Spatial Clustering of Child Malnutrition in Central Java: A Comparative Analysis Using K-Means and DBSCAN

Wiwien Hadikurniawati
Faculty of Information Technology
Satya Wacana Christian University
Salatiga, Indonesia
Faculty of Information Technology and Industry
Universitas Stikubank
Semarang, Indonesia
982022007@student.uksw.edu

Kristoko Dwi Hartomo
Faculty of Information Technology
Satya Wacana Christian University
Salatiga, Indonesia
kristoko@uksw.edu

Irwan Sembiring
Faculty of Information Technology
Satya Wacana Christian University
Salatiga, Indonesia
irwan@uksw.edu

Abstract— ***The issue of malnutrition in children poses a serious challenge in the effort to achieve the well-being and development of a smart generation of children. Central Java is a province on the island of Java with the highest prevalence of stunting, so efforts for improvement and health intervention planning need to focus on areas with malnutrition among children in Central Java. This research aims to identify the spatial patterns and distribution of stunted children in the 35 districts and cities in Central Java using clustering techniques. The data used includes the nutritional status of children in all districts and cities in Central Java. Two clustering methods, K-Means and DBSCAN, were applied to identify groups of districts/cities with stunting characteristics. K-Means resulted in three clusters: low stunting prevalence (11 districts/cities), moderate (18 districts/cities), and high (6 district/cities). DBSCAN grouped 21 districts/cities into one main cluster and identified 14 other districts/cities as outliers. In this study, K-Means outperformed DBSCAN, with a higher Silhouette score (0.403) and a lower Davies-Bouldin Index (0.785).***

Keywords—clustering, K-Means, DBSCAN, malnutrition

I. INTRODUCTION

Malnutrition, referring to insufficient food intake and the presence of infectious diseases, is a major cause of child morbidity and mortality in developing countries [1]. Short-term stunting can have a negative impact on brain growth, intelligence, physical development, and metabolic health. Long-term impacts include decreased cognitive skills, weakened immunity, and increased risk of chronic disease [2]– [4]. This increase can have an impact on the quality of human resources, productivity, and national competitiveness in Indonesia [5].

Stunting, or growth disorders in children characterized by height below age standards, is one of the main indicators of chronic malnutrition. Stunting is an indicator of chronic malnutrition and can affect children's growth and development [6]. Children who experience stunting in the first two years of life have a higher risk of experiencing cognitive impairment, low academic achievement, and low work productivity in adulthood [7].

The prevalence of stunting tends to decrease every year, but it is far from the target set by WHO [2][4]. Indonesia still has a high stunting prevalence rate of 36%. Indonesia's densest population and the largest number of children under five are on the island of Java, which consists of 6 provinces. The six provinces, Central Java, have a stunting prevalence of 20.8%, ranking last for the highest stunting prevalence. This data was obtained based on the Indonesian Nutrition Status Survey by the Ministry of Health in 2022. Provinces on the island of Java that have the lowest prevalence of stunting according to the best ranking: DKI Jakarta (14.8%), DI Yogyakarta (16.4%), East Java (19.2%), Banten (20%), West Java (20.2%) and Central Java (20.8%). Central Java, as one of the provinces with the largest population in Indonesia, faces serious challenges in overcoming the problem of stunting. Geographic variability, socio-economic differences and other factors may contribute to differences in stunting prevalence in various districts/cities in this province. Therefore, it is important to understand the distribution pattern of stunting at the district/city level.

This research aims to identify and understand stunting distribution patterns in districts/cities in Central Java by applying clustering techniques. By identifying geographic clusters of stunting prevalence, stakeholders can plan and allocate resources more efficiently for nutrition interventions. In addition, identifying areas that have a high prevalence of stunting can assist the government in developing more targeted strategies.

Clustering categorizes data into groups based on similarities using specific criteria, offering high precision. This algorithm's maximal and linear nature ensures fast execution and ease of tuning. Identifying geographic stunting clusters enables more efficient resource allocation for nutrition interventions and informs targeted strategy development.

