Vol. 7, No. 4, December 2025

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

K-Means Clustering with Elbow Method for Stunting Risk Detection in Toddlers Using Anthropometric and Nutritional Data

Irma Darmayanti [1], Dhanar Intan Surya Saputra[2], Anugerah Bagus Wijaya[3], Andik Wijanarko[4], Dewi Fortuna[5], Aldrian Firmansyah Putranto[6]

[1,4]Department of Information Technology, Amikom Purwokerto University, Banyumas, Indonesia

[2]Departement of Information System, Amikom Purwokerto University, Banyumas, Indonesia

[3,6]Department of Informatics, Amikom Purwokerto University, Banyumas, Indonesia

[5]Departement of Information System, Telkom University, Bandung, Indonesia

Email: irmada@amikompurwokerto.ac.id[1], dhanarsaputra@amikompurwokerto.ac.id[2], anugerah@amikompurwokerto.ac.id[3], andikwijanarko@amikompurwokerto.ac.id[4], dewifrtuna53@gmail.com[5], ryury345@gmail.com[6]

Received: Oct 12, 2025
Revised: Nov 10, 2025
Accepted: Nov 27, 2025
Published: Dec 12, 2025

Corresponding Author:
Author Name*:
Irma Darmayanti
Email*:
irmada@amikompurwokerto.ac.id

DOI:
10.63158/journalisi.v7i4.1337

Abstract. Stunting remains a critical public health challenge in Indonesia, primarily due to inadequate nutrition and recurrent infections in early childhood. This study aimed to identify patterns of stunting risk by integrating anthropometric and dietary data, specifically sugar consumption, using an unsupervised machine learning approach. A total of 20 toddlers aged 12-59 months from Purwokerto Selatan participated. Anthropometric data (age, weight, height) and dietary intake (sugar consumption, snack frequency) were collected via a caregiver questionnaire. K-Means clustering was applied, with the optimal number of clusters determined using the Elbow Method (K=2). Two clusters were identified: Cluster 0, with a lower risk of stunting, and Cluster 1, with a higher proportion of toddlers at risk. Cross-tabulation with stunting status validated this, showing that Cluster 1 contained more children with "Potential" stunting. Internal validation using the Silhouette score (0.252) and PCA visualization confirmed the clustering's robustness. This study demonstrates the potential of combining anthropometric and dietary data for stunting risk profiling, suggesting a complementary approach for growth monitoring programs and targeted interventions.

Keywords: Stunting, K-Means Clustering, Toddler, Nutrition, Anthropometric Data

1. INTRODUCTION

Stunting, defined as impaired linear growth resulting from prolonged inadequate nutrient intake and recurrent infections, remains a major public health problem in Indonesia [1]. The prevalence of stunting among children under five reached 26.2% in 2022 and 25.8% in 2023 [2], indicating that undernutrition in early childhood is still a critical national concern [3]. Stunting is typically assessed using the length/height-for-age index (LAZ/HAZ), where children with a Z-score between −2 SD and −3 SD are categorized as stunted and those below −3 SD as severely stunted [4]. This condition, which may begin as early as the prenatal period and become clinically evident around the age of two years [4], is associated with long-term consequences, including impaired cognitive development, lower educational attainment, and reduced economic productivity in adulthood [5].

One of the key determinants of stunting is an unbalanced diet that fails to meet the child's nutritional needs over time [5]. In Indonesia, it is common for parents to provide snack foods such as biscuits, candy, and sugar-sweetened packaged drinks as part of toddlers' daily dietary pattern [6],[7]. Many of these commercially available products contain high levels of added sugar, which can disturb the overall balance of nutrient intake. A study by Irma Darmayanti et al. reported that packaged beverages frequently consumed by children, including flavored milk, fall into clusters with high sugar content [8]. Excessive sugar intake may displace more nutrient-dense foods and thereby contribute to deficiencies in essential nutrients required for optimal growth and development [9].

Alongside conventional epidemiological approaches, technology-based methods are increasingly used to analyze complex nutrition and health data. Machine learning has been applied in various health-related studies to uncover hidden patterns in large datasets and to support risk stratification, prediction, and decision-making [9],[10]. In the context of child nutrition, clustering methods can help characterize groups of children who share similar dietary and anthropometric profiles, which may not be easily identified using traditional statistical techniques. Unsupervised learning approaches, particularly clustering algorithms, are well suited to explore patterns of excessive intake of specific nutrients and their potential association with stunting risk in early childhood [11]-[14].

