparison with other algorithms such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is needed, which is more adaptive in recognizing clusters with irregular shapes and is able to identify outlier data [5].

Based on that description, this study raises the topic of “Comparison of K-Means and DBSCAN Algorithms in Regional Segmentation in Indonesia Based on Mortality and Fertility Indicators”. This study aims to find the best clustering algorithm between K-Means and DBSCAN, as well as to group regions in Indonesia based on mortality and fertility indicators. Through this comparison study, it is hoped that a more comprehensive picture of regional mortality and fertility characteristics can be obtained, thereby supporting the formulation of more targeted health policies.

2. LITERATURE REVIEW

2.1 Clustering

Clustering refers to an analysis performed to group a set of objects in such a way that the objects are placed into specific clusters [6]. The results of cluster analysis will group data with similar characteristics into one cluster, while data with different characteristics will be placed into another cluster. This causes the variance between objects within a cluster to be small, but the variance between clusters to be large [7].

2.2 K-Means Clustering Algorithm

The K-Means algorithm is a non-hierarchical cluster analysis technique, in which the number of clusters (k) is determined in advance [7]. K-Means is a distance-based grouping, which works by grouping objects based on the proximity of points to the cluster center (centroid) calculated iteratively [8]. One of the distance calculations often used in cluster analysis is the Euclidean distance, with the formula [9]:

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{jk})^2} \quad (1)$$

Where d_{ik} is the distance of data point i to cluster k , x_{ij} is data point i in column j , and c_{jk} is cluster k in column j .

2.2.1 Silhouette Method

The Silhouette score is a measure that describes how well an object fits its cluster [10]. A score closer to 1 indicates that the data point is in the correct cluster [10]. The Silhouette method for determining the optimal number of clusters is represented in the form of a Silhouette graph and can be seen at the highest value [11].

2.2.2 Elbow Method

The Elbow method evaluates clustering results by comparing the sum of squared errors in each cluster [12]. The Elbow method determines the optimal number of clusters by identifying the inflection point on the graph, which is when the decrease in the Sum of Squared Errors (SSE) is no longer significant, thus forming an elbow-like pattern that is used as a reference for cluster selection [11].

2.3 DBSCAN Clustering Algorithm

The DBSCAN algorithm is a density-based clustering process that groups data based on the number of points around it within a certain radius [13]. An object can be categorized as part of a cluster if it is within a certain radius and meets the minimum number of neighbors that has been determined, while objects that do not meet these criteria are categorized as noise [14]. In addition to being able to detect clusters in large datasets, the DBSCAN algorithm is effective for identifying data with irregular patterns and detecting the presence of outliers [15].

In its application, DBSCAN has two main parameters, namely the epsilon (ϵ) value and min points (minPts). The ϵ value describes the maximum distance for a point to be considered a neighbor, while minPts is the minimum number of neighbors required for a point to be categorized as a core point [16]. The ϵ value is generally determined using the K-Nearest Neighbor distance graph approach, where the selected value is the point on the curve that indicates an inflection point or a sharp increase. The minPts used in theory is $z \times$ the number of variables [14]. However, for data with a small number of objects, minPts is set to a minimum of around 3 or 4 [17].

Based on these parameters, each data point in DBSCAN is grouped into three main categories, namely core points, border points, and noise. Core points are points that have a minimum number of neighbors (minPts) within a radius of ϵ , representing areas with high density. Meanwhile, border points are points that do not meet the core points criteria but are still within a distance of ϵ from at least one core point, so they remain part of the cluster. Noise or outliers are points that do not fall into the previous two categories because they are located in areas with low density, so they are not classified in any cluster [18].

2.4 Evaluation of Clustering Methods

Evaluations were conducted to select the best algorithm for clustering [15]. Several commonly used evaluation metrics are [6]:

2.4.1 Silhouette Score

The Silhouette score measures how well an object is in the right cluster compared to other clusters. A Silhouette value closer to 1 indicates that the clusters formed are better [15]. The Silhouette score can be calculated using the formula [19]:

$$s = \frac{b - a}{\max(a, b)} \quad (2)$$

Where a is the average distance between the observation and all other points in the same cluster, and b is the average distance between the observation and all other points in the nearest cluster [19].

2.4.2 Calinski-Harabasz Index

This index evaluates cluster quality by comparing inter-cluster dispersion with intra-cluster dispersion. The expected result is smaller intra-cluster dispersion and larger inter-cluster dispersion, as this indicates that objects within clusters are dense and compact, and that clusters are clearly separated. The highest CHI value indicates better clustering quality [15]. The CHI value can be calculated using the formula [20]:

$$CHI = \frac{tr(B_k)}{tr(W_k)} \times \frac{N - k}{k - 1} \quad (3)$$

where $tr(B_k)$ is the total amount of inter-cluster variance, $tr(W_k)$ is the total amount of intra-cluster variance, N is the amount of data, and k is the number of clusters formed.