Clustering techniques can map areas that have different stunting prevalence. Clustering can be solved with various algorithms. This research uses the K-Means and DBSCAN algorithms to cluster the nutritional status of toddlers in districts/cities in Central Java. A comparison of the K-Means and DBSCAN algorithms was chosen to cluster the nutritional status of children under five because it provides an interesting perspective, especially on the characteristics and data requirements:

a. Cluster structure

The distribution of data on stunting data is not always hyperbolic or round.

b. Noise

Health data may contain noise or outliers, such as areas with inconsistent reports.

Although much research has been conducted on stunting in Indonesia, very little has focused on geographic cluster analysis at the district/city level, especially in Central Java. In addition, comparisons between different clustering methods, such as K-Means and DBSCAN, in the context of stunting are still rarely carried out. It is hoped that the results of this research can contribute to efforts to reduce stunting cases in Central Java and improve the quality of life of children in the area.

This raises the question of which methods are most effective in identifying high-risk areas and provide the most insight for intervention.

II. RELATED WORK

Clustering is a valuable technique in data mining that is extensively researched and applied in various fields such as computer science, data science, statistics, pattern recognition, artificial intelligence, and machine learning [8]-[10]. There have been many studies using clustering techniques, especially the k-means [11]– [13] and DBSCAN algorithms [14]– [18].

K-Means is a popular, simple clustering algorithm that divides data into clusters by iteratively assigning points to the nearest cluster centroid and updating the centroids. It's widely used in image segmentation, customer segmentation, and more, with readily available resources and implementations in various languages [19][20].

The K-Means algorithm is used to analyze and categorize various aspects of English learning, such as students' proficiency levels or learning preferences. This approach enables them to tailor the teaching system to individual needs, providing personalized resources and support to enhance the students' learning experience [11].
The K-Means algorithm is also used to analyze and classify rural areas into different levels of development based on their electricity consumption [12].

DBSCAN is a widely used clustering algorithm that groups nearby data points based on density. It's unique for not requiring prior cluster knowledge or specific shapes. It identifies variously shaped clusters, detects outliers, and is robust in noisy data. DBSCAN can discover clusters of arbitrary shapes by analyzing density, effectively distinguishing noise from valid clusters to maintain clustering accuracy [15][16][21].

A modified version of DBSCAN has been developed to introduce a reduced set known as the operational set. This subset is periodically updated and used to accurately calculate sample density, eliminating the need to compute distances between all pairs in the dataset [22]. The DBSCAN clustering algorithm is used to classify students based on their characteristics and learning status [17]. The DBSCAN algorithm is used to identify earthquake clustering areas and determine the main regions where earthquakes often occur in Indonesia [14].

III. METHOD

The data used in this research was obtained from the Central Java provincial health survey in 2022-2023. The data includes information regarding the prevalence of stunting in 35 districts/cities in Central Java, including the number of children under five with nutritional status based on gender and age. The data obtained includes data on children under five, male and female, aged 0-24 months and 0-59 months. Other data is data on height based on age which consists of the categories very short, short, stunted, normal, tall. In addition, weight data based on toddlers' height consists of the categories of malnutrition, undernutrition, normal, risk of overnutrition, overnutrition and obesity.

Data clustering using the K-Means algorithm is completed in stages [23]:

a. Determine the number of clusters.
b. Data are allocated into clusters randomly
c. Determine the cluster center from the data in each cluster. If M is the amount of data in the cluster, I is the number of clusters and p is the data dimension, then the centroid of the 1st feature is determined by:

$$K^{i} = \frac{1}{M}\sum_{j=1}^{M} X_j \quad (1)$$

K^i = 1st centroid, M = amount of data in the cluster
X_j = point to j

d. Allocate each data to the nearest centroid. Calculating the distance to the cluster center with Euclidean Distance

$$d = \sqrt{(X_1 - X_2)^2 + (C_1 - C_2)^2} \quad (2)$$

d=Euclidean distance; X_i =point to i; C_j=centroid to j

The number of clusters generated by DBSCAN will depend on the parameter's eps and min samples [21]. Euclidean distance is used for the clustering process by calculating the distance between a point and the centroid (C) that has been randomly determined previously. The formula is written in equation 3:

$$D(X_j, C_i) = \sqrt{\sum_{j=1}^{q}(X_{ij} - C_{ij})2} \quad (3)$$

D = distance from point X to C, X_{ij} = point value 1 in the j^{th} cluster, C_{ij} = centroid value i in the j^{th} cluster
Testing the validity of the grouping method using:
1. Silhouette coefficient which is expressed as:

$$S(h) = \frac{b(h)-a(h)}{(a(h),b(h))} \quad (4)$$

The value a(h) is the average distance between the core point and all points in the same group, and the value b(h) is the average distance between the core point and all points in different groups using Euclidean distance. The Silhoutte coefficient produces a value range of -1 to 1. The better the data grouping in one cluster, the closer the result is to 1 [24][25].