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Several studies in Indonesia have used clustering approaches, especially K-Means, to categorize stunting risk based primarily on anthropometric indicators such as weight, height, and age [15]-[17]. These studies have contributed to identifying groups of children at higher risk of stunting but generally relied on physical growth measurements alone as the basis for classification. As a result, the role of dietary patterns—especially sugar consumption and overall nutritional intake—has not been fully integrated into clustering models for stunting risk. Unlike previous studies that cluster stunting status solely based on anthropometric measures, this study integrates nutritional intake variables to provide a more holistic stunting risk detection model. By combining anthropometric indicators with information on food and sugar consumption patterns, the proposed unsupervised learning approach is expected to capture more nuanced profiles of stunting risk among Indonesian toddlers.

Based on this rationale, the present study aims: (1) to identify patterns of sugar consumption and overall nutritional intake among toddlers using an unsupervised clustering approach; (2) to group toddlers into distinct clusters of stunting risk by jointly considering anthropometric and dietary variables; and (3) to analyze how these clusters relate to stunting-related indicators as a basis for developing more targeted nutrition interventions.

2. METHODS

The research methodology is outlined in Figure 1, which provides a visual representation of the steps followed in the study, from data collection to Cluster Visualization.

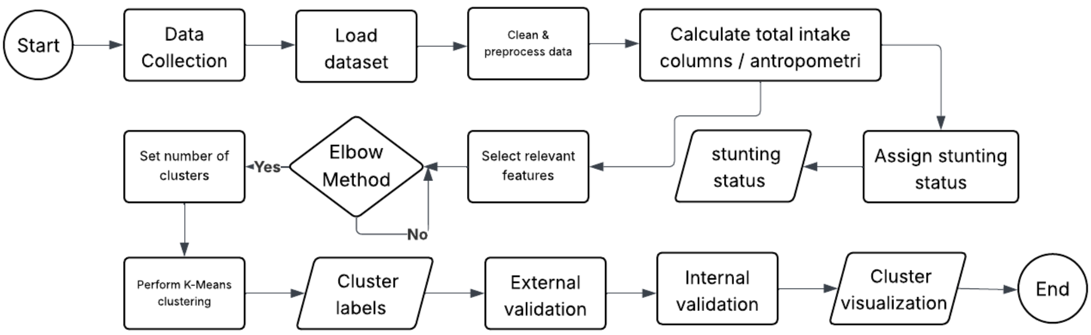

Figure 1. Method Flowchart

The explanation of figure 1 is as follows.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

2.1. Collecting Data

This study employed a quantitative, cross-sectional design conducted among caregivers of children aged 12–59 months residing in the Purwokerto Selatan area, Indonesia. A total of 20 toddlers participated, consisting of 13 boys and 7 girls. Inclusion criteria included: (1) age 12–59 months, (2) residing in Purwokerto Selatan, and (3) caregiver willingness and ability to complete the questionnaire. Toddlers with severe congenital abnormalities affecting growth were excluded.

Data were collected using a self-administered questionnaire distributed widely to caregivers. The first section captured sociodemographic and anthropometric information (age, sex, weight, and height), which were based on the most recent POSYANDU measurements recorded in each child's health card. The second section assessed dietary intake using a 7-day frequency recall focusing on sugar-containing beverages and snacks (e.g., sweetened UHT milk, packaged sweet tea, powdered drinks, sugary snacks). Nutrient intake estimates were later computed using national/international food composition tables and aggregated into weekly total intake variables. Informed consent was obtained at the end of the questionnaire, and only responses with explicit consent were included.

2.2. Load Dataset

All eligible questionnaires were checked for completeness, coded, and imported into Python as a structured dataset. Variable names were standardized for consistency before analysis.

2.3. Clean and Preprocess Data

The dataset was examined for missing values, entry errors, and implausible values. Records missing essential anthropometric data were excluded[18]. For dietary variables with minimal missingness, median imputation was applied. All continuous variables selected for clustering were standardized using Z-score normalization to prevent scale dominance during distance calculations in K-Means.