2.4.3 Davies-Bouldin Index (DBI)

The Davies-Bouldin Index (DBI) measures cluster quality by calculating the average ‘similarity’ between clusters. A smaller DBI value indicates clearly separated clusters without overlap [15]. DBI can be calculated using the following equation [18]:

$$DBI = \frac{1}{M} \sum_{j=1}^M \max R \quad (4)$$

Where M is the number of clusters, and R is the ratio between cluster distances.

3. METHODOLOGY

The data analyzed in this study is secondary data consisting of mortality and fertility indicators for each province in Indonesia. This data was published by the Central Statistics Agency (BPS) and is the result of the 2020 Population Census [2]. Information related to the research variables is presented in **Table 1**:

Table 1. Research Variables

Variable	Description	Scale
X_1	Crude Birth Rate (persons/1000 population)	Numeric
X_2	Maternal Mortality Rate (persons/100000 live births)	Numeric
X_3	Infant Mortality Rate (persons/1000 live births)	Numeric
X_4	Child Mortality Rate (persons/1000 live births)	Numeric
X_5	Under-five Mortality Rate (persons/1000 live births)	Numeric

Based on these variables, the provinces in Indonesia are grouped according to mortality and fertility indicators. The stages of analysis conducted in this study are as follows:

1. Performing cluster analysis using the K-Means algorithm
 - (a) Determining the optimal number of clusters using the Silhouette method and the Elbow method
 - (b) Selecting the optimal number of clusters from the results of the Silhouette method and the Elbow method by comparing Silhouette scores
 - (c) Determining cluster centers
 - (d) Calculating the distance of objects to cluster centers
 - (e) Grouping objects based on the closest distance
2. Performing cluster analysis using the DBSCAN algorithm
 - (a) Determine the minPts parameter value (minimum number of neighbors) and ϵ (radius/neighbors around a point)
 - (b) Classify points that have $\geq$ minPts neighbors within radius ϵ as core points
 - (c) Classify points within radius ϵ but with $<$ minPts neighbors as border points
 - (d) Classify points that are not included in core points and border points as noise points
 - (e) Repeat until all points are classified
3. Evaluate the clustering results and select the best algorithm
4. Draw conclusions

4. RESULT & DISCUSSION

4.1 Descriptive Statistics

Table 2 summarizes the descriptive statistics results of the numerical variables used in this study:

Based on **Table 2**, it can be seen that the average crude birth rate is 18.45 per 1000 population. The standard deviation value of 2.25 is relatively small, indicating that the crude birth rate between regions tends to be uniform and does not differ significantly from other regions. For the maternal mortality rate variable, the average is 210.50 per 100000 live births. The standard deviation for this variable is quite large, at 89.11, indicating that there are quite drastic differences in maternal mortality rates between regions. Infant mortality is also relatively high, with an average of 19.74 per 1000 live births. The data variation is quite wide, with a standard deviation of 7.12, indicating that there are regions with infant mortality rates above the national average. For child mortality (ages 1–4), the average is 3.85 per 1000 live births, with a standard deviation of 2.20. Although

Table 2. Descriptive Statistics

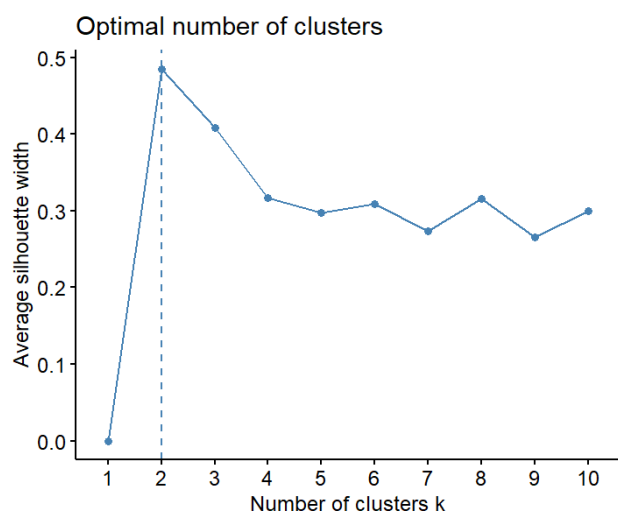
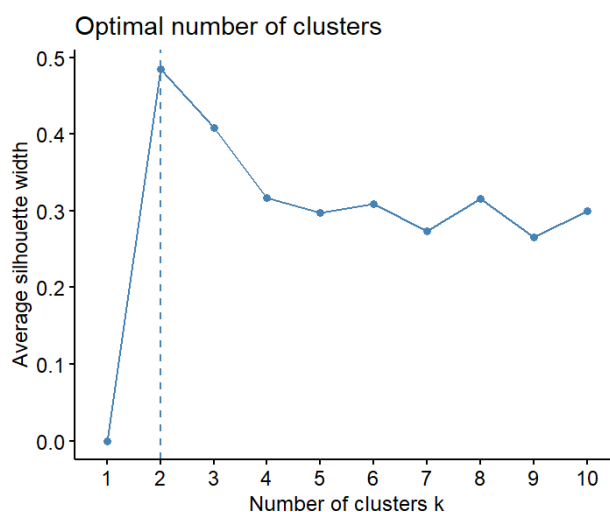
Variable	Mean	Min	Max	Std. Dev
X_1	18.45	13.69	22.84	2.25
X_2	210.50	48.00	565.00	89.11
X_3	19.74	10.38	38.17	7.12
X_4	3.85	1.64	10.88	2.20
X_5	23.58	12.02	49.04	9.29

