Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2025, 2025:15
https://doi.org/10.28919/cmbn/9039
ISSN: 2052-2541

ASSESSING THE PERFORMANCE OF K-MEANS AND DBSCAN CLUSTERING METHODS IN TUBERCULOSIS MAPPING

DEFI YUSTI FAIDAH*, DIANDA DESTIN, FAZILA AZRA ANGGINA, MUHAMMAD IMAMUL CAESAR
Department of Statistics, Universitas Padjadjaran, Bandung 45363, Indonesia

Abstract: Tuberculosis (TB) remains a significant public health challenge, particularly in densely populated and under-resourced areas such as West Java, Indonesia. This study aimed to map and analyze the spatial distribution of TB prevalence by clustering regions based on variables including sanitation quality, population density, and TB rates. Secondary data from official sources were utilized, and clustering methods such as K-means and DBSCAN were employed to group districts and cities into distinct clusters. The K-means method identified five clusters, while DBSCAN formed four clusters with some noise. Performance evaluation using the silhouette index indicated that K-means outperformed DBSCAN. The clustering results revealed that regions with poor sanitation and high TB prevalence require prioritized public health interventions. This analysis underscores the potential of clustering methods to enhance public health planning by identifying areas in critical need of targeted TB control strategies, optimizing resource allocation, and supporting evidence-based decision-making.

Keywords: tuberculosis; K-means; DBSCAN; West Java.

2020 AMS Subject Classification: 62H30.

*Corresponding author
E-mail address: defi.yusti@unpad.ac.id
Received November 26, 2024

1. INTRODUCTION

Tuberculosis (TB) is an infectious disease transmitted through inhalation of droplets from an infected person's saliva. It colonizes the bronchioles or alveoli, with Mycobacterium tuberculosis being the bacterium responsible, first identified by Robert Koch [1]. While primarily affecting the lungs (pulmonary TB), it can also infect other organs (extrapulmonary TB) [2]. Common symptoms include coughing, chest pain, shortness of breath, loss of appetite, weight loss, fever, chills, and fatigue, with coughing being the most efficient way TB aerosols spread [3],[4].

Southeast Asia accounted for nearly half of the global TB cases in 2021, with a total of 4.82 million cases, representing 45.4% of the global total [5]. Eight countries account for approximately 66% of all global cases, with Indonesia (9.2%) being the second highest, following India [6]. Based on the Indonesian health profile in 2021, the number of detected TB cases reached 397,377. This represents an increase compared to 2020, when 351,936 cases were recorded. Provinces with large populations, such as West Java, East Java, and Central Java, reported the highest number of cases [5].

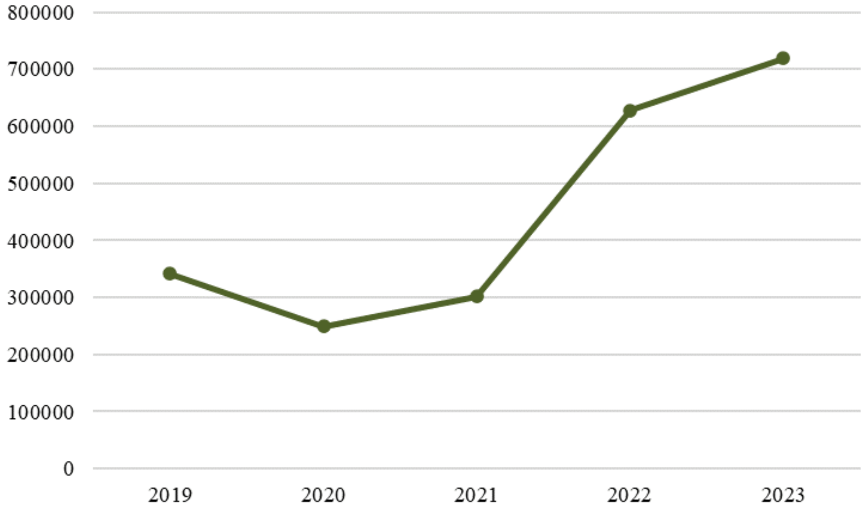

Figure 1. Total Cases of Tuberculosis in West Java 2019-2023

According to Figure 1, in 2023, 211,959 TB cases were reported in West Java out of a total of 718,704 suspected cases. This represents an increase compared to the previous year, when there were 627,481 suspected TB cases in 2022. Based on the annual report of the Tuberculosis

Control Program, West Java is the province with the highest number of TB cases in Indonesia, with the most affected areas being Bogor Regency, Bandung City, and Bekasi City [7]. There are several factors that influence the spread of TB in an area, such as proper sanitation and population density [6]. Previous studies have indicated that factors contributing to the high incidence of TB in urban areas include malnutrition, anemia, poor sanitation, and poverty [8].

