VOL. 11. NO. 1 AUGUST 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v11i1.7007

OPTIMIZATION OF STUNTING INFANT DATA CLUSTERING WITH K-MEANS++ ALGORITHM USING DBI EVALUATION

Efmi Maiyana[1*]; Wizra Aulia[1]

Informatics Management Study Program[1]
AMIK Bukittinggi, Padang, Indonesia[1]
https://amikboekittinggi.ac.id/[1]
efmi_maiyana@yahoo.com*; wizra.ira23@gmail.com

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *Stunting in infants is a serious health issue, particularly in developing countries like Indonesia. This study aims to optimize the clustering of stunting data in infants using the K-Means++ algorithm, evaluated with the Davies-Bouldin Index (DBI) to determine the optimal number of clusters. The stunting data includes variables such as age, gender, weight, and height. The analysis results indicate that the optimal number of clusters is 5, with a DBI value of 0.837986204, confirming the quality of the clustering. This conclusion demonstrates that the combination of these evaluation methods produces effective clustering and provides significant insights into identifying groups of infants with varying stunting risk levels. These findings can serve as a basis for more targeted health interventions in addressing stunting.*

Keywords: *Clustering, Davies-Bouldin Index (DBI), Optimization, Stunting.*

Intisari— *Stunting pada bayi merupakan masalah kesehatan yang serius, terutama di negara berkembang seperti Indonesia. Penelitian ini bertujuan untuk mengoptimalkan pengelompokan data stunting pada bayi menggunakan algoritma K-Means++, yang dievaluasi dengan Indeks Davies-Bouldin (DBI) untuk menentukan jumlah klaster yang optimal. Data stunting mencakup variabel-variabel seperti usia, jenis kelamin, berat badan, dan tinggi badan. Hasil analisis menunjukkan bahwa jumlah klaster yang optimal adalah 5, dengan nilai DBI sebesar 0,837986204, yang menegaskan kualitas pengelompokan tersebut. Kesimpulan ini menunjukkan bahwa kombinasi metode evaluasi ini menghasilkan pengelompokan yang efektif dan memberikan wawasan yang signifikan dalam mengidentifikasi kelompok bayi dengan berbagai tingkat risiko stunting. Temuan ini dapat menjadi dasar bagi intervensi kesehatan yang lebih terarah dalam mengatasi stunting.*

Kata Kunci: *Klastering, Indeks Davies-Bouldin (DBI), Optimisasi, Stuntin.*

INTRODUCTION

Stunting in infants is a serious health issue that affects both physical and cognitive development in the long term [1], [2]. This condition arises from prolonged insufficient nutritional intake, particularly during the first 1,000 days of life [3], [4]. The impact of stunting ranges from physical risks, such as higher susceptibility to infections, to socio-economic consequences, as stunting can impair cognitive abilities and reduce productivity in adulthood [5], [6]. According to surveys in Indonesia, stunting prevalence remains a significant challenge, with consistently high rates [7]. Therefore, identifying groups of infants at risk of stunting is essential as a preliminary step toward targeted intervention.

To address this issue, data mining methods can be applied to analyze infant health data to uncover patterns or risk groups within the dataset [8], [9], [10]. The K-Means++ algorithm, an improvement over the standard K-Means, is an effective clustering technique for performing unsupervised data grouping [11], [12], [13]. K-Means++ optimizes the

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 11. NO. 1 AUGUST 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i1.7007

initial centroid selection process, which enhances cluster quality and accelerates algorithm convergence compared to standard K-Means [14], [15], [16]. In clustering, evaluating cluster quality is crucial, and methods such as the Davies-Bouldin Index (DBI) are commonly used to determine the optimal number of clusters and assess clustering outcomes [17], [18], [19]. DBI measures cluster quality by calculating the ratio between intra-cluster and inter-cluster distances, where lower DBI values indicate better clustering [20], [21].

Several studies have employed clustering techniques to identify stunting patterns across regions. For instance, Fadilah et al. applied the standard K-Means algorithm to cluster districts and cities in Indonesia based on stunting-related factors such as exclusive breastfeeding, basic immunization, and access to sanitation. Their study successfully identified two main clusters but used a fixed number of clusters (K=2) without evaluating multiple configurations or applying clustering evaluation metrics beyond visualization [22]. Meanwhile, Neneng Sayuti Hanapiah et al. conducted research in Cintarasa Village to optimize stunting management by grouping children's ages and supplementary food distribution using K-Means. The K-Means data mining algorithm successfully grouped data on the ages and supplementary food consumption patterns of children in Cintarasa, forming effective clusters based on nutritional needs [23].