the average is relatively low, there are regions with child mortality rates that are nearly three times higher than other regions. Meanwhile, the under-five mortality rate (0–4 years) shows an average of 23.58 per 1000 live births, with a range of 12.02 to 49.04 and a standard deviation of 9.29. This figure confirms that there are significant differences between regions in terms of infant survival. This is indicated by the large range of mortality rates between regions in Indonesia, which indirectly shows that there are areas with very high mortality rates and areas with very low mortality rates. However, this is in contrast to the relatively stable birth rate, which is characterized by a relatively small range of birth rates.

4.2 K-Means Clustering Algorithm

4.2.1 Determining the Optimum Number of Clusters

Figure 1 and **Figure 2** show the Elbow method and the Silhouette method in determining the optimal number of clusters:

**Figure 1.** Silhouette Method Graph**Figure 2.** Elbow Method Graph

Based on the Silhouette graph (**Figure 1**), the optimal number of clusters is 2, while the Elbow graph (**Figure 2**) shows inflection points at $k = 3$ and $k = 5$. To determine the most appropriate number of clusters, a comparison of validation values using Silhouette scores was performed. The number of clusters with the highest Silhouette score will be selected as the final number of clusters in the K-Means algorithm. A summary of the Silhouette score comparison results for the number of clusters 2, 3, and 5 can be seen in **Table 3**:

Table 3. Evaluating the Number of Clusters Using the Silhouette Score

Number of Clusters	Silhouette Score
2	0.49
3	0.41
5	0.34

Based on **Table 3**, the optimal number of clusters that produces the highest Silhouette Score is 2 clusters with a Silhouette Score of 0.49. Therefore, the number of clusters to be used for the K-Means clustering algorithm is 2 clusters.

4.2.2 Forming a K-Means Algorithm Cluster

Figure 3 is a visualization of two clusters formed from the results of K-Means cluster analysis:

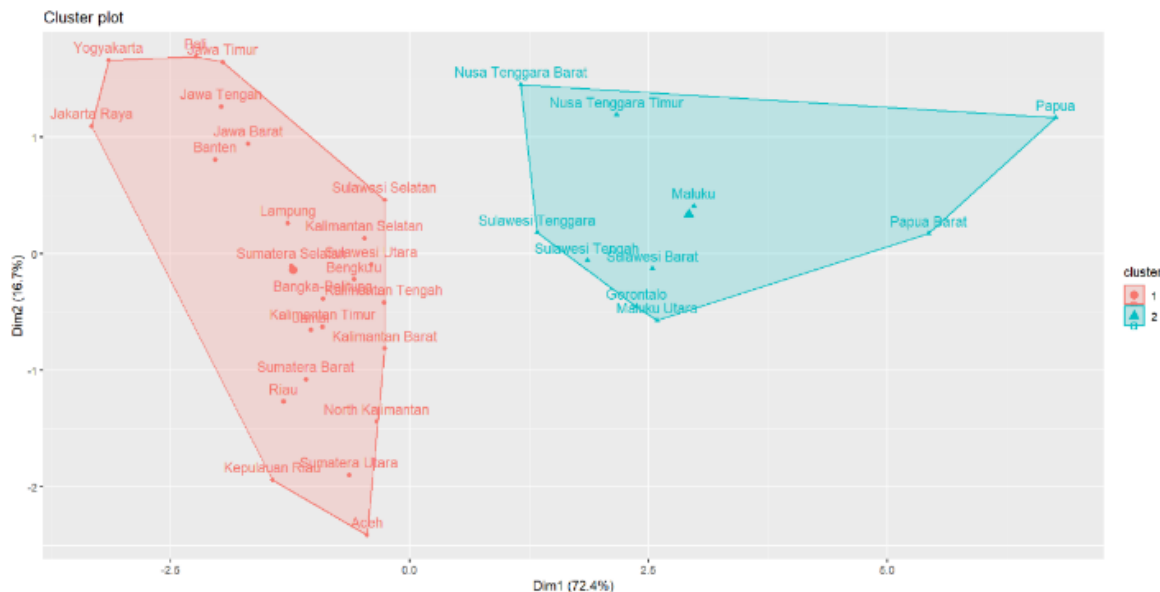


Figure 3. K-Means Clustering Results

Based on **Figure 3**, the results of applying the K-Means algorithm show that grouping regions in Indonesia based on mortality and fertility indicators produces two clearly separate clusters. The first clusters consists mainly of western Indonesia, while the second cluster is dominated by eastern Indonesia.