2.4. Calculate Total Intake Columns / Anthropometric Variables

Seven-day dietary frequency responses were converted into estimated weekly nutrient intakes (sugar, protein, calcium, iron, vitamins, energy) based on food composition tables.

These totals, together with standardized anthropometric indicators (age, weight, height), constituted the primary features considered for clustering.

2.5. Assign Stunting Status

Stunting status was derived using the WHO Child Growth Standards. Each child's Height-for-Age Z-score (HAZ) was computed. Toddlers with HAZ < −2 SD were classified as stunted (or "Potential"), whereas those with HAZ ≥ −2 SD were classified as not stunted ("No Potential"). This label was not used in the clustering algorithm but rather for external validation.

2.6. Select Relevant Features

The final feature set included age, weight, height, total weekly sugar intake, and other nutritional/behavioral variables. All selected features were standardized. The stunting status variable was excluded to maintain the unsupervised nature of the clustering.

2.7. Elbow Method

To determine the optimal number of clusters K, the Elbow Method was applied to the standardized feature set. K-Means models were fitted for a series of candidate cluster numbers, typically $K = 2$ to $K = 10$. For each value of K, the within-cluster sum of squares (WCSS) was computed, representing the total squared distance between each data point and the centroid of its assigned cluster. The WCSS values were then plotted against the corresponding Kvalues to produce an Elbow plot[19]. The "elbow point"—where the rate of decrease in WCSS began to level off—was visually identified and taken as the most reasonable trade-off between model complexity and cluster compactness.

2.8. Set Number of Clusters

Based on the Elbow Method, the number of clusters for the final analysis was set to $K = 2$. This value was used as the primary parameter in the subsequent K-Means clustering procedure. Fixing K at this stage ensured that all further clustering steps and interpretations were consistently based on the same cluster structure.

2.9. Perform K-Means Clustering

K-Means clustering was then performed on the standardized feature set using $K = 2$ and the K-Means++ initialization strategy to enhance convergence stability. The algorithm iteratively assigned each toddler to the nearest cluster centroid based on Euclidean distance, updated the centroid positions as the mean of the points in each cluster, and repeated these steps until convergence criteria were met [20] (no further changes in assignments or minimal change in WCSS, with a maximum of [e.g., 300] iterations). The final K-Means model partitioned the toddlers to representing distinct profiles of growth and nutritional intake.

2.10. Cluster Labels

The output of the K-Means procedure was a cluster label for each toddler (e.g., Cluster 0 and Cluster 1). These labels were added to the dataset as a new categorical variable. Descriptive statistics (means and standard deviations) for anthropometric and intake variables were then calculated within each cluster to provide an initial understanding of the characteristics of the two groups. These cluster labels also served as the basis for subsequent internal and external validation analyses.

2.11. External Validation

Cluster labels were cross-tabulated with stunting status to evaluate alignment between unsupervised clusters and established growth-risk categories. A contingency table was constructed to identify patterns of stunting distribution across clusters.

2.12. Internal Validation

Internal validation of clustering quality was performed using the Silhouette score. For each toddler, the Silhouette coefficient was calculated based on the average distance to all other points within the same cluster and the average distance to points in the nearest neighboring cluster. The Silhouette coefficient ranges from −1 to +1, with higher values indicating better separation between clusters and more coherent cluster membership [21],[22]. The overall quality of the clustering solution was summarized by the mean Silhouette score across all toddlers. This measure provided an internal benchmark of how well the two-cluster solution captured distinct patterns in the data.

2.13. Cluster Visualization

To visualize the cluster structure, Principal Component Analysis (PCA) was applied to reduce dimensionality [23]. The first two principal components (PC1 and PC2) were plotted, with points colored by cluster membership. Variance explained by each component was reported to assess the contribution of each axis in separating clusters.

3. RESULTS AND DISCUSSION

3.1. Determination of the optimal number of clusters

The Elbow Method was applied to the standardized feature set (age, weight, height, total sugar intake frequency, age at complementary feeding initiation, and screen time). For each value of K from 1 to 10, the within-cluster sum of squares (WCSS) was computed and plotted (Figure 2).