2. Davies-Bouldin Index

The Davies-Bouldin Index (DBI) is an evaluation metric for clustering that measures the average similarity between each cluster and the cluster that is most similar to it. Similarity in this context is defined as the ratio of the distance within a cluster to the distance between clusters. A lower DBI value indicates a better clustering model. Formula DBI:

$$DBI = \frac{1}{n}\sum_{i-1}^{n} max \quad i \neq j(\frac{S_i+S_j}{d_{ij}}) \quad (5)$$

n = number of clusters, Si = the average distance between each point in cluster (i) to the centroid of cluster (i), dij = the distance between the centroids of clusters (i) and (j).

Research steps:
a. Normalization

Data cleaning addressed inconsistencies and missing values. Clustering, like K-Means, relies on uniform feature scales. Min-Max Scaling was used to ensure equal feature weight in the analysis.

b. Determination of Number of Clusters
c. Application of Clustering Algorithm

Two clustering methods, K-Means and DBSCAN, were used. K-Means requires determining the number of clusters, which was done using the Elbow method, resulting in three optimal clusters. Districts/cities were then grouped into one of these clusters based on their stunting characteristics (Eq 1 and Eq 2).

In the DBSCAN method, the parameters `eps` and min_samples are selected based on data density analysis (Eq 3). DBSCAN produced one main cluster and several districts/cities identified as outliers.

d. Visualisation and Interpretation
e. Clustering Evaluation

The quality of clustering is evaluated using the Silhouette coefficient and Davies-Bouldin Index. These coefficients measure how similar an object is to its cluster compared to other clusters.

IV. RESULT AND ANALYSIS

The data has been normalized, so now each feature has a mean of 0 and a standard deviation of 1. This normalization is important to ensure that all features have the same weight when clustering. The next step is to determine the number of clusters using the Elbow method. In this method, the K-Means algorithm with various numbers of clusters and measures the within-cluster sum of squares (WCSS). The angled points on the WCSS plot indicate the optimal number of clusters. In this research, 3 clusters are the optimal number for determining clusters (Fig 1).

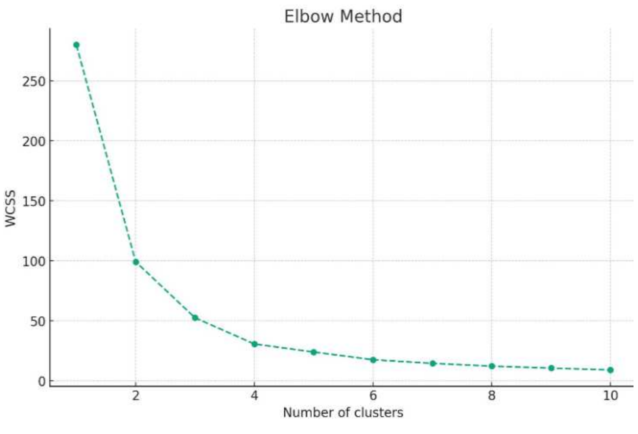

Figure 1. Determine number of K-Means cluster

The clustering results using the K-Means and DBSCAN algorithms for severely lacking and lacking nutrition status in children aged 0-23 months and 0-59 months, classified by gender (female and male), can be seen in Table 1.