Despite their contributions, both studies have limitations. The use of standard K-Means often suffers from poor centroid initialization, potentially leading to suboptimal and unstable clusters. Furthermore, previous works did not explore a range of cluster configurations to determine the optimal K, nor did they incorporate internal validation metrics such as the Silhouette Score or *Calinski-Harabasz* Index alongside the *Davies-Bouldin Index (DBI).* This hinders comprehensive assessment of clustering quality and robustness. Additionally, few studies perform a correlation analysis of input features to validate their relevance to stunting, which is essential for producing interpretable and actionable clustering results.

To address these gaps, this study proposes the use of the K-Means++ algorithm, which enhances centroid initialization and improves clustering stability. The evaluation is conducted across a range of cluster numbers to identify the optimal configuration based on multiple validation *metrics (DBI, Silhouette Score,* and *Calinski-Harabasz Index).* The novelty of this study lies in its integrative approach, combining improved clustering techniques with a multi-metric evaluation strategy to provide more accurate and actionable insights into regional stunting risk patterns.

MATERIALS AND METHODS

1. *Stunting*

Stunting is a condition of impaired growth and development in children under five caused by chronic or recurrent malnutrition, particularly during the first 1,000 days of life [24]. This results in a height-for-age measurement that is below the standard norm for their age and sex. According to the World Health Organization (WHO) standards, stunting is defined as a height or length that is more than two standard deviations below the median growth for children of the same age. Stunting affects not only physical growth but also has long-term impacts, such as impaired cognitive development, reduced immune function, and increased susceptibility to infectious diseases. In Indonesia, the prevalence of stunting remains significant, reaching 37.2% in 2023. Factors contributing to stunting include malnutrition during pregnancy, lack of access to clean water, poor sanitation, and unbalanced diets during early childhood. Addressing stunting requires a holistic approach, encompassing improved dietary practices, exclusive breastfeeding, and health interventions from pregnancy through the first two years of a child's life.

2. *Data Mining*

Data mining is the process of discovering patterns, relationships, or hidden information within large datasets [25]. By utilizing statistical techniques, mathematical algorithms, and machine learning, data mining identifies patterns that are not immediately apparent, enabling faster and more informed decision-making. In this study, data mining is employed to analyze and cluster stunting data in children to identify patterns related to risk factors such as weight, height, and age. Clustering is a frequently used data mining technique that groups entities based on their similar characteristics. One commonly used algorithm for clustering is K-Means, which partitions data into clusters based on their proximity to cluster centroids. K-Means++, an enhancement of the standard K-Means algorithm, improves the efficiency of initial centroid selection, avoiding slow convergence and enhancing clustering results. In this study, data mining is used to identify groups of children at risk of stunting, aiming to support more targeted health intervention planning.

3. K-Means++ Algorithm

The K-Means++ algorithm is an improvement of the K-Means algorithm used for clustering data. The primary objective of this algorithm is to group data into several clusters based on their proximity to cluster centroids [26]. K-Means++ offers advantages in selecting initial centroids more effectively compared to the standard K-Means, which often selects centroids randomly, potentially resulting in suboptimal outcomes.

The first step in the K-Means++ algorithm is to randomly select one point from the dataset as the first centroid. Then, for each remaining data point, the distance to the nearest existing centroid is calculated. The data point with the greatest distance is chosen as the next centroid. This process is repeated until the desired number of centroids is reached. This more strategic selection of centroids helps mitigate the risk of the algorithm being trapped in local optima, a common issue when centroids are selected randomly in the standard K-Means algorithm.

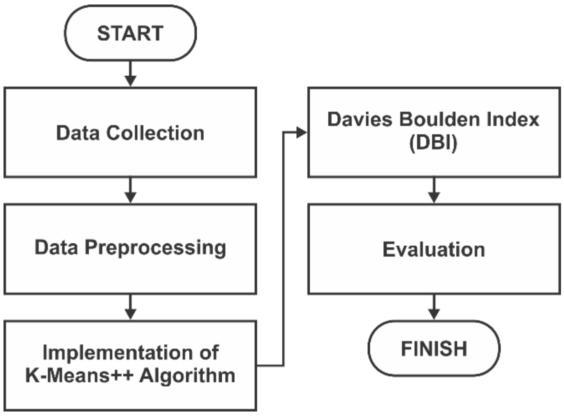
After the initial centroid selection, the algorithm proceeds with steps similar to those in K-Means. Each data point is assigned to the cluster with the closest centroid, and the centroid of each cluster is recalculated based on the mean position of the data points within the cluster. This process is repeated until there is no significant change in the assignment of data points to clusters or the position of centroids, indicating that the algorithm has converged.