4.3 DBSCAN Clustering Algorithm

4.3.1 Determining The minPts and ϵ Parameters

In this study, the amount of data was relatively small, namely 34 objects with five features. When the minPts parameter was set to a value of 4 to 6, DBSCAN only formed one large cluster with one noise point. To address this, trials were conducted with minPts = 2 and minPts = 3. The K-Nearest Neighbor (KNN) graph for determining the epsilon (ϵ) range using minPts 2 and 3 can be seen in **Figure 4** and **Figure 5**:

Based on the KNN graph with minPts = 2 in **Figure 4**, several epsilon values that need to be considered are 1.5, 1.6, and 1.7. Meanwhile, based on the KNN graph with minPts = 3 in **Figure 4.5**, several epsilon values that need to be considered are 1.5, 1.7, and 2.2. To select the best parameter combination, compare the Silhouette Score and noise ratio of all parameter combinations. The combination that produces the highest Silhouette Score and the smallest noise ratio will be selected as the parameters in the DBSCAN clustering algorithm. **Table 4** summarizes the Silhouette Score and noise ratio values for several of these combinations: Based on **Table 4**, it was found that the parameter combinations (minPts = 2, ϵ = 1.6), (minPts = 2, ϵ = 1.7), and (minPts = 3, ϵ = 1.7) produced the highest Silhouette Score value, namely 0.454, with the same low noise ratio, namely 0.059.

In theory, a larger minPts value is generally recommended because it provides stricter density requirements in cluster formation [21]. This can increase the reliability of the cluster structure and reduce the risk of forming clusters that are too small or insignificant. Therefore, the best combination of parameters selected is minPts = 3 with ϵ = 1.7, as it produces the highest Silhouette Score with low and consistent noise, considering the theoretical aspects of the DBSCAN algorithm.

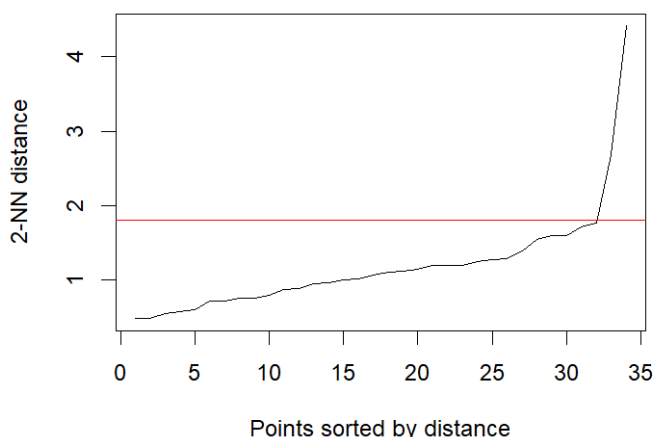


Figure 4. KNN Graph for minPts = 2

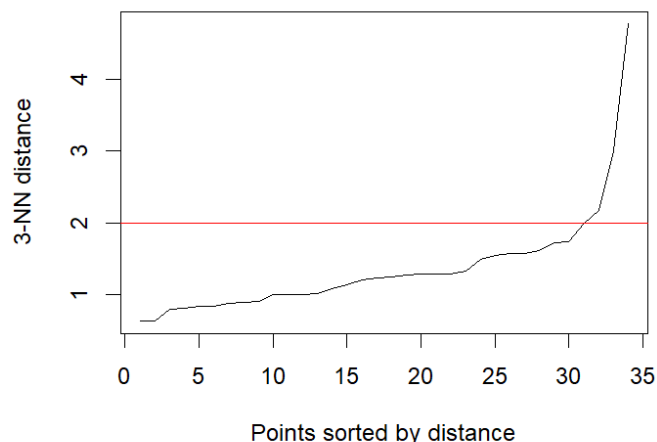


Figure 5. KNN Graph for minPts = 3

Table 4. Selection of the Best minPts and ϵ

minPts	ϵ	k	Noise	Silhouette Score	Noise Ratio
2	1.5	3	3	0.426	0.088
	1.6	2	2	0.454	0.059
	1.7	2	2	0.454	0.059
3	1.5	3	5	0.426	0.088
	1.7	2	2	0.454	0.059
	2.2	1	2	NA	0.059

4.3.2 Forming a DBSCAN Algorithm Cluster

Figure 6 is a visualization of two clusters formed from the results of DBSCAN cluster analysis with minPts = 3 and epsilon (ϵ) = 1.7:

Based on **Figure 6**, the application of the DBSCAN algorithm shows that the grouping of regions in Indonesia based on mortality and fertility indicators forms two clearly separated clusters, with two noise points that are not included in any cluster. The first cluster is dominated by western Indonesia, the second cluster by eastern Indonesia, while Papua and West Papua are identified as noise.

4.4 Evaluation of Clustering Results

The selection of the best clustering method is based on a comparison of the Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index. The results of comparing these three evaluation metrics are presented in **Table 5**:

Table 5. Selecting the Best Clustering Algorithm

Algorithm	Silhouette Score	Davies-Bouldin Index	Calinski-Harabasz Index
K-Means	0.485	0.804	36.34
DBSCAN	0.454	0.753	25.59

Based on **Table 5**, the K-Means clustering algorithm is superior to the DBSCAN algorithm because it has higher Silhouette Score and Calinski-Harabasz Index values than the DBSCAN algorithm. Although the DBI value for the DBSCAN algorithm is smaller, the Silhouette Index provides the best results in evaluating cluster quality and is more resistant to noise and outliers

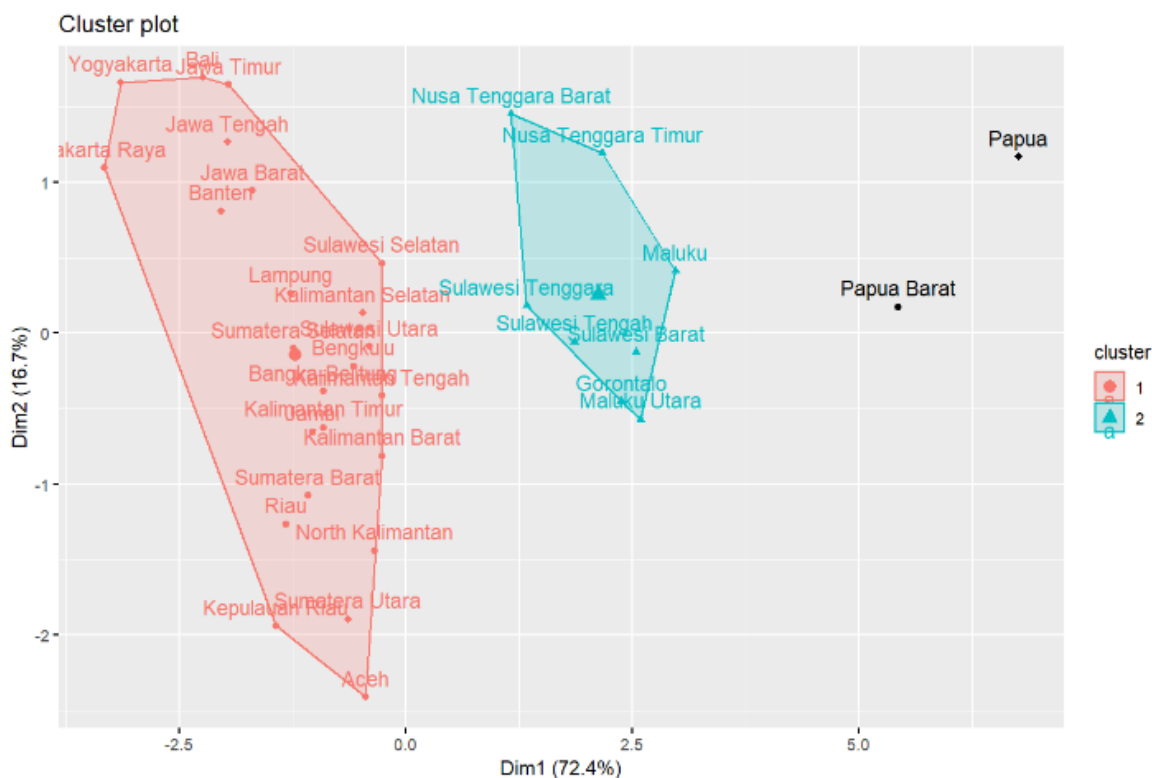


Figure 6. DBSCAN Clustering Results

compared to other indices [22].