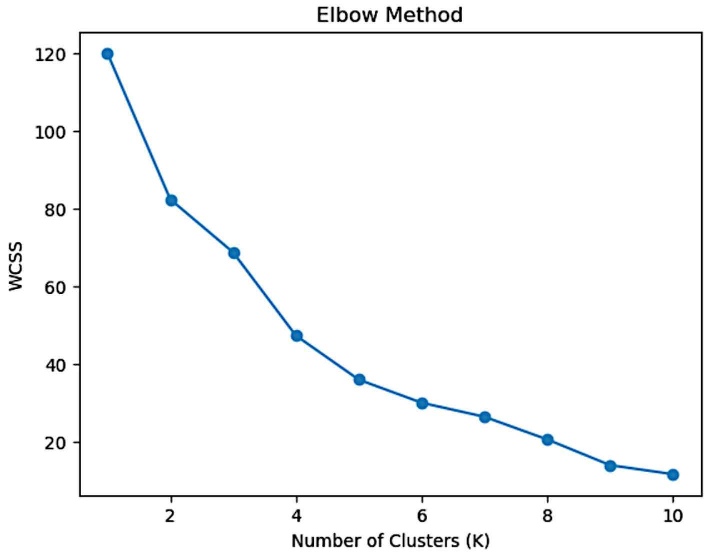

Figure 2. Visualization Result of WCSS

The WCSS curve showed a steep decrease between $K = 1$ and $K = 2$, followed by a much more gradual decline for $K > 2$. This pattern indicates an "elbow" at $K = 2$, suggesting that a two-cluster solution provides an adequate trade-off between model complexity and cluster compactness. Consequently, $K = 2$ was selected as the optimal number of clusters for the subsequent K-Means analysis.

3.2. K-Means clustering results

Using $K = 2$, the K-Means algorithm was fitted to the standardized data. Each toddler was assigned to one of two clusters representing distinct profiles of growth and nutritional intake. Descriptively, one cluster tended to include toddlers with relatively better anthropometric indicators and lower to moderate sugar intake, whereas the other cluster included children with less favorable anthropometric profiles and higher sugar consumption frequencies. The final cluster labels were added to the dataset and used for further internal and external validation. Overall, the two-cluster structure was consistent with the a priori expectation of distinguishing a lower-risk(0) and a higher-risk(1) group in terms of stunting. The results are shown in Table 1.

Table 1. K-Means clustering

Cluster	toddlers
0	6
1	14

3.3. External Validation

To assess whether the clusters were related to established indicators of stunting risk, the K-Means cluster labels were cross-tabulated with the independently derived stunting status ("No Potential" vs "Potential"). The results are shown in Table 2.

Table 2. Crosstab of Cluster × Stunting Status

Cluster	No Potential	Potential
0	5	1
1	7	7

Cluster 0 consisted mainly of toddlers categorized as "No Potential" (5 out of 6; 83.3%), with only 1 child (16.7%) labelled as "Potential" stunting. In contrast, Cluster 1 showed a much higher proportion of children with "Potential" stunting (7 out of 14; 50.0%). Overall, 12 of the 20 toddlers (60.0%) were classified as "No Potential" and 8 (40.0%) as "Potential". These findings indicate that Cluster 1 is enriched with toddlers at higher risk of stunting compared with Cluster 0, supporting the relevance of the unsupervised clustering solution for distinguishing stunting-related risk profiles based on combined anthropometric and dietary information.

3.4. Internal Validation Using Silhouette Score

Internal validation using the Silhouette coefficient yielded an overall Silhouette score of 0.252 for the two-cluster solution. This value indicates a moderate separation between clusters: most toddlers are closer to members of their own cluster than to the neighboring cluster, although some degree of overlap remains. Considering the small sample size (N = 20) and the multidimensional nature of the anthropometric and dietary variables, this level of separation is acceptable and suggests that the identified clusters reflect a meaningful underlying structure in the data rather than random noise.

3.5. PCA-based visualization of cluster structure

Principal Component Analysis (PCA) was used to project the standardized features onto two principal components for visualization. The first principal component (PC1) explained 54.5% of the total variance, while the second component (PC2) explained 21.0%, so that together PC1 and PC2 accounted for 75.5% of the variance in the data.