The primary advantage of K-Means++ lies in its improved initial centroid selection, leading to more stable and consistent clustering results, as well as fewer iterations required for convergence. This makes K-Means++ more efficient in computational time and more accurate in forming distinct clusters, particularly useful for complex data analysis, such as in this study involving stunting data in children.

4. Clustering Evaluation

Clustering evaluation is a critical step in assessing the quality of data partitioning achieved by algorithms such as K-Means++. The evaluation ensures that data within the same cluster exhibit high similarity (cohesion) while data in different clusters demonstrate clear distinctions (separation) [27]. A common evaluation metric for measuring clustering quality is the Davies-Bouldin Index (DBI). The Davies-Bouldin Index (DBI) is an internal evaluation metric that measures the average similarity ratio between clusters based on the distance between cluster centroids and the average intra-cluster distances. A smaller DBI value indicates better clustering quality, as a low DBI signifies high cohesion within clusters and clear separation between clusters. DBI is calculated as the average of the ratios between the centroid-to-centroid distances of different clusters and the average intra-cluster distances. A non-negative DBI value close to zero suggests that the clusters produced by the algorithm are well-separated. In this study, the Davies-Bouldin Index (DBI) was chosen as the primary evaluation method to assess the effectiveness of K-Means++ clustering on stunting data in children. DBI provides a detailed understanding of clustering quality based on inter-centroid distances and cluster cohesion. By optimizing the DBI value, we can ensure that the clustering results feature well-separated and highly cohesive clusters, which is essential for accurately identifying risk groups for stunting and enabling effective interventions.

Research Stages



Source: (Research Results, 2025)

Figure 1. Research Stages

The image illustrates the research stages involved in implementing the K-Means++ algorithm for data analysis, using the Davies-Bouldin Index (DBI) as an evaluation metric. The process begins with Data Collection, where relevant data for the study, such as age and weight of infants, is gathered. Next is the Data Preprocessing stage, where the collected data is prepared to ensure cleanliness and usability. This includes steps such as normalization, handling missing data, and detecting outliers.

Following this is the Implementation of K-Means++ Algorithm, where the K-Means++ algorithm is applied to partition the data into clusters based on shared characteristics. To identify the optimal number of clusters, the Davies-Bouldin Index (DBI) is used, assessing clustering validity by measuring cluster compactness and separation.

The results of DBI evaluation are then analyzed in the Evaluation phase to determine the

VOL. 11. NO. 1 AUGUST 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i1.7007

effectiveness of the K-Means++ algorithm in generating optimal clusters.

Dataset

This study utilizes a dataset of 500 stunting risk samples, accessed via the Kaggle platform at the following link: https://www.kaggle.com/datasets/harnelia/faktor-stunting.

The dataset includes several critical variables influencing stunting risk, such as gender, age, birth weight, birth length, current weight, and current length. This comprehensive data enables analysis and clustering of infants based on factors related to their nutritional status and growth.

These variables facilitate the implementation of clustering techniques to explore potential patterns in groups of infants at high risk of stunting. The insights gained can support more effective decision-making in addressing stunting. A sample of the raw data is displayed in Table 1.

Table 1. Raw Data Sample Of Stunting Risk

Id	Gender	Age	Birth Weight	Birth Length	Body Weight	Body Length
A1	1	17	3	49	10	72.2
A2	0	11	2.9	49	2.9	65
A3	1	16	2.9	49	8.5	72.2
A4	1	31	2.8	49	6.4	63
A5	1	15	3.1	49	10.5	49
A6	0	11	2.8	49	8.5	65
A7	1	35	2.8	49	10.5	72.2
A8	0	17	2.8	49	8	63
A9	0	10	2.7	49	8.4	73.5
A10	0	16	2.8	49	8.5	65
A11	0	11	2.8	49	10	68.3
A12	1	13	2.9	50	10	69
A13	1	44	3	49	7.1	72.2
A14	1	18	2.8	50	7.2	65
A15	1	13	2.8	48	7.7	65
A16	0	13	2.8	49	10.5	72.2
A17	1	7	2.3	50	6.4	68.3
A18	1	16	2.7	50	2.9	69
...	...	...	...	...	...	...
A500	1	10	3.1	50	9	73.5