TABLE 1. Clustering Results using K-Means and DBSCAN

No	A	B	C	D	E	F	G	H	K-Means Cluster	DBSCAN Cluster
1	136	87	318	215	1224	953	3337	2600	1	-1
2	140	84	272	165	873	633	2072	1577	1	-1
3	79	45	185	142	437	290	1199	818	0	1
4	46	28	77	55	493	315	1153	765	2	-1
5	21	16	37	22	650	511	1729	1241	0	-1

Table caption:

A. Severely Lacking 0 - 23 Months Male
B. Severely Lacking 0 - 23 Months Female
C. Severely Lacking 0 - 59 Months Male
D. Severely Lacking 0 - 59 Months Female
E. Lacking 0 – 23 Months Male
F. Lacking 0 – 23 Months Female
G. Lacking 0 – 59 Months Male
H. Lacking 0 – 59 Months Female

In the application of the K-Means method, the nutritional status data of children under five are divided into three different clusters:

1. Cluster 0: Consists of 18 districts/cities, representing areas with moderate stunting prevalence.
2. Cluster 1: Consists of 6 districts/cities, representing areas with high stunting prevalence.
3. Cluster 2: Consists of 11 districts/cities, representing areas with low stunting prevalence.

Cluster 0 shows that the districts/cities in this cluster have a relatively moderate number of children under five with severely lacking and lacking status compared to other clusters. They are in the middle between clusters 1 and 2 in terms of the number of children under five with that status. Cluster 1 shows that the districts/cities in this cluster have the highest number of children under five with severely lacking and lacking status among all clusters. This may indicate areas with a higher prevalence of stunting or malnutrition. Cluster 2 shows that the districts/cities in this cluster have the lowest number of children under five with the status of severely lacking and lacking. This may indicate areas with relatively better nutritional conditions for children under five.

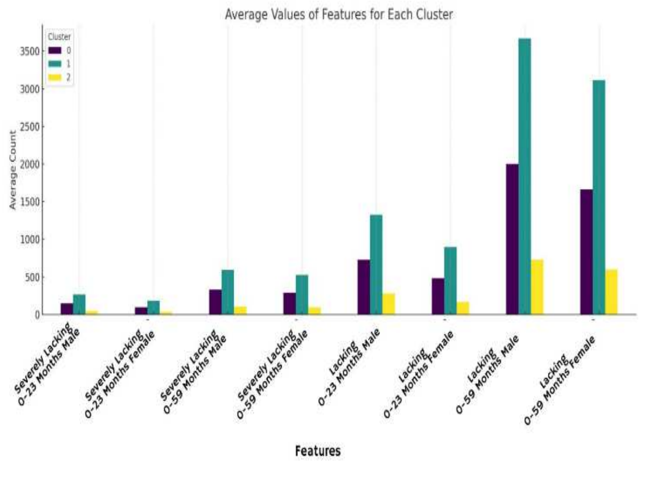

Figure 2. Clustering using K-Means

Figure 2 explains that Cluster 0 areas include: Purbalingga District (D), Banjarnegara D, Wonosobo D, Magelang, Boyolali D, Sukoharjo D, Wonogiri D, Sragen D, Grobogan D, Blora D, Rembang D, Pati D, Semarang D, Temanggung D, Kendal City (C), Pekalongan C, Pemalang C, Brebes C. Districts in this cluster have a number of children under five with severely lacking and lacking statuses that are in the middle between clusters 1 and 2. Although they have a lower prevalence of stunting compared to cluster 1, they still have a number of children under five with this status. which is quite significant. Cluster 1 includes: Cilacap D, Banyumas D, Kebumen D, Klaten D, Jepara D, Tegal D. Cluster 2 includes: Purworejo D, Karanganyar D, Kudus D, Demak D, Batang D, Magelang C, Surakarta C, Salatiga C, Semarang C, Pekalongan C, Tegal C.

As a non-parametric clustering algorithm, DBSCAN does not require determining the number of clusters first, but it is necessary to determine several main parameters when using DBSCAN. These parameters can be determined:

1. eps: the maximum distance between two samples for one to be a neighbor of the other.
2. min_samples: the minimum number of samples in the environment to consider a data point as a center point.

The choice of parameter values can have a substantial impact on the clustering results. One method for determining the eps value is by utilizing the Nearest Neighbors Distance technique. When examining the Nearest Neighbors Distance plot, we search for points where a significant increase in distance occurs, resembling an elbow. This point can be regarded as the optimal value for eps. In the chart, it appears that the elbow point is located around 1.0 - 1.5 (Fig 3).

The value of -1 in the DBSCAN clusters (Table 1) indicates that the data point is considered noise or an outlier by the algorithm and does not belong to any cluster. In these results, it appears that many data points are considered as noise.