In densely populated areas with unsuitable living conditions, poor sanitation is common, which facilitates the transmission of TB through the air when people cough or sneeze. In contrast, a clean and healthy environment with adequate sanitation can help reduce TB transmission. Tuberculosis is becoming an increasingly significant public health issue and must be treated promptly. To achieve this, it is crucial to identify the areas with the highest TB cases, allowing for targeted attention and prioritization in TB treatment. One method that can be used is K-Means clustering. The aim of this method is to categorize districts and cities in West Java into distinct clusters, thereby enhancing the effectiveness of TB management. It is hoped that the government can collaborate with relevant parties to reduce the number of TB cases in West Java [9].

2. MATERIALS AND METHODS

2.1. Data

The data used in this study is secondary data, obtained from data collection organizations such as the official website of the West Java Provincial Statistics Agency (https://jabar.bps.go.id/en) and Open Data Jabar (https://opendata.jabarprov.go.id/id). The variables analyzed in this study include the Prevalence Rate of Tuberculosis $\left(X_1\right)$, Percentage of Proper Sanitation $\left(X_2\right)$ and Population Density (Individuals per square kilometer) $\left(X_3\right)$.

2.1.1 Prevalence Rate of Tuberculosis

Prevalence rate of TB refers to the number of cases of TB within a population at some specific time points, denoted as the rate per 100.000 people, including cases of TB in HIV sufferers [10]. This study explores the prevalence rate of TB for every province in West Java.

$$Prevalence = \frac{Number\ of\ cases\ TB\ within\ a\ population}{Total\ individuals} x100\% \quad (1)$$

2.1.2. Percentage of Proper Sanitation

Water and sanitation are closely interconnected. Where there is clean water, there is also waste generated. In Indonesia, rivers serve as the primary source of clean water for a significant portion of the population. Unfortunately, the primary source of river water pollution in Indonesia is household or domestic waste. This underscores the strong correlation between water quality and sanitation standards, where water quality is largely determined by the quality of sanitation systems. Poorer sanitation conditions lead to a corresponding decline in water quality [11]. This situation also impacts environmental health, potentially leading to various diseases. This study focuses on the percentage of adequate sanitation in West Java [12].

2.1.3. Population Density

Population density is calculated from the number of individuals or inhabitants per square kilometer [13]. A large population is correlated with also a large possibility of encountering TB sufferers [12]. In this study, it was measured from every province in West Java.

2.2. Data Standardization

The process of standardization refers to the establishment of technical specifications that define a uniform design or set of criteria for a product, service, or process. These specifications are intended to ensure consistency and compatibility across different systems, products, or processes. Some standardization processes can be highly complex and technical, particularly when they are applied in industries that involve intricate designs or processes [14]. In the context of data, standardization is the practice of ensuring that datasets are aligned and conform to a consistent format or set of rules. When dealing with large datasets from various sources, synchronization challenges arise [15]. This is because data can come in different formats, units, or structures, making it difficult to integrate, analyze, or compare them without first standardizing them. The goal of data standardization is to streamline the process of data analysis by ensuring that all datasets are compatible and comparable, which is particularly crucial when managing large-scale datasets or conducting complex analyses [16]. Standardization at different levels such as metadata, data formats and licensing is essential to enable broad data integration, data exchange and interoperability with the overall goal to foster data-driven innovation [17].

2.3. VIF (Variance Inflator Factor)

In regression analysis, multicollinearity occurs when there is a high level of inter-correlation or inter-associations among the independent variables. Multicollinearity can be identified through the high value of Variance Inflation Factor (VIF), which represents the regression value of a particular predictor variable on other independent variables.

$$VIF = \frac{1}{1-R^2} \tag{2}$$

where R^2 is the coefficient of determination. In cases where the Variance Inflation Factor (VIF) is greater than 10, it indicates the presence of multicollinearity within the data [18].

2.4. K-Means Clustering

K-means clustering analysis is known for its easy and simple algorithm [19]. The aim in k- means clustering is to categorize a given dataset into k-distinct clusters, where the value of k is fixed in advance. The algorithm consists of two distinct phases, the first phase involves the definition of k centroids of each cluster. Meanwhile, the second phase is to allocate each point of the given data set to the nearest centroid [20]. The formulation of k-means clustering itself can be written as [16]:

$$J = \sum_{j=1}^{k}\sum_{i=1}^{n} x_i^{(j)} - C_j^2 \tag{3}$$

where k is the number of clusters, n is the number of observations, x_i is the i -th observation, and C_{ji} is the centroid of the cluster.

These following steps are the algorithm of K-means clustering [19]:

1. Choose the number of clusters (k) randomly.
2. K-mean cluster will divide data points into subsets and allocate the centroids to each subset regarding the number of clusters.
3. Choose the number of clusters for the second time and k-mean will use the same method as step number 2 to compute again.