Source: (Research Results, 2025)

Data Preprocessing

Before analysis is conducted, the data will go through a preprocessing process that includes several important stages. First, data cleaning is carried out to handle missing values using imputation methods or by removing incomplete data. Second, data normalization is performed to adjust the data scale to the same standard to ensure more accurate clustering results. Finally, outlier identification and handling are conducted because outliers can significantly affect clustering results, making it necessary to address them so that the analysis is more representative and valid. The processed data after preprocessing can be seen in Table 2.

Table 2. Stunting Risk Data Sample After Preprocessing

Id	Gender	Age	Birth Weight	Birth Length	Body Weight	Body Length
A1	1	0.261904762	0.909090909	0.5	0.934210526	0.530892449
A2	0	0.119047619	0.818181818	0.5	0	0.366132723
A3	1	0.238095238	0.818181818	0.5	0.736842105	0.530892449
A4	1	0.595238095	0.727272727	0.5	0.460526316	0.320366133
A5	1	0.214285714	1	0.5	1	0
A6	0	0.119047619	0.727272727	0.5	0.736842105	0.366132723
A7	1	0.69047619	0.727272727	0.5	1	0.530892449
A8	0	0.261904762	0.727272727	0.5	0.671052632	0.320366133
A9	0	0.095238095	0.636363636	0.5	0.723684211	0.560640732
A10	0	0.238095238	0.727272727	0.5	0.736842105	0.366132723
A11	0	0.119047619	0.727272727	0.5	0.934210526	0.441647597
A12	1	0.166666667	0.818181818	1	0.934210526	0.457665904
A13	1	0.904761905	0.909090909	0.5	0.552631579	0.530892449
A14	1	0.285714286	0.727272727	1	0.565789474	0.366132723
A15	1	0.166666667	0.727272727	0	0.631578947	0.366132723
A16	0	0.166666667	0.727272727	0.5	1	0.530892449
A17	1	0.023809524	0.272727273	1	0.460526316	0.441647597
A18	1	0.238095238	0.636363636	1	0	0.457665904
...	...	...	...	...	...	...
A500	1	0.095238095	1	1	0.802631579	0.560640732

Source: (Research Results, 2025)

The dataset was first examined for missing values and outliers. Missing values were addressed using mean imputation, while outliers were detected through boxplot analysis. Data normalization was conducted using Min-Max scaling to ensure that all features were on the same scale and to avoid bias in distance-based clustering. An exploratory data analysis (EDA) was also performed to understand the data distribution and relationships among variables, although detailed correlation analysis is recommended for future work.

Implementation of K-Means++ Algorithm

In the implementation of the K-Means++ algorithm, the clustering process begins by calculating the Euclidean distance between each data point and the centroids that have been formed, as well as the initial centroid selection aimed at optimizing cluster formation. The Euclidean distance is used to measure the proximity between a data point and the centroid of each cluster. This distance determines which cluster is closest to a particular data point, allowing the data to be grouped into the most appropriate cluster. The formula for Euclidean Distance between two points $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ is as follows:

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

With x_i and y_i being the coordinates of the data point and the centroid in the iii-th dimension, and nnn representing the number of dimensions of the data, this formula calculates the linear distance between the data point and the centroid. The smaller the distance, the closer the data point is to the centroid, and the higher the likelihood of the data point being assigned to that cluster.

The next step involves the selection of the initial centroids in K-Means++. K-Means++ improves clustering accuracy by strategically choosing the initial centroids, unlike the standard K-Means, which selects centroids randomly. The steps for selecting the initial centroids in K-Means++ are as follows:

1) Select one data point randomly from the dataset as the first centroid.
2) For each unselected data point x, calculate the distance between the point and the already selected centroid. Determine the smallest distance between the data point x and the selected centroids, and record this distance as $D(x)^2$.
3) Determine the probability of selecting each data point as the next centroid based on the $D(x)^2$ values calculated. Points that are farther from the existing centroids will have a higher probability of being selected as the next centroid. The formula for determining the probability of selecting a data point xxx as the next centroid is:

$$P(x) = \frac{D(x)^2}{\sum_{x \in X} D(x)^2} \quad (2)$$

4) Select the next data point as a centroid based on the calculated probabilities. Repeat this process until the desired number of centroids is reached.

where c_j is the new centroid of cluster j, and n_j is the number of data points in the cluster. This process is repeated by calculating the Euclidean distance, grouping the data, and updating the centroids until convergence is achieved, i.e., when there are no significant changes in the centroid positions or the iteration reaches the maximum limit. The results of the initial centroid points using the K-Means++ algorithm can be seen in Table 3.