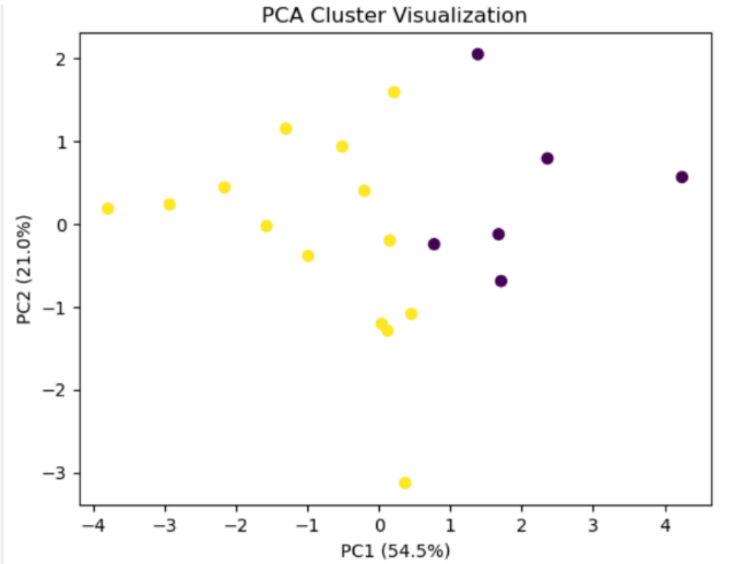

Figure 3. PCA Visualization

The scatter plot of toddlers in the PC1–PC2 space, colored by cluster membership (Figure 3), shows that the two clusters are primarily separated along PC1. Cluster 1 tends to occupy the region with higher (positive) PC1 values, whereas Cluster 0 is more dispersed and occupies a broader range, including negative PC1 values. This pattern suggests that PC1—which represents a composite gradient of anthropometric measures and sugar-related intake—plays a major role in differentiating the two groups. The PCA visualization corroborates the numerical validation results: although there is some overlap between clusters, the two-cluster solution captures distinct profiles of toddlers that align with differences in stunting risk.

3.6. Discussion

This study utilized an unsupervised machine learning approach to explore stunting risk patterns in toddlers by integrating both anthropometric data and dietary intake, with a particular focus on sugar consumption. The K-Means clustering algorithm, with K=2, successfully identified two distinct groups based on these variables, providing valuable insights into stunting risk. The optimal number of clusters was determined using the Elbow Method, which suggested K=2 as the most appropriate number of clusters. The Silhouette score of 0.252 indicated a moderate level of separation between the clusters, suggesting that the identified groups were distinct but not completely isolated. This moderate separation can be attributed to the multidimensional nature of the data, where variations in both dietary and growth factors contribute to the clustering results.

The cross-tabulation with stunting status showed a significant relationship between the clusters and the risk of stunting. Cluster 0, which consisted primarily of toddlers with lower sugar consumption and better anthropometric measurements, contained 83.3% of the children classified as having "No Potential" stunting. In contrast, Cluster 1, which had higher sugar consumption and poorer anthropometric indicators, included 50% of the toddlers categorized as "Potential" stunting. These findings underscore the importance of incorporating both dietary and anthropometric data in identifying stunting risk. The results suggest that toddlers in Cluster 1 are more likely to be at risk of stunting due to their nutritional intake, particularly excessive sugar consumption, which aligns with previous research linking poor dietary patterns to stunting.

The Principal Component Analysis (PCA) provided further validation of the clustering results by reducing the dimensionality of the data and enabling a clearer visual representation of the clusters. The first principal component (PC1) explained 54.5% of the total variance in the data, indicating that it captured a significant portion of the underlying structure in the anthropometric and dietary data. PC1 primarily reflected a combination of growth-related variables and sugar intake, with toddlers in Cluster 1 generally occupying the higher end of this component. This visual representation corroborated the numerical validation results, showing that the two clusters were distinct, with only slight overlap, reinforcing the robustness of the clustering solution.

The integration of anthropometric and dietary data, including sugar consumption, provides a more holistic approach to stunting risk detection compared to traditional methods that rely solely on growth measurements. This combined approach has the potential to improve early identification of children at risk of stunting, allowing for more targeted and effective interventions. For instance, toddlers in Cluster 1, identified as being at higher risk, could benefit from focused interventions aimed at improving their nutrition, particularly by reducing sugar consumption and promoting nutrient-dense foods. Programs such as POSYANDU, which already monitor child growth in Indonesia, could incorporate these findings to further refine their risk assessment and intervention strategies.