4. Iterate steps 2 and 3 until the cluster arithmetic stops changing.

2.5. DBSCAN

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a popular clustering algorithm in machine learning and data analysis. The fundamental concept of Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is to identify a structure that accurately reflects the density distribution of a given dataset. Unlike K-means clustering, density-based clustering methods do not assume parametric distributions or rely on variance. This makes DBSCAN capable of discovering arbitrarily shaped clusters, handling varying levels of noise, detecting outliers in the data space, and operating without prior knowledge of the number of clusters. DBSCAN has been widely applied across various fields, including civil engineering, chemistry, spectroscopy, social sciences, medical diagnostics, remote sensing, computer vision, automatic identification systems, and anomaly detection [21][22].

These following steps are the algorithm of DBSCAN [23]:

1. Specify the Epsilon (Eps) and MinPoints (MinPts) values.
2. Determine the value of initial point (p) randomly.
3. Calculate the Eps value or calculate the distance of each point that has density to point p with the following Euclidean Distance formula

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \tag{4}$$

Where D_e is the euclidean distance, (x, y) is the data point, and (s, t) is the center point with i being the amount of data.

4. A cluster is formed if the point has enough epsilon more than the minimum points, then the point is the center point.
5. Repeat steps 3 and 4 until all points are counted. Proceed to the next point when there is no point that has density with respect to p or the initial point.

3. RESULTS

The clustering analysis and variable-based mapping for West Java were conducted using R Studio software. The variables analyzed included the percentage of proper sanitation, population density, and the TB prevalence rate. The results of this analysis, visualized in Figure 2, provide insights into the spatial distribution of TB across different regions in West Java, highlighting areas with varying levels of risk based on these key factors.

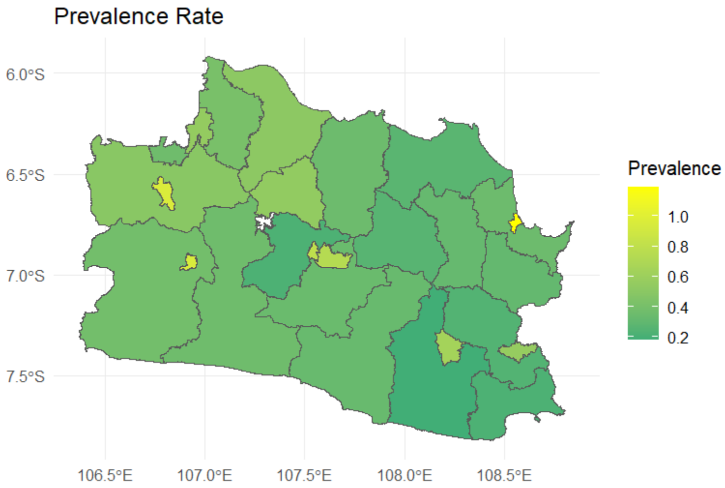

Figure 2. Mapping of District/City in West Java based on Prevalence Rate of Tuberculosis

Figure 2 illustrates the distribution of tuberculosis (TB) prevalence rates across various districts and cities in West Java. The data shows that Cirebon City has the highest TB prevalence rate, recorded at 1,194 cases per 100,000 population. This finding underscores a significant public health concern in the area, indicating a need for targeted interventions and resources to address the TB outbreak.