Table 3. Initial Centroid Results Of K-Means++

Centroid	Id	Highest Probability
C1	A4	(Randomly Selected)
C2	A142	0.005151142
C3	A192	0.012478458
C4	A428	0.005632282
C5	A378	0.004450477
C6	A331	0.005606862
C7	A489	0.005170229
C8	A125	0.005231879
C9	A164	0.005242804
C10	A143	0.005056038

Source: (Research Results, 2025)

The initial centroid point is selected based on calculations from the K-Means++ algorithm.

Evaluation Using Davies-Bouldin Indeks (DBI)

The Davies-Bouldin Index (DBI) is an internal evaluation metric used to assess the quality of clustering results generated by clustering algorithms, such as K-Means++. DBI measures how well the clusters are formed by evaluating the cohesion (compactness) within each cluster and the separation between clusters. In DBI, a lower value indicates better cluster quality, where the clusters are more compact internally and more distinct from each other.

DBI is calculated as the average of the ratio between the centroid distances of clusters and the average distances between points within a cluster. The formula for DBI for k clusters is as follows:

$$DBI = \frac{1}{k} \sum_{i=1}^{k} max_{i \neq j} (R_{i,j}) \quad (3)$$

Based on equation (3), K represents the total number of clusters. A lower DBI value indicates better clustering performance. This study draws conclusions based on the DBI value, where a low DBI value suggests that the resulting clusters are of high quality, and the method effectively evaluates the optimal number of clusters.

VOL. 11. NO. 1 AUGUST 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i1.7007

RESULTS AND DISCUSSION

Clustering with K-Means++

After applying the K-Means++ algorithm to determine the initial cluster centroids more accurately, the clustering process continues using the K-Means method with the number of clusters varying from K=2 to K=10. Validation of the clustering results is conducted by calculating the Davies-Bouldin Index (DBI) for each cluster configuration. Information regarding the initial cluster centroids generated by the K-Means++ algorithm for the range of K=2 to K=10 is presented in Table 4.

Table 4. Cluster Validation with DBI

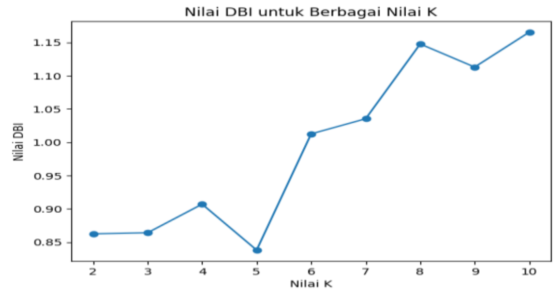
K	ID	Gender	Age	Birth Weight	Birth Length	Body Weight	Body Length
C1	A8	0	0.261904762	0.727272727	0.5	0.671052632	0.320366133
C2	A23	1	0.166666667	0.818181818	0.5	0.802631579	0.457665904
C3	A53	0	0.142857143	0	0.5	0.565789474	0.530892449
C4	A66	1	0.976190476	0.909090909	0.5	0.802631579	0.320366133
C5	A144	1	0.19047619	0.818181818	0.5	0.802631579	0.320366133
C6	A228	1	0.047619048	0.272727273	1	0.460526316	0.441647597
C7	A242	0	0.166666667	0.727272727	0.5	0.631578947	0.366132723
C8	A273	1	0.071428571	0.727272727	1	0	0.560640732
C9	A302	0	0.119047619	0.727272727	0.5	0.631578947	0
C10	A308	0	0.238095238	0.727272727	0.5	0.723684211	0

Source: (Research Results, 2025)

Table 4 shows that the initial centroids have been determined using the K-Means++ algorithm. Subsequently, clustering is performed using the K-Means algorithm for each K value, ranging from K=2 to K=10, utilizing the initial centroids generated in Table 4. The clustering results have been obtained using the K-Means algorithm with the number of clusters varying from K=2 to K=10.