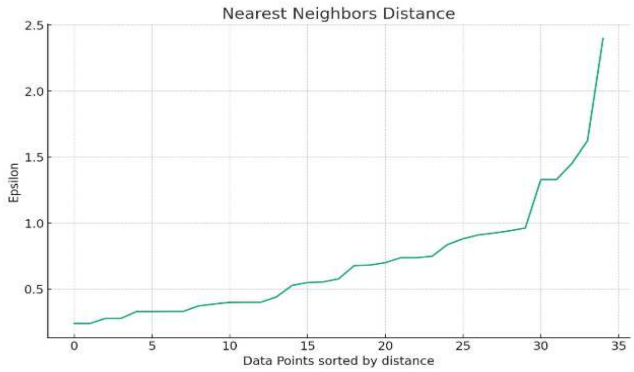

Figure 3. Determine number of DBSCAN cluster

DBSCAN classified 21 districts/cities into one main cluster and 14 other districts/cities were identified as outliers, showing unique or different stunting characteristics from the majority of other regions. Clustering results with DBSCAN show two clusters:

1. Cluster 0: Consists of 21 districts/cities.

Cluster 0 includes: Purbalingga D, Banjarnegara D, Purworejo D, Wonosobo D, Boyolali D, Sukoharjo D, Wonogiri D, Karanganyar D, Sragen D, Grobogan D, Blora D, Rembang D, Kudus D, Semarang D, Temanggung D, Kendal C, Batang C, Pekalongan C, Pemalang C, Brebes C, Semarang C. Districts/cities in this cluster have a number of children under five with severely lacking and lacking status which is below the outlier cluster average (-1). This shows that the prevalence of stunting in this area is lower than the outlier cluster

2. Cluster -1: Consists of 14 districts/cities.

A cluster value of -1 indicates data points that are considered noise or outliers by DBSCAN. Cluster -1 includes: Cilacap D, Banyumas D, Kebumen D, Magelang D, Klaten D, Pati D, Jepara D, Tegal D, Magelang C, Surakarta C, Salatiga C, Pekalongan C, Tegal C. Districts/cities considered as outliers by DBSCAN show a higher prevalence of stunting. They have a much higher number of children under five with severely lacking and insufficient status compared to cluster 0. This shows that these areas may need more attention and nutritional intervention. DBSCAN, on the other hand, identifies most of the data as one main cluster and some other data as outliers, which provides a different perspective on the data distribution.

To visualize clusters, it is necessary to reduce the data dimensionality to 2 dimensions using Principal Component Analysis (PCA) dimension reduction technique. The steps are as follows:

1. Dimension Reduction: Using PCA to reduce the data dimension to 2 dimensions.
2. Cluster Visualization: Visualizing the clustering results using K-Means and DBSCAN.

The cluster visualization results are shown in Figure 4.

These insights show that although DBSCAN considers some districts/cities to be outliers, they actually have characteristics that are very important in the context of stunting. Thus, clustering with DBSCAN provides a different perspective compared to K-Means. This is very useful for identifying areas that require special attention. The evaluation of this research's clustering is performed by calculating the Silhouette Score and Davies-Bouldin Index.

The evaluation results with the Silhouette Score for both clustering methods are as follows:

a. The K-Means Clustering: Silhouette Score of (0.403). This indicates that K-Means did a reasonably good job in grouping the data, as the score is closer to 1.
b. The DBSCAN Clustering: Silhouette Score of (0.204). While this score is lower compared to K-Means, it still shows that objects within the cluster have better similarity to other objects in the same cluster than to objects in different clusters.

The visualization of the Silhouette Score can be seen in Figure 5 and 6.