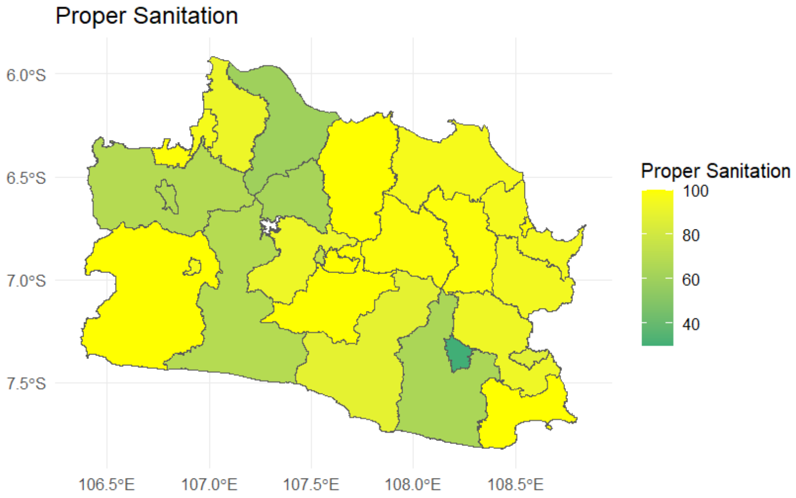

Figure 3. Mapping of District/City in West Java based on Proper Sanitation

Based on Figure 3, 19 areas in West Java have percentage of proper sanitation within 80-100%, 6 areas within 60- 80%, 1 area within 40-60%, and 1 area with the lowest percentage of proper sanitation which is Tasikmalaya City at 29.66%. 19 areas in West Java have percentage of proper sanitation within 80-100%, 6 areas within 60-80%, 1 area within 40-60%, and 1 area with the lowest percentage of proper sanitation which is Tasik City at 29,66%. There exist numerous conclusions that can be derived from the 2 figures above, including the correlation between the TB prevalence and proper sanitation. Figure 2 (the first map) shows the prevalence of TB, while figure 3 (the second map) represents proper sanitation coverage. Areas with higher TB prevalence (represented in yellow) seem to overlap with regions that have poorer sanitation conditions (also represented in yellow). This finding proves a potential correlation between poor sanitation and higher TB cases, which several studies have also shown that the deterioration of living and health conditions of a population contributes to the reproduction of TB disease [24]. In other words, areas with poor sanitation may facilitate the spread of TB due to poor hygiene practices and limited access to clean water, which can exacerbate health risks.

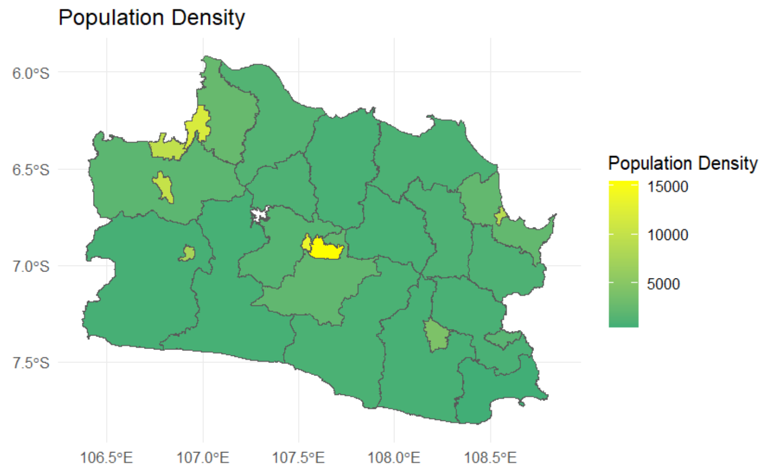

Figure 4. Mapping of District/City in West Java based on Population Density

The area with the highest population density in West Java is in Bandung City at 15.421 individuals / kilometer^2. The area with the highest population density in West Java is in Bandung City at 15.421 individuals / kilometer2. There is also a correlation between population density and TB prevalence. Figure 4 (third map) highlights the aspect of population density, where higher density areas (represented in yellow) appear to align with some of the areas that have a higher prevalence of TB. The reason is because high population density increases the likelihood of TB spreading through contact with the sufferers [25].

In conclusion, all of the variables, which include TB prevalence, proper sanitation, and population density, seem interconnected to each other. Areas with high population density and poor sanitation are likely to experience a higher prevalence of TB. Poor sanitation conditions may contribute to the health vulnerabilities of people in these densely populated areas, making it easier for diseases like TB to spread. Improving sanitation infrastructure and addressing overcrowding in these areas could be critical steps in reducing the TB burden in West Java.

3.1. Multicollinearity

Before starting clustering, a multicollinearity test needs to be performed to determine the relationships between variables using the Variance Inflation Factor (VIF) value. Multicollinearity occurs when two or more independent variables are highly correlated with each other, which can lead to instability in the estimation of regression coefficients and can make the model less reliable. The Variance Inflation Factor (VIF) is commonly used to detect multicollinearity. It quantifies how much the variance of a regression coefficient is inflated due to collinearity with other variables.

Table 1. VIF Value

Variabel	**VIF**
Prevalence of Tuberculosis	2.167
Proper Sanitation	1.105
Population Density	2.103

This step is essential before performing clustering because highly correlated variables can distort the results, making the clustering process less stable or less meaningful. Based on table 1, it can be found that no variable has a VIF value more than 10 so there is no multicollinearity, and all variables can be used.

3.2. K-Means

In the analysis using the k-means method, it is necessary to determine the number of clusters first. Determining the number of clusters uses the Average Silhouette Method to determine the optimal number of clusters.