Evaluation with Davies-Bouldin Indeks (DBI)

The clustering results are evaluated using the Davies-Bouldin Index (DBI). The DBI value serves as an indicator to assess the quality of clustering. Clusters are considered optimal if they have the smallest DBI value. A visualization of cluster validation using DBI can be seen in Figure 2.



Source: (Research Results, 2025)
Figure 2. DBI Chart to Determine Optimal Number of Clusters

From Figure 2, we get the DBI value results from all clusters which can be seen in Table 5:

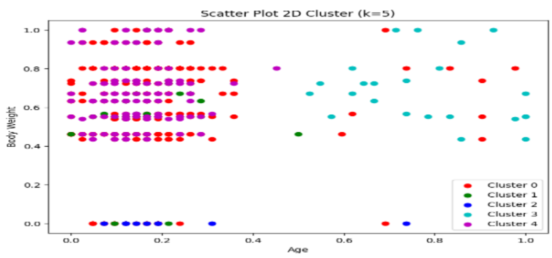
Table 5. Cluster Validation with DBI

K	DBI Value
K=2	0.86233495
K=3	0.86400863
K=4	0.906926438
K=5	0.837986204
K=6	1.012528682
K=7	1.035213981
K=8	1.147575347
K=9	1.112910126
K=10	1.165254112

Source: (Research Results, 2025)

Table 5 shows that the DBI value from the K-Means++ clustering results for K=5 achieves an average of 0.837986204, which is lower than the DBI values for other cluster counts. This lower DBI value indicates that clustering with K-Means++ produces better cluster quality.

Next, Figure 3 presents a 2D scatter plot of the clustering results using the K-Means++ algorithm with the optimal number of clusters, K=5.



Source: (Research Results, 2025)
Figure 3. Clustering Results of K-Means++ K=5 Algorithm Based on 2 Variables

The image above is a two-dimensional scatter plot illustrating the division of data into five clusters (K=5), which was selected as the optimal number of clusters based on the Davies-Bouldin Index (DBI) evaluation. The horizontal axis represents the Age feature, while the vertical axis represents the Body Weight feature. Each point in the scatter plot represents an individual data entry, and the color of the point indicates the cluster to which the data belongs.

In this visualization, each cluster is represented by a different color for easier visual identification: red (Cluster 0), green (Cluster 1), blue (Cluster 2), cyan (Cluster 3), and purple (Cluster 4). The clustering shows distinct distribution patterns based on age and body weight. For instance, Cluster 3 (cyan) appears to include infants with higher age and body weight compared to other clusters. Conversely, Cluster 0 (red) tends to cover data with a wider age range but lower body weight.

This visualization demonstrates that the K-Means++ algorithm successfully identified optimal initial centroids, resulting in a more structured cluster division. These clusters can be used to analyze the characteristics of infant groups based on age and body weight, aiding in the identification of stunting patterns and determining the necessary intervention priorities. The clustering provides critical insights for researchers and practitioners in understanding the distribution of stunting data among infants within the dataset.

Compared to the results of Fadilah et al., who employed standard K-Means to cluster districts/cities in Indonesia based on stunting risk factors using a fixed K=2 determined by the Elbow method, this study extends the approach by evaluating a broader range of cluster numbers and applying K-Means++ for improved centroid initialization. The DBI value achieved in this study (0.837) with K=5 demonstrates better cohesion and separation, addressing the previous study's limitation of relying solely on SSE without incorporating internal validity indices like DBI.

Furthermore, unlike the approach by Hanapiah et al., which applied K-Means to group supplementary feeding types based on children's age in a specific rural area (Desa Cintarasa) without cluster validity analysis, this study generalizes clustering across broader regional features and supports the results with robust evaluation. By optimizing both the number of clusters and centroid initialization, K-Means++ in this study provides more granular and reliable groupings for stunting risk assessment.

This study has several limitations. First, it only uses internal validation (DBI) without combining it with other indices such as Silhouette Score or Calinski-Harabasz. Second, the dataset is limited to a specific region and may not represent the full diversity of stunting risk factors in different demographics. Future research should compare multiple clustering algorithms (e.g., Agglomerative, DBSCAN) and employ multi-metric evaluations. It is also recommended to perform feature selection or correlation analysis to ensure that only the most relevant features are included in the clustering process.

CONCLUSION

This study successfully identified the optimal number of clusters within stunting infant data using the K-Means++ algorithm, validated through the Davies-Bouldin Index (DBI). The analysis results indicated that the optimal number of clusters is five, with a DBI score of 0.837986204, signifying adequate separation between clusters. With this approach, this study is expected to contribute to the development of more optimal data analysis methods in healthcare and provide valuable insights for public health policy-making.