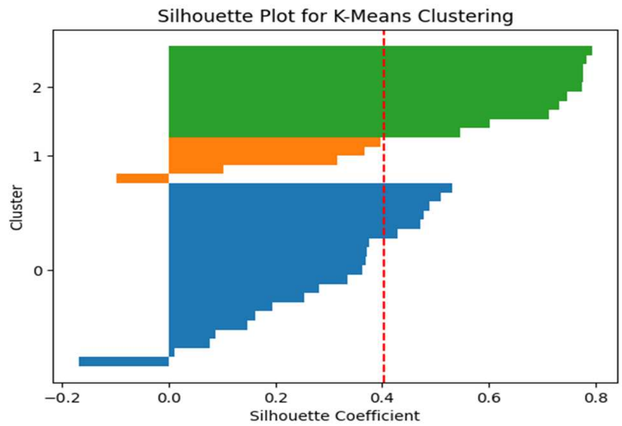

Figure 5. Visualization of the Silhouette Score for K-Means

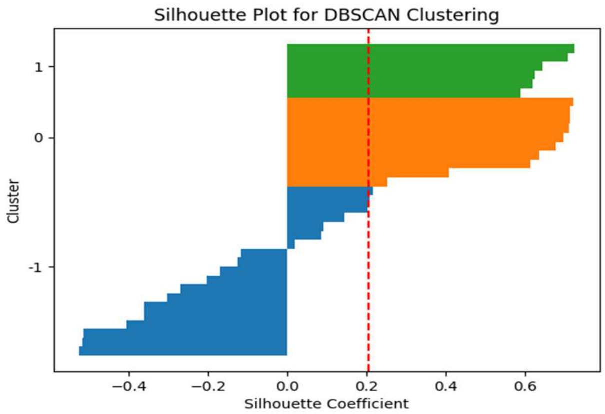

Figure 6. Visualization of the Silhouette Score for DBSCAN

The Davies-Bouldin Index results for both clustering methods are as follows: the K-Means clustering has a Davies-Bouldin Index of 0.785, and DBSCAN clustering has a Davies-Bouldin Index of 1.067. In the Davies-Bouldin Index, a lower value indicates better clustering performance. Based on this metric, the K-Means provides better clustering results compared to DBSCAN for these children under five stunting dataset. The complete clustering evaluation results with the K-Means and DBSCAN algorithms are presented in Table 2

TABLE 2. Results of K-Means and DBSCAN Clustering Evaluation.

	Silhouette Score	Davies-Bouldin Index
K-Means	0.403	0.785
DBSCAN	0.204	1.067

The evaluation results in table 2 indicate that K-Means is a superior algorithm compared to DBSCAN for clustering child malnutrition in Central Java. The K-Means algorithm has a higher Silhouette score than DBSCAN (0.403) and a lower Davies-Bouldin Index than DBSCAN (0.785).

Although K-Means provides a clearer division between 3 clusters with low, moderate, and high stunting prevalence, the DBSCAN provides additional insight by identifying districts/cities that have unique stunting prevalence as outliers. The identification of districts/cities as outliers by DBSCAN provides additional perspective. For example, the city of Surakarta is considered an outlier due to its very low prevalence of children under five with severely lacking status, indicating the possibility of successful nutrition improvement interventions in the city.

V. CONCLUSION

This paper presents an analysis of child nutrition status mapping in all districts and cities in Central Java Province by comparing the K-Means and DBSCAN clustering algorithms. The clustering methods used are the K-Means and DBSCAN algorithms, evaluated using the silhouette score and Davies-Bouldin Index parameters. The evaluation results with these two parameters indicate that the K-Means algorithm outperforms DBSCAN in this study.

Simulation results show that the K-Means algorithm identifies three main clusters, while the DBSCAN algorithm divides the data into two clusters. This research provides a valuable contribution to the government for planning and implementing more effective nutrition interventions in Central Java.

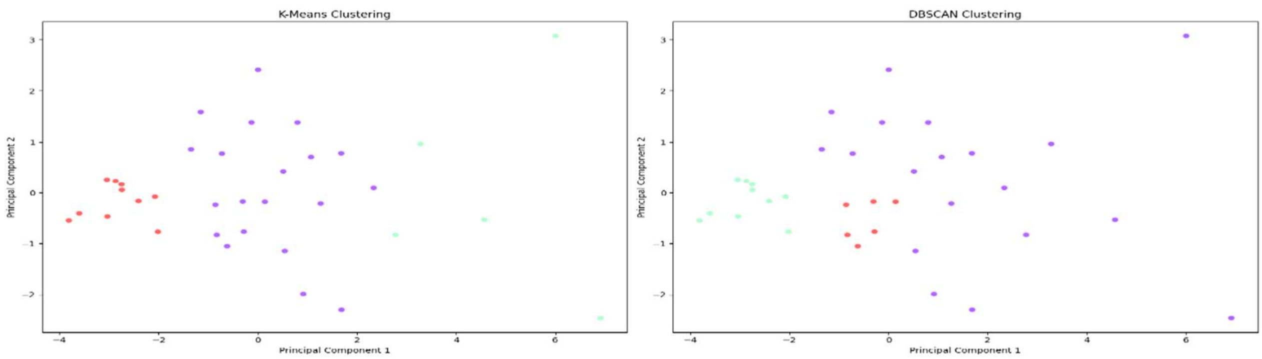

Figure 4. The Cluster Visualization Results

REFERENCES

[1] J. D. Moyer, D. K. Bohl, C. Petry, A. Scott, J. R. Solórzano, and R. Kuhn, "The persistent global burden of severe acute malnutrition: Cross-country estimates, models and forecasts," *Glob Transit*, vol. 2, pp. 167–179, Jan. 2020, doi: 10.1016/j.glt.2020.07.004.

[2] A. Saleh, S. Syahrul, V. Hadju, I. Andriani, and I. Restika, "Role of Maternal in Preventing Stunting: a Systematic Review," *Gac Sanit*, vol. 35, pp. S576–S582, Jan. 2021, doi: 10.1016/j.gaceta.2021.10.087.

[3] S. H. Quamme and P. O. Iversen, "Prevalence of child stunting in Sub-Saharan Africa and its risk factors," *Clinical Nutrition Open Science*, vol. 42. Elsevier B.V., pp. 49–61, Apr. 01, 2022. doi: 10.1016/j.nutos.2022.01.009.

[4] C. Berti and A. La Vecchia, "Temporal trend of child stunting prevalence and Food and Nutritional Surveillance System," *J Pediatr (Rio J)*, 2022, doi: 10.1016/j.jped.2022.10.001.

[5] W. Hadikurniawati, K. Dwi Hartomo, I. Sembiring, H. Dwi Purnomo, and A. Iriani, "Triangular fuzzy numbers-based MADM for selecting pregnant mothers at risk of stunting," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 3, pp. 579–585, Jun. 2023, doi: 10.29207/resti.v7i3.4966.

[6] V. T. Siy Van *et al.*, "Predicting undernutrition among elementary schoolchildren in the Philippines using machine learning algorithms," *Nutrition*, vol. 96, Apr. 2022, doi: 10.1016/j.nut.2 021.111571.

[7] H. Shi *et al.*, "Explainable machine learning model for predicting the occurrence of postoperative malnutrition in children with

congenital heart disease," *Clinical Nutrition*, vol. 41, no. 1, pp. 202–210, Jan. 2022, doi: 10.1016/j.clnu.2021.11.006.

[8] K. L. H. S. Perera and M. D. R. L. Silva, "A Comparative Analysis of Clustering Techniques with Feature Selection for Predicting Breast Cancer Recurrence," in *21st International Conference on Advances in ICT for Emerging Regions, ICter 2021 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 81–86. doi: 10.1109/ICter53630.2021.9774799.

[9] V. Chang, M. A. Ganatra, K. Hall, L. Golightly, and Q. A. Xu, "An assessment of machine learning models and algorithms for early prediction and diagnosis of diabetes using health indicators," *Healthcare Analytics*, vol. 2, p. 100118, Nov. 2022, doi: 10.1016/j.health.2022.100118.

[10] A. E. Ezugwu *et al.*, "A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects," *Engineering Applications of Artificial Intelligence*, vol. 110. Elsevier Ltd, Apr. 01, 2022. doi: 10.1016/j.engappai.2022.104743.

[11] S. Gao, C. Feng, and Y. Feng, "Design of English Teaching System Based on KMean Clustering Algorithm," in *2nd IEEE International Conference on Advanced Technologies in Intelligent Control, Environment, Computing and Communication Engineering, ICATIECE 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICATIECE56365.2022.10047138.

[12] P. Li, H. Li, S. Wang, W. Zu, H. Zhang, and J. Wang, "Rural Developing Level Clustering Based on KMEANS from Electricity Perspective," in *2021 IEEE Conference on Telecommunications, Optics and Computer Science, TOCS 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 324–328. doi: 10.1109/TOCS53301.2021.9688988.