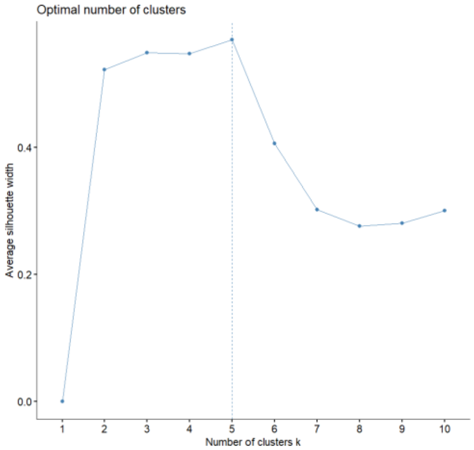

Figure 5. Optimal number of clusters

The average silhouette approach determines the observations of average silhouette for various values of k. The optimal number of clusters is 5 because it has the maximum average silhouette (0.5698) over a range of possible values for k. Cluster 1 consists of 3 sub-districts, cluster 2 consists of 4 sub-districts, cluster 3 consists of 1 sub-district, cluster 4 consists of 14 sub-districts, and cluster 5 consists of 5 sub-districts.

Table 2. Mean of Each Cluster

Cluster	**Prevalence of Tuberculosis**	**Proper Sanitation**	**Population Density**
1	0.55	98.5	12310
2	0.975	84.3	10042
3	0.64	29.7	4120
4	0.326	95.5	1233
5	0.416	64.9	1130

Cluster 2 shows the highest prevalence of tuberculosis with 0.975. Cluster 4 shows the lowest prevalence of 0.326. This indicates that the transmission rate is relatively low compared to other clusters. Cluster 1 has the best sanitation rate of 98.5%, indicating that almost all cluster areas have access to adequate sanitation. Cluster 1 also has a high population density of 12,310, indicating that the area is very dense. Clusters with high TB prevalence, such as Cluster 2, tend to have lower sanitation levels compared to clusters with lower prevalence. Cluster grouping can be seen in Figure 6.

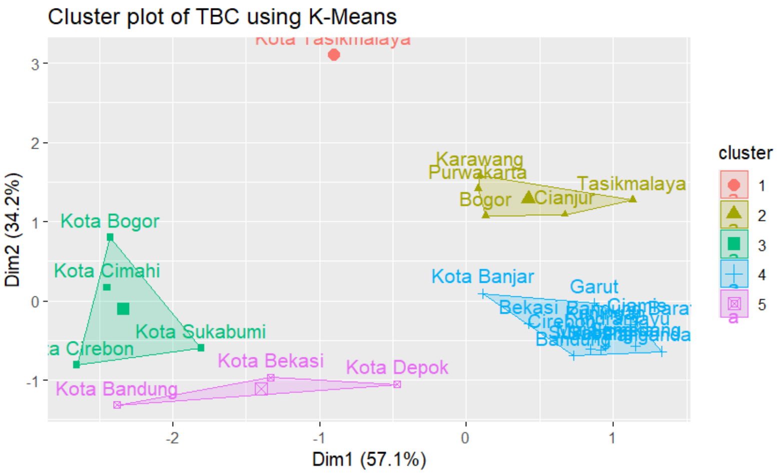

Figure 6. K-means visualization plot with $k = 5$

Figure 6 shows the grouping of districts/cities into 5 clusters that can be distinguished based on their color. Then the cluster mapping can be seen in figure 7.

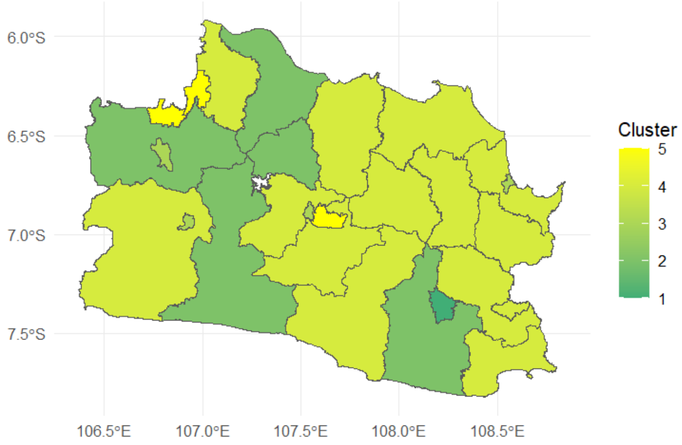

Figure 7. Cluster map of tuberculosis in West Java using K-means

Figure 7 above shows the distribution of TB clusters in West Java. Different colors indicate different clusters. The greener area means that the districts/cities are included in cluster 5. While the yellow color shows districts/cities that are in cluster 1.

3.4 DBSCAN

In the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm, the Epsilon (ε) and Minimum Points (MinPts) parameters are essential in determining clustering and identifying noise. In this study, MinPts = 2 is determined while Epsilon is generated using the KNN (K- Nearest Neighbors) method shown in figure 8.