REFERENCE

[1] N. Puspita Dewi, M. Mariyani, S. Alaydrus, V. Anggi, M. Magfirah, and I. Kurnia Utami, “Penyuluhan Stunting : Upaya Meminimalisir Tingginya Angka Stunting di Desa Sawidago,” *J. DiMas*, vol. 5, no. 2, pp. 64–67, 2023, doi: 10.53359/dimas.v5i2.69.

[2] R. Khoiriyah, T. Mahmudiono, and S. R. Nadhiroh, “Indirect Determinant Effects of Stunting with Toddler Stunting Incident in Papua New Guinea in 2018,” *Amerta Nutr.*, vol. 8, no. 3, pp. 433–440, 2024, doi: 10.20473/amnt.v8i3.2024.433-440.

[3] A. Syahrin, P. Candriasih, and D. E. Kusumawati, “Gambaran Pengetahuan Stunting pada Ibu Balita Stunting,” *Svasta Harena J. Ilm. Gizi*, vol. 3, no. 2, pp. 46–51, 2023, doi: https://doi.org/10.33860/shjig.v3i2.3401.

[4] E. Palupi, I. Y. Sari, I. Prawesti, and S. Febrianti, “Menuju Generasi Emas Tanpa Stunting Melalui Edukasi Pencegahan Stunting Pada Ibu Hamil Di Mlatibaru Semarang,” *JAMAS J. Abdi Masy.*, 2024, doi: https://doi.org/10.62085/jms.v2i1.70.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 11. NO. 1 AUGUST 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i1.7007

[5] E. Mutiara, N. I. Hasibuan, S. Yunita, K. Saleh, and N. Akmal, “Community Empowerment Through Healthy Kitchens Overcomes Stunting Towards Stunting-Free Villages,” *Din. J. Pengabdi. Kpd. Masy.*, vol. 8, no. 5, pp. 1426–1434, 2024, doi: https://doi.org/10.31849/dinamisia.v8i5.23486.

[6] Hesti Nur Sahadatilah, Hasan Mukmin, Rini Setiawati, and MA Achlami, “Program Dapur Sehat Atasi Stunting (DASHAT) dalam Mempercepat Penurunan Stunting,” *Media Publ. Promosi Kesehat. Indones.*, vol. 6, no. 12, pp. 2599–2604, 2023, doi: 10.56338/mppki.v6i12.4382.

[7] Tugiono, A. Alhafiz, and Hafizah, “Sistem Cerdas Mendiagnosa Stunting pada Anak Menggunakan Mesin Inferensi,” *J. Inf. dan Teknol.*, vol. 4, no. 4, pp. 7–10, 2022, doi: 10.37034/jidt.v4i4.237.

[8] A. U. Rehman, R. M. Saleem, Z. Shafi, M. Imran, M. Pradhan, and H. M. Alzoubi, “Analysis of Income on the Basis of Occupation using Data Mining,” *2022 Int. Conf. Bus. Anal. Technol. Secur. ICBATS 2022*, no. February, 2022, doi: 10.1109/ICBATS54253.2022.9759040.

[9] P. Edastama, A. Dudhat, and G. Maulani, “Use of Data Warehouse and Data Mining for Academic Data : A Case Study at a National University,” *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 2, pp. 206–215, 2021, doi: 10.34306/ijcitsm.v1i2.55.

[10] J. Yang *et al.*, “Brief introduction of medical database and data mining technology in big data era,” *J. Evid. Based. Med.*, vol. 13, no. 1, pp. 57–69, 2020, doi: 10.1111/jebm.12373.

[11] L. Wei, H. Ma, Y. Yin, and C. Geng, “Kmeans-CM Algorithm with Spectral Angle Mapper for Hyperspectral Image Classification,” *IEEE Access*, vol. 11, no. March, pp. 26566–26576, 2023, doi: 10.1109/ACCESS.2023.3257859.

[12] Z. Ji, Y. Wang, J. Su, and J. Wu, “Research on insurance policy formulation based on ARIMA-KMEANS algorithm,” *Highlights Business, Econ. Manag.*, vol. 36, pp. 52–58, 2024, doi: 10.54097/491e2j22.

[13] N. Z. Salih and W. Khalaf, “Prediction of student’s performance through educational data mining techniques,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 22, no. 3, pp. 1708–1715, 2021, doi: 10.11591/ijeecs.v22.i3.pp1708-1715.