However, the study had several limitations that should be considered when interpreting the results. First, the sample size was relatively small (N=20), which may limit the generalizability of the findings. The small sample size also affects the statistical power of the analysis, and future studies should include larger and more diverse populations to strengthen the validity of the clustering model. Additionally, the reliance on caregiver-reported dietary intake via a 7-day recall introduces the potential for recall bias, which could affect the accuracy of the dietary data. To address this, future research could incorporate more objective methods of dietary assessment, such as food diaries or direct observation. Finally, other factors that influence stunting risk, such as socioeconomic status, infection history, and access to healthcare, were not considered in this study. Incorporating these additional factors could further enhance the accuracy and applicability of the model.

4. CONCLUSION

This study demonstrates that the unsupervised machine learning approach, specifically using the K-Means clustering technique, effectively identifies patterns of stunting risk in toddlers by integrating anthropometric data and sugar-related nutritional intake. The Elbow Method identified K=2 as the optimal number of clusters, distinguishing children with varying levels of stunting risk. Internal validation through the Silhouette score and

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

PCA confirmed the robustness of the clustering results. These findings suggest that combining basic anthropometric and dietary data with machine learning provides a more comprehensive approach to stunting risk detection than relying on growth indicators alone. This method could complement existing growth monitoring programs, such as POSYANDU, by identifying higher-risk subgroups for targeted interventions.

However, this study is limited by its small sample size (N=20) and reliance on caregiver-reported dietary intake, which may introduce bias. Future research should apply this method to larger and more diverse populations, incorporate additional variables such as socioeconomic status and infection history, and use longitudinal designs to assess the long-term predictive value of these clusters. Additionally, exploring other clustering algorithms or hybrid models combining unsupervised and supervised learning could further enhance the robustness and predictive accuracy of stunting risk profiling.

ACKNOWLEDGMENT

This research was funded by the LPPM Universitas Amikom Purwokerto under the scheme for the 2025 fiscal year.

REFERENCES

[1] N. I. Irsanti and I. Rodiyah, “Peran Posyandu Edelways Dalam Upaya Penanganan Stunting Di Desa Grogol Kecamatan Tulangan Kabupaten Sidoarjo,” *J. Publicuho*, vol. 8, no. 3, pp. 1943–1962, 2025.

[2] Mivtahurrahimah, S. K. M., Epid, M., & Sevtiyani, I., "Unhealthy lifestyle and limited health monitoring as key risk factors for diabetes mellitus: Evidence from the 2023 Indonesian health survey," *J. Epidemiol. Kesehat. Komunitas*, vol. 10, no. 3, 2025.

[3] Nugroho, M. R., Sasongko, R. N., & Kristiawan, M., "Faktor-faktor yang mempengaruhi kejadian stunting pada anak usia dini di Indonesia," *J. Obsesi: J. Pendidik. Anak Usia Dini*, vol. 5, no. 2, pp. 2269-2276, 2021.

[4] K. BKPK, *Buku Saku Hasil Studi Status Gizi Indonesia (SSGI) Tahun 2021*. Kementrian Kesehatan Republik Indonesia, 2021.

[5] A. Daracantika, Ainin, and Besral, "Systematic Literature Review: Pengaruh Negatif Stunting terhadap Perkembangan Kognitif Anak," *J. BIKFOKES*, vol. 1, no. 2, 2021, doi: 10.7454/bikfokes.v1i2.1012.

[6] Siregar, P. A. S., "Urgensi penerapan kebijakan cukai atas minuman berpemanis dalam kemasan (MBDK) di Indonesia," *ETHNOGRAPHY: J. Design, Soc. Sci. Hum. Stud.*, vol. 2, no. 1, pp. 01-11, 2025.

[7] D. Sulistiawati, "Agensi anak dalam pembentukan kebiasaan jajan balita dengan status gizi kurang di Rawa Bogo, Bekasi," *Antropol. Indones.*, vol. 44, no. 1, pp. 1–15, 2023, doi: 10.7454/jai.v44i1.1021.

[8] I. Darmayanti, D. Intan, S. Saputra, and I. Saputri, "Clustering Sugar Content in Children ' s Snacks for Diabetes Prevention Using Unsupervised Learning," vol. 6, no. 4, pp. 2923–2936, 2024, doi: 10.51519/journalisi.v6i4.932.