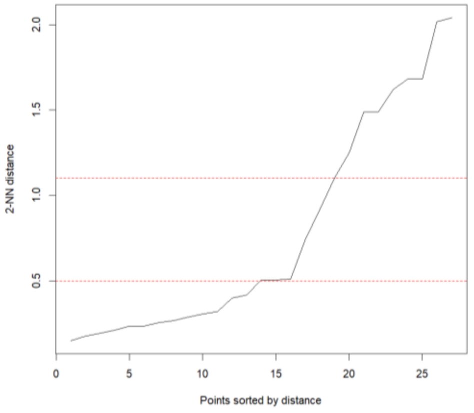

Figure 8. K-NN Distance plot

Based on figure 8, there is a sharp change around Eps 0.5 to 1.1. Therefore, a simulation should be conducted between Eps 0.5 to 1.1 with MinPts 2 which will be compared using the Silhouette Index.

Table 3. Simulation of Eps 0.5 – 1.1 with MinPts = 2

MinPts	**Eps**	**Silhouette**
2	0.50	0.3990577
	0.55	0.3990577
	0.60	0.4430330
	0.65	0.4430330
	0.70	0.4543092
	0.75	0.5031921
	0.80	0.5031921
	0.85	0.5031921
	0.90	0.4898660
	0.95	0.4898660
	1.00	0.4898660
	1.05	0.5134396
	1.10	0.4510421

Table 3 shows the largest Silhouette value of 0.5134 obtained from Eps = 1.05 and MinPts = 2. Therefore, in the study using DBSCAN, the parameters Eps = 1.05 and MinPts = 2 will be used to obtain 4 clusters and 4 noises.

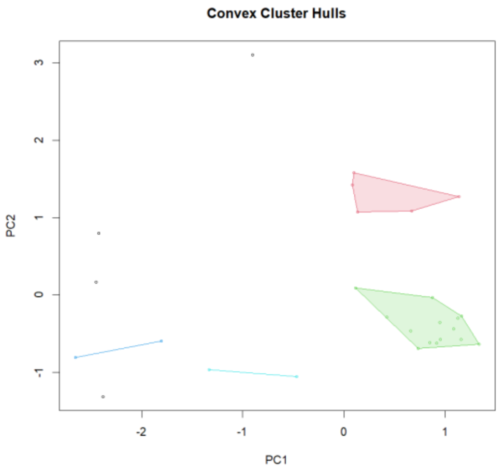

Figure 9. DBSCAN visualization plot with MinPts = 2 and Eps = 1.05

Figure 9 shows the clustering results using DBSCAN. Cluster 1 consists of 5 sub-districts, cluster 2 consists of 14 sub-districts, cluster 3 consists of 2 sub-districts, cluster 4 consists of 2 sub-districts, and 4 noisy sub-districts.

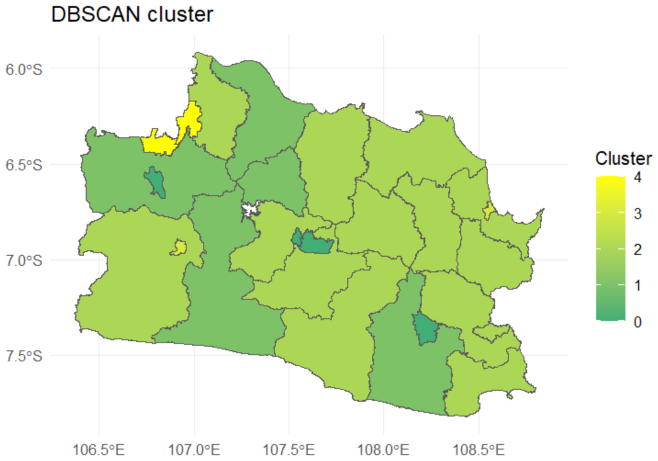

Figure 10. Cluster map of tuberculosis in West Java using DBSCAN

Cluster mapping in figure 10 shows the distribution of TB clusters in West Java using the DBSCAN method. The greenest color shows noise while the yellow color shows cluster 4. The two clustering methods, K-means and DBSCAN, will be compared using the silhouette index value. The highest value indicates the best method to use.

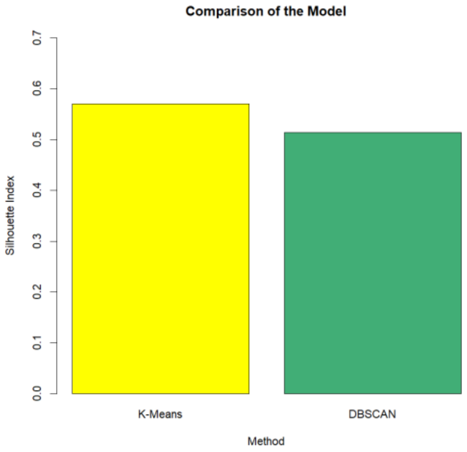

Figure 11. Comparison of silhouette index value

Based on figure 11, it is obtained that the K-means algorithm has a greater silhouette value than the DBSCAN algorithm, which is 0.5698. Therefore, in this study, the k-means algorithm produces a more optimal cluster than DBSCAN.

Table 4. Clustering and characteristic each cluster

Cluster	Characteristic	District/City
1	Exhibits moderate tuberculosis prevalence with very high population density and excellent access to proper sanitation.	Bandung City, Bekasi City, Depok City
2	Displays a very high tuberculosis prevalence, coupled with high population density and good access to proper sanitation.	Bogor City, Sukabumi City, Cirebon City, Cimahi City
3	Characterized by high tuberculosis prevalence, moderate population density, and very poor access to proper sanitation.	Tasikmalaya City
4	Shows low tuberculosis prevalence, low population density, and excellent access to proper sanitation.	Sukabumi, Bandung, Garut, Ciamis, Kuningan, Cirebon, Majalengka, Sumedang, Indramayu, Subang, Bekasi, Bandung Barat, Pangandaran, Banjar City
5	Features moderate tuberculosis prevalence, low population density, and moderate access to proper sanitation.	Bogor, Cianjur, Tasikmalaya, Purwakarta, Karawang

4. CONCLUSION

This study performs clustering analysis and variable-based mapping in West Java, focusing on proper sanitation, population density, and tuberculosis prevalence. The clustering methods that have been applied K-means and DBSCAN offered distinct insights into the spatial distribution of these variables across districts and cities. The K-means method, with an optimal number of five clusters, provided a clearer and more actionable grouping as indicated by its higher silhouette index value (0.5698), compared to DBSCAN (0.5134), which resulted in four clusters with some noise. The analysis revealed critical areas of concern, such as Cirebon City with the highest TB prevalence and Tasikmalaya City with the lowest sanitation levels, highlighting the need for targeted public health interventions. Ultimately, the K-means algorithm was found to be more effective in forming meaningful clusters, which can be leveraged for policymaking and resource allocation to combat tuberculosis and improve sanitation in West Java. Based on the cluster analysis, special attention should be directed towards Cluster 2 (comprising cities with very high TB prevalence and high population density) and Cluster 3 (characterized by high TB prevalence but very poor sanitation access), as these clusters are at high risk for disease spread and lack sufficient sanitation infrastructure. Efforts to improve sanitation and control TB should be prioritized in these areas. Continuous monitoring and periodic data analysis are also recommended to ensure that interventions remain relevant and effective over time.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests

REFERENCES

[1] T.N. Susilawati, R. Larasati, A Recent Update of the Diagnostic Methods for Tuberculosis and Their Applicability in Indonesia: A Narrative Review, Med. J. Indones. 28 (2019), 284–91. https://doi.org/10.13181/mji.v28i3.2589.

[2] N. Tandirogang, W.U. Mappalotteng, E.N. Raharjo, et al. The Spatial Analysis of Extrapulmonary Tuberculosis Spreading and Its Interactions with Pulmonary Tuberculosis in Samarinda, East Kalimantan, Indonesia, Infect. Dis. Rep. 12 (2020), 8727. https://doi.org/10.4081/idr.2020.8727.

[3] N.L.P.H. Wedari, I.W.A. Pranata, N.N.S. Budayanti, et al. Tuberculosis Cases Comparison in Developed Country (Australia) and Developing Country (Indonesia): A Comprehensive Review from Clinical, Epidemiological, and Microbiological Aspects, Intisari Sains Medis 12 (2021), 421–426. https://doi.org/10.15562/ism.v12i2.1034.

[4] G. Churchyard, P. Kim, N.S. Shah, et al. What We Know About Tuberculosis Transmission: An Overview, J. Infect. Dis. 216 (2017), S629–S635. https://doi.org/10.1093/infdis/jix362.

[5] S. Kaaffah, I.Y. Kusuma, F.S. Renaldi, et al. Knowledge, Attitudes, and Perceptions of Tuberculosis in Indonesia: A Multi-Center Cross-Sectional Study, Infect. Drug Resist. 16 (2023), 1787–1800. https://doi.org/10.2147/IDR.S404171.

[6] A.S. Azzahrain, A.N. Afifah, L.N. Yamani, Detection of Tuberculosis in Toddlers and Its Risk Factor at East Perak Health Center Surabaya, J. Kesehat. Lingkung. 15 (2023), 92–98. https://doi.org/10.20473/jkl.v15i2.2023.92-98.

[7] U. Nurjaman, O. Setiani, Sulistiyani, Risk Factors for the Incidence of Tuberculosis in Children at Sumedang District, West Java, Indonesia, Int. J. Adv. Res. 7 (2019), 612–618. https://doi.org/10.21474/IJAR01/9080.

[8] E. Karyadi, C.E. West, R.H. Nelwan, et al. Social aspects of patients with pulmonary tuberculosis in Indonesia, Southeast Asian J. Trop. Med. Public Health, 33 (2002), 338-345.

[9] D.Y. Faidah, G. Darmawan, B. Tantular, et al. Spatial Clusters and Determinants of the High Incidence of Diarrhea in Children, Commun. Math. Biol. Neurosci. 2024 (2024), 80. https://doi.org/10.28919/cmbn/8669.

[10] J. Wang, C. Wang, W. Zhang, Data Analysis and Forecasting of Tuberculosis Prevalence Rates for Smart Healthcare Based on a Novel Combination Model, Appl. Sci. 8 (2018), 1693. https://doi.org/10.3390/app8091693.

[11] A.S. Suryani, Pembangunan Air Bersih Dan Sanitasi Saat Pandemi Covid-19, Aspirasi: J. Masalah-Masalah Sos. 11 (2020), 199–214. https://doi.org/10.46807/aspirasi.v11i2.1757.

[12] M. Ardiyanti, S. Sulistyawati, Y. Puratmaja, Spatial Analysis of Tuberculosis, Population and Housing Density in Yogyakarta City 2017-2018, Epidemiol. Soc. Health Rev. 3 (2021), 28–35. https://doi.org/10.26555/eshr.v3i1.3629.

[13] I.F.U. Muzayanah, H.H. Lean, D. Hartono, K.D. Indraswari, R. Partama, Population Density and Energy Consumption: A Study in Indonesian Provinces, Heliyon 8 (2022), e10634. https://doi.org/10.1016/j.heliyon.2022.e10634.

[14] M.A. Lemley, Intellectual Property Rights and Standard-Setting Organizations, SSRN (2003). https://doi.org/10.2139/ssrn.310122.

[15] M. Schreiber, J. Metternich, Data Value Chains in Manufacturing: Data-Based Process Transparency through Traceability and Process Mining, Procedia CIRP 107 (2022), 629–634. https://doi.org/10.1016/j.procir.2022.05.037.

[16] E. Curry, A. Metzger, S. Zillner, et al. eds., The Elements of Big Data Value: Foundations of the Research and Innovation Ecosystem, Springer, Cham, 2021. https://doi.org/10.1007/978-3-030-68176-0.

[17] M. Gal, D.L. Rubinfeld, Data Standardization, SSRN (2019). https://doi.org/10.2139/ssrn.3326377.

[18] M.A. Raheem, N.S. Udoh, A.T. Gbolahan, Choosing Appropriate Regression Model in the Presence of Multicolinearity, Open J. Stat. 09 (2019), 159–168. https://doi.org/10.4236/ojs.2019.92012.

[19] A. Ali, C. Sheng-Chang, Characterization of Well Logs Using K-Mean Cluster Analysis, J. Pet. Explor. Prod. Technol. 10 (2020), 2245–2256. https://doi.org/10.1007/s13202-020-00895-4.

[20] K.A.A. Nazeer, M.P. Sebastian, Improving the Accuracy and Efficiency of the k-means Clustering Algorithm, in: Proceedings of the World Congress on Engineering, 2009.

[21] M. Hahsler, M. Piekenbrock, D. Doran, Dbscan: Fast Density-Based Clustering with R, J. Stat. Softw. 91 (2019), 1–30. https://doi.org/10.18637/jss.v091.i01.

[22] A.A. Bushra, G. Yi, Comparative Analysis Review of Pioneering DBSCAN and Successive Density-Based Clustering Algorithms, IEEE Access 9 (2021), 87918–87935. https://doi.org/10.1109/ACCESS.2021.3089036.

[23] B. Biantara, T. Rohana, A.R. Juwita, Perbandingan Algoritma K-Means dan DBSCAN untuk Pengelompokan Data Penyebaran Covid-19 Seluruh Kecamatan di Provinsi Jawa Barat, Sci. Stud. J. Inf. Technol. Sci. IV (2023), 88–94.

[24] F. Monteiro De Castro Fernandes, A.F. Couto Junior, J.U. Braga, et al. Environmental and Social Effects on the Incidence of Tuberculosis in Three Brazilian Municipalities and in Federal District, J. Infect. Dev. Ctries. 15 (2021), 1139–1146. https://doi.org/10.3855/jidc.13674.

[25] M. Ardiyanti, S. Sulistyawati, Y. Puratmaja, Spatial Analysis of Tuberculosis, Population and Housing Density in Yogyakarta City 2017-2018, Epidemiol. Soc. Health Rev. 3 (2021), 28–35. https://doi.org/10.26555/eshr.v3i1.3629.