4.5 Interpretation of Results

Figure 7 shows a map of the distribution of K-Means algorithm clusters based on mortality and fertility indicators in Indonesia:

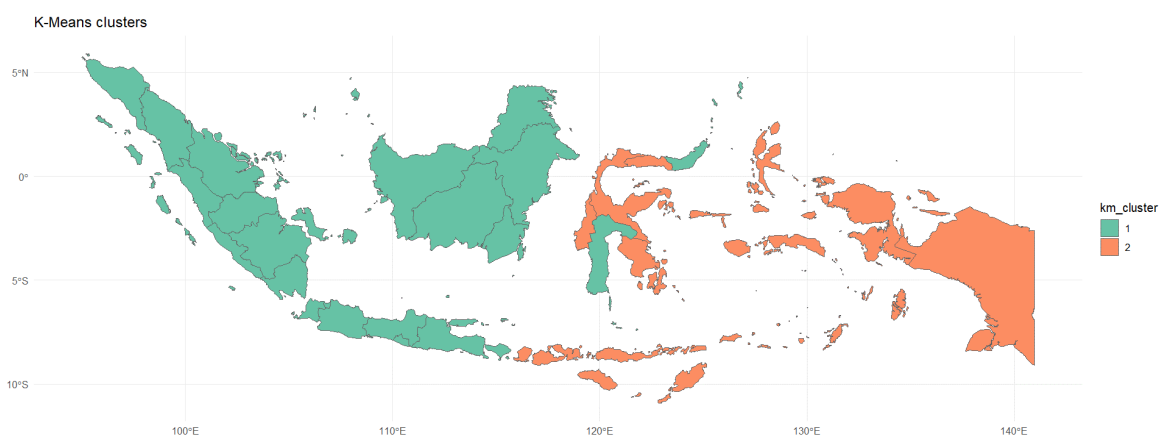


Figure 7. K-Means Algorithm Cluster Distribution Map Based on Mortality and Fertility Indicators

The provinces resulting from the K-Means clustering algorithm are summarized in **Table 6**:

Table 6. K-Means Clustering Results Area

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

The characteristics of each cluster formed can be seen in **Table 7**:

Table 7. Characteristics of Each Cluster

Cluster	X_1	X_2	X_3	X_4	X_5
Indonesia	18.5	210.5	19.7	3.9	23.6
1	17.5	172.0	15.7	2.7	18.4
2	20.7	303.0	29.4	6.7	36.0

Based on **Table 7**, it can be seen that regions included in cluster 1 have lower crude birth rates and lower maternal, infant, child, and toddler mortality rates than the national average. Conversely, cluster 2 shows values for these indicators that are higher than the national average. Mapping the distribution of clusters also found that most of western Indonesia is included in cluster 1, while most of eastern Indonesia is included in cluster 2. Thus, it can be concluded that western Indonesia has low mortality and fertility rates, while eastern Indonesia has high mortality and fertility rates.

4.6 Discussion

This study shows that the K-Means algorithm performs better than DBSCAN in clustering Indonesian regions based on mortality and fertility indicators. The application of the K-Means algorithm produced two main clusters. The first cluster was dominated by western Indonesia, which had a relatively low mortality rate, while the second cluster covered most of eastern Indonesia, which had higher than average national mortality and fertility rates.

These findings are in line with research conducted by Alya Azzahra and Arie Wahyu Wijayanto in 2022 [23], which concluded that western and central Indonesia are grouped in a cluster with good to very good quality of maternal health services. Adequate maternal health services play a crucial role in reducing maternal and infant mortality rates during childbirth.

Meanwhile, the characteristics of the second cluster describe regions with high mortality and fertility rates. These results are consistent with the 2020 study by Adina Astasia and You Ari Faeni [24], which used data from the Indonesian Demographic and Health Survey (IDHS) with a K-Means and biplot approach. The study shows that provinces such as Papua, Maluku, North Maluku, and West Papua have low maternal health quality and minimal awareness of the importance of visiting official health facilities for check-ups.

Additionally, a literature review conducted by Tyastuti Prima Hapsari and Lutfi Agus Salim in 2023 [25] found that regular pregnancy check-ups can minimize maternal mortality rates. Another study by Intan Permata Sari et al. [26] also concluded that access to health facilities has a significant effect on mortality rates. These results reinforce that disparities in health services are closely correlated with maternal mortality rates. Conceptually, this supports the research findings that eastern regions with limited access to health services tend to have higher mortality rates than western regions with better access to services.

Furthermore, a national-scale ecological study by Supriyatiningsih Wenang et al. published in PLOS ONE [27] also shows that high maternal mortality rates are often associated with disparities in access to maternal health services, particularly in terms of the availability of facilities, medical personnel, and infrastructure, which are generally still limited in remote areas. Thus, the

characteristics of the eastern cluster, which shows high mortality and fertility rates, can be explained by limited infrastructure and uneven distribution of health workers. Conversely, western regions with better health services tend to have lower mortality rates.

Overall, the results of this study emphasize the importance of equitable access to and quality of health services throughout Indonesia. The government needs to strengthen its health decentralization policies by ensuring the availability of medical personnel, improving health facilities, and developing transportation infrastructure in the eastern region. These efforts have the potential to not only reduce maternal and infant mortality rates, but also promote more equitable and sustainable health development throughout Indonesia.

The strength of this study lies in the use of two clustering algorithms, namely K-Means and DBSCAN, which enable a comparison of the effectiveness of the methods in a complex socio-demographic context and provide a comparative perspective that is rarely raised in similar studies. This approach also demonstrates the algorithms' ability to map regional inequalities visually and measurably. However, this study has limitations in terms of the number of variables used and the aggregate nature of the data per province, meaning that other socioeconomic factors such as education levels, income, and availability of health workers were not included in the analysis. For future research, it is recommended to develop models by adding socioeconomic, environmental, and infrastructure variables, as well as applying spatial clustering or hybrid machine learning approaches to obtain more precise and detailed results.

5. CONCLUSION

Based on the analysis results, it was concluded that the K-Means algorithm was superior to DBSCAN in clustering regions in Indonesia based on mortality and fertility indicators. The K-Means algorithm produced a higher Silhouette Score and Calinski-Harabasz Index than the DBSCAN algorithm. Based on the results of regional clustering analysis using the K-Means algorithm, two clusters were obtained. Most of western Indonesia is included in cluster 1, which has lower average crude birth rates, maternal, infant, child, and toddler mortality rates than the national average. Meanwhile, most of eastern Indonesia is included in cluster 2, which has higher average mortality and fertility indicators. Based on these results, it is recommended that the government pay more attention to regions in cluster 2 and provide more in-depth interventions to reduce maternal and child mortality rates.

REFERENCES

- [1] D. T. Saputro and W. P. Sucihermayanti, "Penerapan Klasterisasi Menggunakan K-Means untuk Menentukan Tingkat Kesehatan Bayi dan Balita di Kabupaten Bengkulu Utara," *JBI*, vol. 12, no. 2, pp. 146–155, Nov. 2021. [Online]. Available: <https://ojs.uaajy.ac.id/index.php/jbi/article/view/4861>
- [2] Badan Pusat Statistik, *Statistik Demografi Indonesia (Hasil Sensus Penduduk 2020)*. Jakarta: Badan Pusat Statistik, 2024.
- [3] E. Sabna and M. Mitra, "Analysis of K-Means Clustering in Grouping Child Mortality Rate in Indonesia," *JOISSE*, vol. 1, no. 1, pp. 27–31, Mar. 2024.
- [4] M. Syairaji, D. S. Nurdiati, B. S. Wiratama, Z. D. Prüst, K. W. M. Bloemenkamp, and K. J. C. Verschueren, "Trends and causes of maternal mortality in Indonesia: a systematic review," *BMC Pregnancy Childbirth*, vol. 24, no. 1, p. 515, Jul. 2024.
- [5] M. Fathur Rizki and F. Sulianta, "Analisis Klasterisasi Data Peserta Asuransi PT Xyz Menggunakan Metode Denisty-Based Spatial Clustering of Applications with Noise (DBSCAN)," *Jurnal Algoritma*, vol. 22, no. 1, pp. 993–1001, May 2025.
- [6] D. Jollyta, M. Siddik, H. Mawengkang, and S. Efendi, "Teknik evaluasi cluster," in *Teknik Evaluasi Cluster Solusi Menggunakan Python dan Rapidminer*. Yogyakarta: Deepublish, 2021, ch. I, pp. 1–9.
- [7] D. A. Setiady and H. Leong, "Implementation of k-means algorithm elbow method and silhouette coefficient for rainfall classification," *PROXIES*, vol. 4, no. 1, pp. 18–25, 2020.

- [8] T. Tendean and W. Purba, “Analisis cluster provinsi indonesia berdasarkan produksi bahan pangan menggunakan algoritma k-means,” *Jurnal Sains dan Teknologi*, vol. 1, no. 2, pp. 5–11, 2020.
- [9] K. B. Aditya, Y. Setiawan, and D. Puspitaningrum, “Sistem Informasi Geografis Pemetaan Faktor-Faktor yang Mempengaruhi Angka Kematian Ibu (AKI) dan Angka Kematian Bayi (AKB) dengan Metode K-Means Clustering (Studi Kasus: Provinsi Bengkulu),” *J. Teknik Informatika*, vol. 10, no. 1, pp. 59–66, Jan. 2018.
- [10] E. M. Fitri, R. R. Suryono, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, and A. Wantoro, “Klasterisasi Data Penjualan Berdasarkan Wilayah Menggunakan Metode K-Means Pada PT XYZ,” *Komputasi*, vol. 11, no. 2, pp. 157–168, Oct. 2023.
- [11] D. Widyadhana, R. B. Hastuti, I. Kharisudin, and F. Fauzi, “Perbandingan analisis klaster k-means dan average linkage untuk pengklasteran kemiskinan di provinsi jawa tengah,” in *PRISMA, Prosiding Seminar Nasional Matematika 4*, 2021, pp. 584–594.
- [12] M. Sholeh and K. Aeni, “Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means,” *STRING*, vol. 8, no. 1, p. 56, Aug. 2023.
- [13] I. N. Simbolon and P. D. Friskila, “Analisis dan Evaluasi Algoritma DBSCAN (Density-Based Spatial Clustering of Applications with Noise) pada Tuberkulosis,” *JITET*, vol. 12, no. 3S1, Oct. 2024.
- [14] A. M. A. Tambunan, U. Khaira, and B. F. Hutabarat, “Implementasi Algoritma DBSCAN untuk Klasterisasi Koperasi Berdasarkan Indikator Keuangan,” *Naratif*, vol. 7, no. 1, pp. 38–55, Jun. 2025.
- [15] M. R. Mubarak, A. T. J. Harjanto, and R. Renaldy, “Peningkatan Performa DBSCAN dengan Reduksi Dimensi Principal Component Analysis (PCA) dalam Klasterisasi Tingkat Kemiskinan di Indonesia,” *JINTEKS*, vol. 7, no. 3, pp. 1176–1184, Aug. 2025.
- [16] I. N. Setiawan, D. Krismawati, S. Pramana, and E. Tanur, “Klasterisasi wilayah rentan bencana alam berupa gerakan tanah dan gempa bumi di indonesia,” in *Prosiding Seminar Nasional Official Statistics 2022*, 2022, pp. 669–676.
- [17] R. Adha, N. Nurhaliza, U. Soleha, and M. Mustakim, “Perbandingan algoritma dbscan dan k-means clustering untuk pengelompokan kasus covid-19 di dunia,” *Jurnal Sains, Teknologi dan Industri*, vol. 18, no. 2, pp. 206–211, 2021, e-ISSN: 2407-0939.
- [18] R. Fadhlillah, B. D. F. Zamzami, G. A. Karim, A. Rachman, K. L. Azzira, F. D. Irawati, R. D. Permana, and R. M. Badri, “Analisis klasterisasi kepadatan penduduk kabupaten muara enim menggunakan algoritma dbscan,” *PROSIDING SEMINAR NASIONAL SAINS DATA*, vol. 4, no. 1, pp. 982–992, 2024.
- [19] T. D. Harjanto, A. Vatesia, and R. Faurina, “Analisis penetapan skala prioritas penanganan balita stunting menggunakan metode dbscan clustering,” *Jurnal Rekursif*, vol. 9, no. 1, pp. 30–42, 2021.
- [20] A. M. Sikana and A. W. Wijayanto, “Analisis Perbandingan Pengelompokan Indeks Pembangunan Manusia Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarchical Clustering,” *jik*, vol. 14, no. 2, p. 66, Sep. 2021.
- [21] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD’96)*. Portland, Oregon: AAAI Press, 1996, pp. 226–231.
- [22] N. E. Kondruk, “A Comparative Study of Cluster Validity Indices,” *RIC*, vol. 0, no. 4, pp. 59–67, Nov. 2019.
- [23] A. Azzahra and A. W. Wijayanto, “Perbandingan agglomerative hierarchical dan k-means dalam pengelompokan provinsi berdasarkan pelayanan kesehatan maternal,” *SISTEMASI: Jurnal Sistem Informasi*, vol. 11, no. 2, pp. 481–495, 2022, e-ISSN: 2540-9719.

- [24] You Ari Faeni and Adina Astasia, “Maternal Mortality Risk di Indonesia dengan Analisis Biplot dan Kluster Menggunakan Data Survei Kesehatan dan Demografi Indonesia (SDKI) Tahun 2017,” *JSC*, vol. 3, no. 2, pp. 144–152, Aug. 2020.
- [25] T. P. Hapsari and L. A. Salim, “Efektivitas Asuhan Antenatal sebagai Upaya untuk Mencegah Komplikasi Obstetri yang Berdampak terhadap Kematian Ibu: Literature Review,” *j. keperawatan. kesehat. masy. cendekia. utama.*, vol. 12, no. 2, p. 115, Jul. 2023.
- [26] I. P. Sari, C. A. Sucirahayu, S. A. Hafilda, S. N. Sari, V. Safithri, F. Fitria, J. Febriana, and H. Hasyim, “Faktor penyebab angka kematian ibu dan angka kematian bayi serta strategi penurunan kasus (studi kasus di negara berkembang): Systematic review,” *PREPOTIF: Jurnal Kesehatan Masyarakat*, vol. 7, no. 3, pp. 16 578–16 593, 2023.
- [27] S. Wenang, O. Emilia, A. Wahyuni, A. Afdal, and J. Haier, “Obstetrics care in Indonesia: Determinants of maternal mortality and stillbirth rates,” *PLoS ONE*, vol. 19, no. 7, p. e0303590, Jul. 2024.