[9] S. L. Aila, F. F. Dieny, A. Candra, and H. S. Wijayanti, "Added Sugars Consumption Decreased Iron and Zinc Intake among Children Aged 24-59 Months in Central Java Konsumsi Gula Tambahan Menurunkan Asupan Zat Besi dan Seng pada Anak," *Amerta Nutr.*, vol. 7, no. 2, pp. 47–57, 2023, doi: 10.20473/amnt.v7i2SP.2023.47.

[10] W. Jin, "Research on Machine Learning and Its Algorithms and Development," *J. Phys. Conf. Ser.*, 2020, doi: 10.1088/1742-6596/1544/1/012003.

[11] Nurhayati, Busman, and R. P. Iswara, "Pengembangan Algoritma Unsupervised Learning Technique Pada Big Data Analysis Di Media Sosial Sebagai Media Promosi," *J. Tek. Inform.*, vol. 12, no. 1, pp. 79–96, 2019.

[12] N. Chapwanya and K. N. Gorejena, "Hybrid Unsupervised Machine Learning for Insurance Fraud Detection : PCA-XGBoost-LOF and Isolation Forest," *J. Inf. Syst. Informatics*, vol. 7, no. 1, pp. 941–959, 2025, doi: 10.51519/journalisi.v7i1.958.

[13] R. S. Nurhalizah and R. Ardianto, "Analisis Supervised dan Unsupervised Learning pada Machine Learning : Systematic Literature Review," *J. Ilmu Komput. dan Inform.*, vol. 4, no. 1, pp. 61–72, 2024.

[14] J. Yu *et al.*, "Dietary Sugar Research in Preschoolers: Methodological, Genetic, and Cardiometabolic Considerations.," *Rev. Cardiovasc. Med.*, vol. 24, no. 9, p. 259, Sep. 2023, doi: 10.31083/j.rcm2409259.

[15] Apriyani, P., Dikananda, A. R., & Ali, I., "Penerapan algoritma K-Means dalam klasterisasi kasus stunting balita desa Tegalwangi," *Hello World J. Ilmu Komput.*, vol. 2, no. 1, pp. 20-33, 2023.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[16] D. Cytry, S. Defit, and G. Nurcahyo, "Penerapan Metode K-Means dalam Klasterisasi Status Desa terhadap Keluarga Beresiko Stunting," *J. KomtekInfo*, pp. 122–127, Sep. 2023, doi: 10.35134/komtekinfo.v10i3.423.

[17] I. P. Sari, Al-Khowarizmi, O. K. Sulaiman, and D. Apdilah, "Implementation of Data Classification Using K-Means Algorithm in Clustering Stunting Cases," *J. Comput. Sci. Inf. Technol. Telecommun. Eng.*, vol. 4, no. 2, pp. 402–412, 2023, doi: 10.30596/jcositte.v4i2.15765.

[18] M. A. Aziz, L. Amalia, and I. Darmayanti, "Comparison of K-Medoids Algorithm with K-Means on Number of Student Dropped Out," *1st Int. Conf. Smart Technol. Appl. Informatics, Eng.*, pp. 53–58, 2022.

[19] C. Shi, B. Wei, S. Wei, W. Wang, H. Liu, and J. Liu, "A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm," *EURASIP J. Wirel. Commun. Netw.*, 2021, doi: 10.1186/s13638-021-01910-w.

[20] I. Darmayanti, D. Mustofa, N. Hidayati, and I. Saputri, "K - Means and Fuzzy C - Means Cluster Food Nutrients for Innovative Diabetes Risk Assessment," vol. 13, pp. 2175–2182, 2024.

[21] M. Shutaywi and N. N. Kachouie, "Silhouette analysis for performance evaluation in machine learning with applications to clustering," *Entropy*, vol. 23, no. 6, pp. 1–17, 2021, doi: 10.3390/e23060759.

[22] K. R. Shahapure and C. Nicholas, "Cluster Quality Analysis Using Silhouette Score," in *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, 2020, pp. 747–748. doi: 10.1109/DSAA49011.2020.00096.

[23] D. Wulandari, T. Prahasto, and V. Gunawan, "Penerapan Principal Component Analysis untuk Mereduksi Dimensi Data Penerapan Teknologi Informasi dan Komunikasi untuk Pendidikan di Sekolah," vol. 02, pp. 91–96, 2016, doi: 10.21456/vol6iss2pp91-96.