[13] A.Feng, "Automotive Product Analysis Based on MP-DP-Kmeans Clustering," in *2023 6th International Conference on Artificial Intelligence and Big Data, ICAIBD 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 305–311. doi: 10.1109/ICAIBD57115.2023.10206277.

[14] Mustakim, E. Rahmi, M. R. Mundzir, S. T. Rizaldi, Okfalisa, and I. Maita, "Comparison of DBSCAN and PCA-DBSCAN Algorithm for Grouping Earthquake Area," in *2021 International Congress of Advanced Technology and Engineering, ICOTEN 2021*, Institute of Electrical and Electronics Engineers Inc., Jul. 2021. doi: 10.1109/ICOTEN52080.2021.9493497.

[15] Z. Zhou, L. Luo, Z. Li, M. Chao, W. Tang, and Y. Wang, "Harmonic pollution zoning method based on improved DBSCAN clustering," in *China International Conference on Electricity Distribution, CICED*, IEEE Computer Society, 2022, pp. 1099–1103. doi: 10.1109/CICED56215.2022.9929020.

[16] D. Deng, "Research on Anomaly Detection Method Based on DBSCAN Clustering Algorithm," in *Proceedings - 2020 5th International Conference on Information Science, Computer Technology and Transportation, ISCTT 2020*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020, pp. 439–442. doi: 10.1109/ISCTT51595.2020.00083.

[17] H. Du, S. Chen, H. Niu, and Y. Li, "Application of DBSCAN clustering algorithm in evaluating students' learning status," in *Proceedings - 2021 17th International Conference on Computational Intelligence and Security, CIS 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 372–376. doi: 10.1109/CIS54983.2021.00084.

[18] L. Jing and M. Guo, "Weighted DBSCAN Algorithm for Discovery of Spatial Features in Industries Management," in *IMCEC 2021 - IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference*, Institute of Electrical and Electronics Engineers Inc., Jun. 2021, pp. 383–386. doi: 10.1109/IMCEC51613.2021.9482143.

[19] A. Feng, "Automotive Product Analysis Based on MP-DP-Kmeans Clustering," in *2023 6th International Conference on Artificial Intelligence and Big Data, ICAIBD 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 305–311. doi: 10.1109/ICAIBD57115.2023.10206277.

[20] L. E. Bule and N.-K. C. Nair, "Smart Meter Customer Data Synchronization Technique by KMean Clustering," in *2023 IEEE PES GTD International Conference and Exposition (GTD)*, IEEE, May 2023, pp. 409–413. doi: 10.1109/GTD49768.2023.00103.

[21] D. Deng, "DBSCAN Clustering Algorithm Based on Density," in *Proceedings - 2020 7th International Forum on Electrical Engineering and Automation, IFEEA 2020*, Institute of Electrical and Electronics Engineers Inc., Sep. 2020, pp. 949–953. doi: 10.1109/IFEEA51475.2020.00199.

[22] N. Hanafi and H. Saadatfar, "A fast DBSCAN algorithm for big data based on efficient density calculation," *Expert Syst Appl*, vol. 203, Oct. 2022, doi: 10.1016/j.eswa.2022.117501.

[23] A. Yusuf, "K-Means Clustering Based on Distance Measures: Stunting Prevalence Clustering in South Kalimantan," in *2022 5th International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 706–710. doi: 10.1109/ISRITI56927.2022.10052925.

[24] G. Jin-Heng, L. Jia-Xiang, Z. Zhen-Chang, and L. Han-Yu, "CDBSCAN: Density clustering based on silhouette coefficient constraints," in *Proceedings - 2022 International Conference on Computer Engineering and Artificial Intelligence, ICCEAI 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 600–605. doi: 10.1109/ICCEAI55464.2022.00128.

[25] H. B. Tambunan, D. H. Barus, J. Hartono, A. S. Alam, D. A. Nugraha, and H. H. H. Usman, "Electrical peak load clustering analysis using K-means algorithm and silhouette coefficient," in *Proceeding - 2nd International Conference on Technology and Policy in Electric Power and Energy, ICT-PEP 2020*, Institute of Electrical and Electronics Engineers Inc., Sep. 2020, pp. 258–262. doi: 10.1109/ICT-PEP50916.2020.9249773.