[14] S. Siet, S. Peng, S. Ilkhomjon, M. Kang, and D. S. Park, “Enhancing Sequence Movie Recommendation System Using Deep Learning and KMeans,” *Appl. Sci.*, vol. 14, no. 6, 2024, doi: 10.3390/app14062505.

[15] M. Guntara and N. Lutfi, “Optimasi Cacah Klaster pada Klasterisasi dengan Algoritma KMeans Menggunakan Silhouette Coeficient dan Elbow Method,” *JuTI “Jurnal Teknol. Informasi,”* vol. 2, no. 1, p. 43, 2023, doi: 10.26798/juti.v2i1.944.

[16] W. T. Wu *et al.*, “Data mining in clinical big data: the frequently used databases, steps, and methodological models,” *Mil. Med. Res.*, vol. 8, no. 1, pp. 1–12, 2021, doi: 10.1186/s40779-021-00338-z.

[17] I. Faturohim and A. Baita, “Penerapan Algoritma Kmeans Clustering Untuk Mengelompokkan Kerawanan Rob Di Daerah Pekalongan,” *JuTI “Jurnal Teknol. Informasi,”* vol. 1, no. 1, p. 12, 2022, doi: 10.26798/juti.v1i1.624.

[18] L. Dong, “Research on online interactive teaching mode of college vocal music based on adaptive Kmeans algorithm,” *Appl. Math. Nonlinear Sci.*, vol. 8, no. 2, pp. 3383–3392, 2023, doi: https://doi.org/10.2478/amns-2024-2873.

[19] M. Gao *et al.*, “Identification Method of Electrical Load for Electrical Appliances Based on K-Means ++ and GCN,” *IEEE Access*, vol. 9, pp. 27026–27037, 2021, doi: 10.1109/ACCESS.2021.3057722.

[20] M. Mughnyanti, S. Efendi, and M. Zarlis, “Analysis of determining centroid clustering x-means algorithm with davies-bouldin index evaluation,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 725, no. 1, 2020, doi: 10.1088/1757-899X/725/1/012128.

[21] F. Nurulhikmah and D. N. E. Abdi, “Classification of Foods Based on Nutritional Content Using K- Means and DBSCAN Clustering Methods,” *TEKNIKA*, vol. 13, no. November, pp. 481–486, 2024, doi: 10.34148/teknika.v13i3.1067.

[22] A. Fadilah, M. N. Pangestu, S. Lumbanbatu, and S. Defiyanti, “Pengelompokan Kabupaten/Kota Di Indonesia Berdasarkan Faktor Penyebab Stunting Pada Balita Menggunakan Algoritma K-Means,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 6, no. 2, p. 223, 2022, doi: 10.26798/jiko.v6i2.581.

[23] N. Sayuti Hanapiah, N. Suarna, and W. Prihartono, “Meningkatkan Penanganan Stunting Pada Anak Melalui Klasifikasi Pemberian Makanan Tambahan Berdasarkan Usia Dengan Metode K-Means Di Desa Cintarasa,” *JATI (Jurnal Mhs. Tek.*

Inform., vol. 7, no. 6, pp. 3484–3493, 2024, doi: 10.36040/jati.v7i6.8213.

[24] R. R. Akbar, W. Kartika, and M. Khairunnisa, "The Effect of Stunting on Child Growth and Development," *Sci. J.*, vol. 2, no. 4, pp. 153–160, 2023, doi: 10.56260/sciena.v2i4.118.

[25] H. Thakkar, V. Shah, H. Yagnik, and M. Shah, "Comparative anatomization of data mining and fuzzy logic techniques used in diabetes prognosis," *Clin. eHealth*, vol. 4, no. 2021, pp. 12–23, 2021, doi: 10.1016/j.ceh.2020.11.001.

[26] S. Daoudi, C. M. A. Zouaoui, M. C. El-Mezouar, and N. Taleb, "Parallelization of the K-Means++ clustering algorithm," *Ing. des Syst. d'Information*, vol. 26, no. 1, pp. 59–66, 2021, doi: 10.18280/isi.260106.

[27] E. Hassan *et al.*, "A Hybrid K-Means++ and Particle Swarm Optimization Approach for Enhanced Document Clustering," *IEEE Access*, vol. 13, no. March, pp. 48818–48840, 2025, doi: 10.1109/ACCESS.2025.3535226.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri